Yanmei Li
Sumei Zhang

Applied Research Methods in Urban and Regional Planning

MOREMEDIA

Springer

Applied Research Methods in Urban and Regional Planning

Yanmei Li • Sumei Zhang

Applied Research Methods in Urban and Regional Planning

Springer

Yanmei Li
Florida Atlantic University
Boca Raton, FL, USA

Sumei Zhang
University of Louisville
Louisville, KY, USA

ISBN 978-3-030-93573-3 ISBN 978-3-030-93574-0 (eBook)
https://doi.org/10.1007/978-3-030-93574-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book has grown out of years of teaching and researching in urban planning and urban studies.

It covers the research process in urban planning from conception to completion, such as identifying the research problems in planning, literature review, data collection and presentation, descriptive data analysis, and reporting findings. It also explains specialized methods in planning and its sub-disciplines. These specific topics include field research methods, qualitative data analysis, economic and demographic analysis, evaluation research, and methods in sub-disciplines such as land use planning, transportation planning, environmental planning, and housing analysis. The book is written in a style that is easy to understand. Rather than presenting complicated mathematical formulas and procedures, the book focuses on explaining rationales and illustrating how methods should be used, using real-world examples. Throughout, we use boxed knowledge points to aid the readers in understanding the key concepts.

The main objective of the book is to help researchers, practitioners, and advanced undergraduate students and graduate students grasp the fundamentals of research methods in urban and regional planning. Readers will gain knowledge in research design, data collection, and data analysis, and communicating key findings to target audiences. Additionally, readers will grasp specific methods in land use, urban economics, demographics, transportation, environmental planning, and housing analysis. We hope that planners can use this wide variety of research methods to help communities handle their social, economic, and environmental opportunities and challenges.

To help the readers grasp the content of the book, the authors compiled a series of Excel files accompanying Chaps. 10, 12, and 15. Along with these data files, a companion document including learning outcomes, chapter highlights, key concepts, review questions, and additional readings ("Electronic Supplementary Materials") will be available per related chapter of the online book on SpringerLink.

Boca Raton, FL, USA — Yanmei Li
Louisville, KY, USA — Sumei Zhang

Contents

1 Research Methods and the Planning Process 1
The Planning Process 2
Research Methods and the Planning Process 3
Smart Cities, Machine Learning, and the Future of Urban Analytics. . . 6
Structure of the Book 7
References. 12

2 Identifying the Research Problem 13
Research Problems Identified from Sub-disciplines of Planning 15
Literature Search for Research Questions. 19
Research Questions Identified from Data 20
Research Questions Identified Through Other Sources 20
References. 21

3 Research Design 23
Project Statement 25
Literature Review 26
Sources of Literature. 26
Systematic Quantitative Literature Review. 28
Writing a Narrative Literature Review 30
Theorizing the Key Concepts 31
Citation Formats 33
Research Methodology 34
Research Ethics. 35
References. 36

4 Data Collection 37
Data and Variables. 38
Variables 40
Observations 41
Census Data Versus Sample Data 42
Sampling Techniques 43

Probability Sampling 43
Non-probability Sampling 46
Data Sources 50
Secondary Data Sources 51
Primary Data Sources 54
Data Collection Through Social Media, Remote Sensing, Real-Time Tracking, and Other Technologies 63
References 64

5 Data Presentation 67
Data, Information, and Geographic Information 67
Information and Planning 71
Data Presentation 72
Audiences 72
Table, Graph, and Map 73
Ten Rules for an Effective Data Presentation 78
Planning Ethics 87
References 88

6 Communicating the Research Findings 89
Effective Communication 89
Plan the Communication 90
Technical Writing Skills 91
Types of Communication Conduit 93
Emails 93
Memorandum 94
Staff Report 96
Research Papers 97
Research or Planning Reports 97
Infographics 99
Proofreading 100
PowerPoint Presentation 102
Public Speaking Skills 103
Engaging the Audience 103
Facilitation Skills 105
Storytelling Skills 106
References 107

7 Statistical Analysis 109
Summary Statistics 110
Measures of Location 110
Frequency Table 111
Measures of Dispersion 112
Probability 115
Basic Concepts 116
Using Probability for Planning Issues 118

Probability Distributions . . . 118
The Central Limit Theorem . . . 125
Confidence Interval . . . 125
Hypothesis Test . . . 126
General Steps of a Hypothesis Test . . . 126
One Sample z-Test of Population Mean . . . 129
One Sample t-Test for Population Mean . . . 130
Two Sample t-Test for Difference . . . 132
Two-Sample t-Test . . . 133
Type I Vs. Type II Error . . . 137
The Simple Linear Regression . . . 139
Dependent Vs. Independent Variable . . . 139
The Concept of Correlation . . . 139
Simple Linear Regression . . . 139
Regression Analysis in Excel . . . 140
Spatial Statistics . . . 143
Appendix 1: Standard Normal Distribution . . . 145
Appendix 2: T-Distribution . . . 147
References . . . 148

8 Qualitative Data Analysis . . . 149
Discovering Patterns in Qualitative Data . . . 150
Frequencies . . . 151
Magnitude . . . 151
Structure . . . 152
Processes . . . 152
Causes and Consequences . . . 152
Types of Qualitative Data . . . 152
Observations and Ethnographic Data . . . 152
Interview Data . . . 153
Focus Groups or Design Charrette . . . 154
Documents and Artifacts . . . 155
Analyzing Qualitative Data . . . 155
Cross-Case Analysis . . . 155
Grounded Theory Method . . . 156
Constant Comparative Method . . . 156
Keywords-in-Context Analysis . . . 158
Conversation Analysis . . . 158
Word Count Analysis . . . 158
Classical Content Analysis . . . 159
Semiotics and Domain Analysis . . . 159
Taxonomic Analysis . . . 159
Componential Analysis . . . 160
Qualitative Data Processing . . . 160
Coding . . . 161

Memoing . . . 161
Concept Mapping . . . 162
Common Software for Qualitative Data Analysis . . . 163
References . . . 164

9 Demographic Analysis . . . 167
Demographic Data . . . 167
Resident-based Versus Employment-based Data . . . 167
The Decennial Census . . . 168
The American Community Survey . . . 168
Describing the Population . . . 171
Size . . . 171
Composition and Characteristics . . . 171
Population Changes . . . 177
Distribution . . . 178
Choosing a Benchmark . . . 178
Extrapolation Analysis . . . 179
Estimates Versus Projections . . . 180
Extrapolation Versus Interpolation . . . 180
Popular Growth Trends . . . 182
The Cohort-component Method . . . 189
Age Cohort . . . 190
The Mortality Component . . . 190
The Fertility Component . . . 193
The Migration Component . . . 194
Set Up a Cohort-Component Model . . . 196
Population Projection Quality Assessment . . . 198
References . . . 199

10 Analyzing the Local Economy . . . 201
The Concept of Economy . . . 201
Economic Activity Classification . . . 202
U.S. Economic Data and Limitations . . . 203
Assessing an Economy . . . 206
The Economic Base Theory . . . 207
Basic vs. Non-basic Employment . . . 208
Base Multiplier . . . 209
Calculating Basic and Non-basic Employment – Quick Approaches . . . 210
Location Quotients . . . 212
Calculating Base Multiplier: The Location Quotient Approach . . . 214
Shift-Share Analysis . . . 215
Set Up a Shift-Share Analysis . . . 217
National Share . . . 218
Industry Mix . . . 221
Regional Shift . . . 221
Employment Decomposition . . . 222

Employment Projection 223
Input-Output Analysis 226
The Transaction Table 226
Input-Output Coefficients 227
Input-Output Matrix 228
The Leontief Matrix 229
Direct, Indirect, and Induced Effects 231
Multiplier Analyses 232
References 234

11 Evaluation Research 237
Fundamentals of Budgeting 238
Ex-ante Evaluation Methods 239
Site Feasibility Analysis 240
Regulatory and Political Feasibility Analysis 242
Market Feasibility Analysis 242
Financial Feasibility Analysis 248
Ex-post Analysis 256
Cost-Oriented Approaches 257
With-and Without Comparisons 262
Actual-Versus-Planned Performance Comparisons 262
Experimental Methods 262
Quasi-Experimental Methods 263
Economic Impact Analysis 263
Social Impact Analysis 263
Environmental Impact Analysis 264
Data Science, Technology, and Evaluation Research 265
References 266

12 Urban Land Use Analysis 269
Land Use Classification 269
Land Use Inventory Analysis 274
Land Use Incompatibility 275
Land Use Indicators 277
Density 277
Diversity 278
Design 279
Destination Accessibility 280
Land Use Suitability Analysis 280
Future Land Development Projection 282
Analytical Programs and Models 286
References 287

13 Transportation Planning Analysis 289
Key Concepts in Transportation Planning 290
Transportation and Mobility 291

Functional Classification of Streets and Roads. 291
Level of Service 293
Transportation Demand Management (TDM) Modeling 293
Trip Generation, Distribution, Mode Choice, and Route Selection . . 294
Parking Analysis 296
Fundamentals of Traffic Data Analysis. 298
Annual Average Daily Traffic (AADT) and Vehicle Miles Traveled (VMT). 298
K-Factor 298
Capacity and LOS 299
Critical Speed, Maximum Flow Rate, and Critical Density 299
Number of Signalized Intersections 299
Transit Data Analysis 301
Ridership and Boarding Estimation 301
Passenger Miles 302
Load Profiles 302
Schedule Adherence Analysis 302
Other Types of Transit Analysis 303
Cycling, Walking, and Other Alternative Mode Analysis 303
Travel Behavior and Safety Analysis 303
Capital Improvement Planning 305
Long-Range Transportation Planning 306
The Future of Mobility 308
References. 309

14 Environmental Analysis 311
Environmental Elements 312
Measures of the Environment 317
The Environmental Inventory Analysis. 318
The Environmentally Sensitive Area Analysis 320
The Environmental Carrying Capacity 323
The Ecological Footprint 323
The Environmental Impact Assessment 326
References. 329

15 Housing Analysis. 331
Land Economics 332
Housing Needs Assessment 334
Housing Supply Analysis 334
Provision of Affordable Housing 337
Fair Market Rent (FMR) and Income Limits 338
Low Income Housing Tax Credit (LIHTC) Program 338
Public Housing Projects 341
Section 8 Housing Choice Voucher Program 341
Housing Demand Analysis 342
The Gap: Housing Affordability Analysis 343

Housing Finance 351
Mortgages and Secondary Mortgage Market 351
Project Financing 355
Pro-forma and Cash Flow 355
Analyzing Housing Market Conditions 357
Housing Construction 357
Housing Price 358
Housing Tenure 359
Housing Vacancy Rates 359
Housing Affordability Analysis 359
Using GIS to Analyze Housing Market Conditions 360
References 360

Chapter 1
Research Methods and the Planning Process

Urban and regional planning has been practiced by cities since the early civilization periods (Morris, 2013). Urban settlements formed on the basis of trade, service, and manufacturing activities. Surplus of rural labor migrated to cities to look for jobs and settled in those cities. The earliest cities in famous civilizations, for example, the Roman Empire, follow a set layout with administrative centers, markets, artisan workshops, and residences in their own zones. Some of the cities may have advanced sewage and water treatment systems.

However, as a formal discipline, urban and regional planning was not established until the late nineteenth century, when the demand for sanitary, orderly, and affordable living environment has become the norm of urban living. Over the years, urban and regional planning has evolved from physical planning, such as land use, zoning, and building codes, to social, economic, and environmental planning. Urban form also evolves from early monocentric cities to polycentric cities. Increased complexity of the urban system demands more sophisticated planning research methods. The planning process has become more scientific and complex. Data collection and analysis has become a key component of the planning process and analytical reasoning has become the foundation of the plan-making process (Wachs, 1986). The methodological skills valued highly by professional planners are ***writing***, ***research***, and ***synthesis*** (Feldt, 1986). Broadly speaking, planning methods should cover basic ***analytical methods*** used in an academic or research setting to generate general knowledge, and the methods to be used in the entire planning and urban design process (applied methods which are more policy-oriented), such as public engagement methods, and negotiation and conflict resolution skills (Baum, 2021; Sawicki, 1989). Due to the limitation in scope, this book will only cover the applied analytical methods, such as demographic analysis, economic analysis, evaluation research, and methods used in planning sub-disciplines such as land use, environmental analysis, transportation

---

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_1].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_1

planning analysis, and housing analysis. This chapter will start with describing the planning process and then highlight the importance of data collection and analysis in the process.

The Planning Process

The planning process as adopted by most planners is often based on the theory of rational comprehensive planning theory. However, with advances in planning tools and technology, the rational comprehensive planning approach has become obsolete. Regardless of the approaches of planning, the basic planning process was described in Fig. 1.1.

Based on Fig. 1.1, the typical planning process is a dynamic process starting from identifying issues, to state goals and objectives, and then collect and interpret data. After data collection and analysis plans are prepared, adopted, and implemented. During the implementation process, new issues might arise, therefore calling for an upward iterative process to improve planning efficiency and effectiveness.

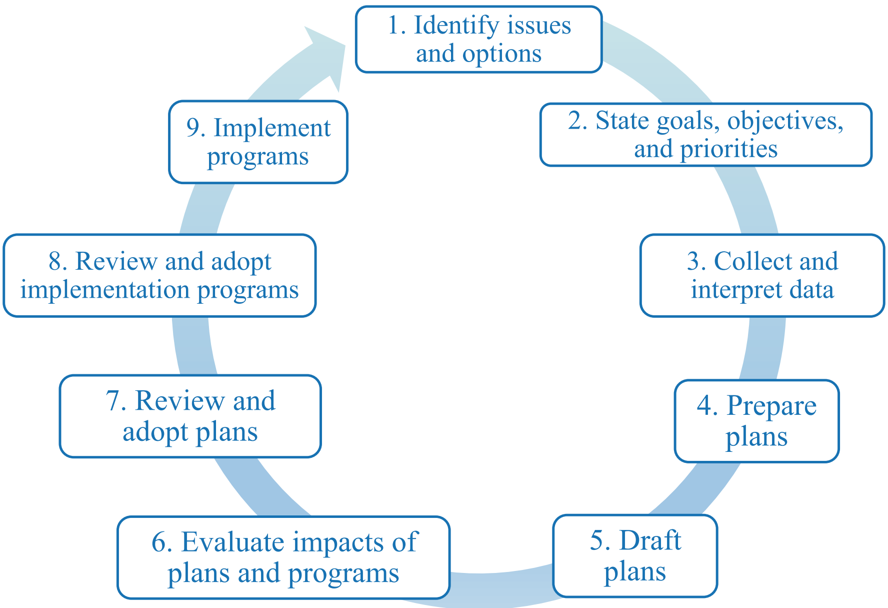


Fig. 1.1 The planning process. (Adapted from (American Planning Association, 2006))

Research Methods and the Planning Process

Research design, which involves the process from the initiation of research to completion, was deemed one of the most important generic skills of planners for entry-level job positions (Schon et al., 1976). The other very important skills are *writing*, *synthesis*, *interaction*, and *consultation*. These skills are deemed very important by more than half of the professional planners surveyed (Schon et al., 1976), while communication skills are the most important among various planning practitioners, followed by the ability to formulate problems and design methods to answer them (Constant & Forkenbrock, 1986; Guzzetta & Bollens, 2003; Ozawa & Selzer, 1999; Seltzer & Ozawa, 2002). Special skills, such as information skills, site planning skills, economic analytic techniques, and quantitative data analysis skills, are deemed very important by about one third of the survey respondents (Constant & Forkenbrock, 1986; Schon et al., 1976). A most recent survey indicates that 97% of the planning practitioners think that research skills are most important, compared to 91.3% planning educators (Miller, 2019). Practitioners also think that time management is very important, while planning educators value more on synthesis and visual communication. Regardless of the differences in views among the planning educators and the planning practitioners, the following skills are regarded the most important for both groups of survey respondents:

- Writing
- Oral communication
- Attention to detail on work products
- Willingness to accept constructive criticism
- Teamwork and collaboration
- Problem solving

The importance of writing and oral communication skills is further affirmed in another study conducted by Greenlee, Edwards, & Anthony (2015). Although planning practitioners stress generic skills more than technical skills, 78.5% think that using spreadsheet to conduct data analysis is very important, 67.5% think that data acquisition from field observations is very important, 74% think acquisition of secondary data from existing sources is very important. About 49% think that descriptive statistics is very important, and only 20.5% think inferential statistics, where statistic models are used to explain data and data relationships, is very important. More than 50% think that specialized knowledge in land use, environmental planning, transportation planning, and community development planning are very important (Miller, 2019), among which land use terminology and regulations are deemed the most useful.

Regarding daily planning tasks that planning practitioners face, Fig. 1.2 ranks percentage planners answering tasks routinely or occasionally used (Constant & Forkenbrock, 1986):

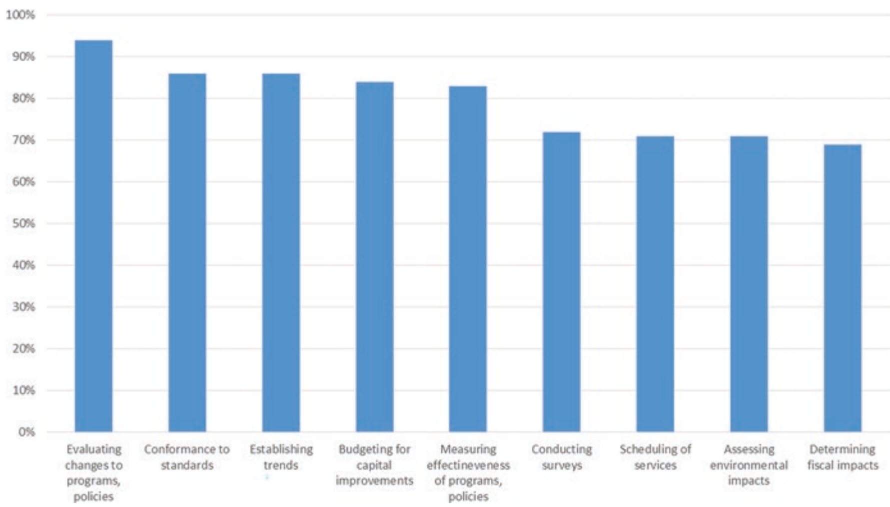


Fig. 1.2 Planning tasks by percentage planners. (Data Source: Constant & Forkenbrock, 1986)

Accordingly, practicing planners view the following analytic techniques most important and relevant to practice (Constant & Forkenbrock, 1986; Kaufman & Simons, 1992):

- Simple regression analysis
- Computer applications
- Cost-benefit analysis
- Descriptive statistics
- Population projection
- Survey research
- Economic base analysis
- Input-output analysis
- Cost-revenue analysis
- Trip-generation modeling
- Market area analysis
- Financial analysis
- Economic impact analysis
- ***Budget preparation***
- ***Capital improvement***
- ***Decision analysis***
- ***Issues analysis***
- ***Scheduling***

The bold and italic methods are those highly used in the planning practice while not sufficiently taught in planning schools (Kaufman & Simons, 1992). The less important techniques include shift-share analysis, multivariate statistics, gravity modeling, time series analysis, inferential statistics, present value analysis, modal split modeling, network analysis, risk analysis, linear programming, scenario construction, factor analysis, and life-cycle costing analysis. The least used methods are econometric modeling and non-linear programming, which are advanced inferential statistical modeling techniques used in forecasting, network optimization, and other more sophisticated applications.

Based on these studies about skills sets valued by planning practitioners, it is imperative to explore applied methods, such as research design, data collection, and data analysis using community demographic data, economic data, land use, housing, transportation, environmental analysis, and other types of data, and for the purpose of day-to-day planning functions either in a public planning agency, or a private planning consulting firm.

Data collection and analysis is one of the key steps in formulating research questions and answering them in urban planning. However, collecting data and analyzing them is not a single independent process. Nearly all steps in the planning process require certain levels of data and analysis. For example, when defining research problems, one needs to find supporting evidence to prove whether a problem is real. The supporting evidence is often based on data, although the analysis may not require advanced statistical methods. Some analysis methods, such as the back- of-the-envelop analysis, issue paper, and the eyeball analysis, are simple analysis methods to estimate the magnitude of an issue or a problem. Back-of-the-envelop calculations use simple mathematical methods, such as additions, subtractions, etc., to calculate the issues at hand. The data sources can be based on published reference sources, surveys, guessing, experts, and other outlets (Patton et al., 2013). Issue paper analysis can be combined with the back-of-the-envelop analysis to identify whether an issue is worth researchable analyses, i.e., further data collection and analysis.

The process of defining a problem requires basic data support. For example, you went to a city to travel. After arriving at the airport, you wondered how to get to the hotel that you would be staying. The city does not have convenient public transit systems. There are a few ways to meet your transportation needs in the city, Uber, shuttle buses, or taxies. Uber and Lyft are called ride hailing services and did not emerge into the transportation market until during the recent years. So you booked your trip on Uber, and soon met the driver and started your trip. On the way to the hotel, you noticed horrific traffic congestion, and it was not the rush hour. Within the days you stayed in the city, you experienced inconvenience of going around and about. You then wondered if traffic congestion is an issue for the city and whether the lack of public transit is one key reason to such congestion. To help you probe your curiosity, which data do you need to identify the issues and problems?

In this situation, you may need a quick search to find out the demographic characteristics of the city, the commuting time of the city, the statistics in public transit, the annual average daily traffic, the visitors coming to the city, and other pertinent information. Many people, including planners, often claim that an issue is real, simply based on subjective assumptions. In this example, you analyzed the information and found that the traffic congestion was caused by a major event happening in the city, attracting thousands of visitors to the city. After the event, the city's traffic is usually very relaxed and there is no significant delay or congestion.

If traffic is indeed an issue, then further analysis is often needed to measure how severe the issue is, and what factors are contributing to the issue. In this case, more data collection and analysis is often needed. When proposing solutions to resolving the issue, further collection of data, such as community needs analysis, will often help planners find the best solution to the research problems.

The above example further demonstrates the importance of incorporating research methods centering around data collection and analysis into each step of the planning process. For example, evaluation research to assess the effectiveness of plan implementation is often used after the plan is implemented, while evaluation research about the estimated costs and benefits of a planning project is often needed before the implementation of the project. Regardless of the types of data needed, for example, numerical counts or interview transcripts, research methods are throughout the entire planning process.

Smart Cities, Machine Learning, and the Future of Urban Analytics

Technological and paradigm discourses are constantly evolving in urban and regional planning. The most evident advances are in transportation planning (such as the use of remote sensing in traffic and travel demand management and monitoring, artificial intelligence, computing technology, autonomous vehicles, flying cars, etc.), smart housing (e.g. connected appliances and internet of things for remotely controlled housing environment), energy efficiency and new energy (e.g. energy efficient appliances, solar energy, wind energy, marine energy, geothermal energy, etc.), healthy homes (e.g. furnishing materials, furniture, and paint with non-toxic and low volatile organic compounds), social media (e.g. Facebook, Nextdoor, Twitter, LinkedIn, Reddit, and Instagram), e-governance, 3D urban simulations (e.g. Remix, UrbanSim, TerraSim, etc.), urban planning mobile applications (e.g. smart parking apps, way-finding apps, etc.), and big data (data collected from real time monitoring and tracing using technologies such as the internet, remote sensing, GPS (Global Positioning System), and mobile devices). These new concepts and technology have added a myriad of dimensions of accuracy and elevated horizons to the traditional data collection, analysis, and communication methods. Figure 1.3 demonstrates the continuum of data collection and analysis in the age of Internet of Things (IoT). Although big data and machine learning is not the focus of this book, we hope readers will be able to integrate newer technologies and methods into the more traditional analytical methods in urban and regional planning.

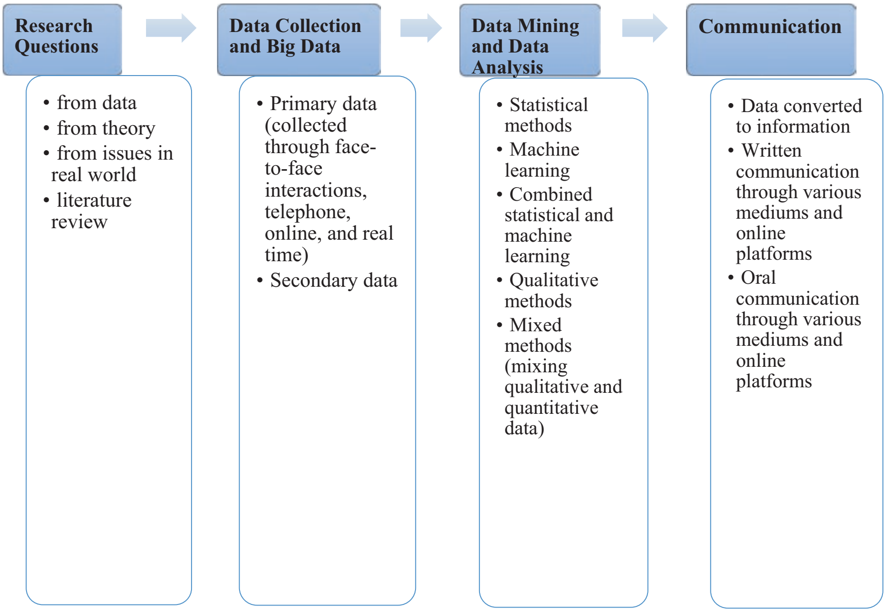


Fig. 1.3 Data analysis continuum

Structure of the Book

This book is written for practicing urban planners, upper-level undergraduate students, graduate students, and scholars or researchers who are interested in research methods. The book will draw upon numerous planning examples and explain how methods presented in this book can be used in real world practice. The book is divided into three major parts, Part I covers research design and data collection methods, along with formats and techniques communicating research findings. Part II are data analysis methods applicable to all planning sub-disciplines, including basic statistical analysis, qualitative data analysis, demographic analysis, analyzing the local economy, and evaluation research. Part III covers specific methods in some planning sub-disciplines, such as land use, transportation, environmental planning, and housing. The three major parts of the book comprise 15 chapters.

Chapter 1 introduces the definition of research methods, and how they relate to the urban and regional planning process. Although there are different approaches to resolving planning issues or making a plan, the basic process of planning goes from problem definition, data collection, data analysis, to reporting findings and using the findings for planning purposes. The unique characteristics and process of planning require basic knowledge in research design, data collection, and data analysis. Most of the data needed for planning are specific to a locale. These data include demographic data, business data, transportation, environmental, and other pertinent information.

Primary and secondary data collection methods thus need to be carefully formulated before collecting the data. Statistical or simple tabulation methods are necessary in conducting descriptive and inferential analyses. All these indicate the critical role research methods play in planning. While conducting research, it is essential and often mandatory for researchers to follow strict ethical principles of research.

Chapter 2 explains the procedures to identify the research problem. The procedure of research involves systematically investigating a subject matter to reveal facts or reach new conclusions. Research can be theory driven or problem driven. Most of the studies in urban and regional planning attempt to investigate issues or problems either to: (1) identity and define the problem, and/or (2) probe further to answer research questions after identifying, defining, and operationalizing the problem. This chapter introduces methods, such as factual data mining, back of envelope analysis, quick research, and issues unique to a sub-discipline of planning, to identify the research problem.

After identifying the research problem and the research questions, Chap. 3 introduces methods to design the research. Research design is the blueprint of how to conduct research from conception to completion. It requires careful crafts to ensure success. The initial step of research design is to theorize key concepts of the research questions, operationalize the variables used to measure the key concepts, and carefully identify the levels of measurements for all the key variables. After theorization of the key concepts, a thorough literature search and synthetization is imperative to explore extant studies related to the research questions. The purpose of literature review is to retrieve ideas, replicate studies, or fill the gap for issues and theories that extant research has (or has not) investigated.

Chapter 4 then discusses the data collection process by identifying sources of data, variables, and number of observations needed. The data sources in urban and regional planning research include primary and secondary data, either based on probability or non-probability sampling methods. Secondary data sources are abundant, ranging from international organizations, such as the United Nations, to national and local organizations, such as the federal government, professional associations, state government, and local government and entities. Key methods to gather primary data will be explored in detail to provide tools for planners to gather field and survey data.

Chapter 5 explores several issues related to data manipulation and presentation. Planners deal with spatial data and need to understand the spatial hierarchy among common administrative and census-designated boundaries, such as counties, census tracts, and block groups. Data can be presented as a table, a graph, or a map. The choice depends on the purpose of the planning project, the nature of the data, and the strength and weakness of a specific presentation method. At last, this chapter presents ten rules of effective presentation for planning practitioners for the purpose of facilitating a planning process. These rules encourage planners to consider their audiences' preferences and limitations, the purpose of the planning project, the pitfalls of the software, and the internal logic within the data. This chapter also introduces technical information about FIPS codes, pivot table analyses, and common graphical presentations in Excel.

Chapter 6 introduces various methods of communicating the research findings. Regardless of the research outcome, researchers need to convey the results and disseminate the findings to the academic community or to the public. Writing the research paper, report, or plan requires careful planning and strong written skills. This chapter explains the fundamental differences and similarities of a paper, a report, or a plan, and present the structure of these different dissemination platforms. Tips and pitfalls to avoid will be covered in this chapter as well. The chapter then goes into the communication skills needed for all planners, with a particular focus on memoranda, staff report, oral presentation, presentation design and crafting, and dissemination of results through public meetings or storytelling events.

Chapter 7 introduces common statistical methods and the applications of these methods in urban issues. Summary statistics describe main features of a set of data with a common theme. For instance, when dealing with the data on housing transactions of a community, summary statistics can help planners quickly understand the existing housing market conditions. Probability theories offer another set of useful tools for planners to identify patterns and make plans accordingly. The chapter first introduces methods that describe a dataset, including summary statistics, measures of location and dispersion, and the frequency table. Then this chapter proceeds with the introduction of the probability theory, common probability distributions, and the central limit theorem. These statistical theories build up the foundation for constructing a confidence interval and conducting a hypothesis test. This chapter introduces various hypothesis testing methods (e.g. z-test, t-test) with great details and real-world examples. This chapter also covers the methods of simple linear regression and the basic concepts in spatial statistics.

Chapter 8 turns to analyzing qualitative data, such as open-ended survey questions, planning documents, and narrative data collected from storytelling, planning workshops, public meetings, public hearings, planning forums, or focus groups. Practicing planners collect these types of data regularly and they are often the foundation of community needs analysis. Analyzing these data requires specialized methods. This chapter introduces methods to analyze qualitative data and conduct content analysis. Identifying trends and patterns of the data is the key to analyzing qualitative data. Related software, such as Atlas.ti, will be briefly explored to help researchers analyze complex qualitative data with complicated content or a large number of observations.

Chapter 9 focuses on demographic analyses. This chapter distinguishes between residence-based and employment-based data and introduces major demographic residence-based data sources. As for how to describe a population, this chapter presents a series of indicators that measure or explore the size, composition, change, and distribution of the population. This chapter introduces two of methods to project a future population: extrapolation methods and the cohort component method. Extrapolation methods are to explore trends. It is highly data dependent. This chapter covers common growth trends, such as the linear growth, the exponential, and the logistic trends. The Cohort-Component analysis decomposes population growth into three components, the mortality (death) component, the fertility (birth) component, and the migration component, and project future population accordingly.

Chapter 10 introduces methods to describe and analyze a local economy and to project future economic activities. This chapter starts with exploring the complexity of the concept of economy. Economic activities can be classified in different ways. The North American Industry Classification System (NAICS) is an industry-based system, while the U.S. Standard Occupational Classification (SOC) is based on occupations. In this chapter, we introduce common economic indicators, as well as their data sources. The Economic Basic theory explains why an economy grows or re-grows. It differentiates between basic and non-basic employment and presents the role of export-oriented industries in stimulating the growth of a local economy through the base multiplier analysis. Location quotients compare the industries in a local economy to their corresponding ones in the national economy and identify the strengths and weaknesses of the economy. Then the chapter proceeds with the Shift-Share analysis. The concepts of national share, industry mix, and regional shift are explained, and real-world examples are presented to illustrate the calculations. At last, this chapter introduces the basic concepts of an Input-Output analysis, including the transaction table, input-output coefficients, Leontief coefficients, direct, indirect, and induced effects, and output, employment, and income multipliers.

Chapter 11 focuses on evaluation research, which is frequently used in planning and public policy. Evaluation research is often divided into two phases. The first phase is called ex-ante evaluation research and the second phase is the ex-post evaluation research. Ex-ante evaluation research is often referred to feasibility studies prior to implementing a planning project or activity, which often investigates the regulatory, financial, market, and political feasibility of implementing the proposed projects and activities. Data sources and tools of feasibility study are identified in this chapter. The second phase of evaluation research is the ex-post research, which evaluates the outcomes after implementing a planning project or activity. Before-after comparisons, experimental or quasi-experimental methods, goal achievement matrixes, and other measurements are often used in ex-post policy and planning evaluation research. The chapter also explains the differences among cost-benefit analysis, cost-effectiveness analysis, and cost-revenue analysis.

Chapter 12 introduces concepts and methods for land use analyses, including land use classification systems, land use inventory and compatibility analyses, various measures for identifying land use patterns, the land use suitability analysis, and the land use demand analysis. We cover five common land use classification systems, including the Anderson Land Classification Scheme, the Cadastral Land Use Classification, the traditional urban land use classification, the Land based Classification Standards, and the Urban Transects. Different systems serve different purposes of urban studies. A land use inventory analysis has analytical purposes of understanding the current level of land supply and explaining how land has been used in the city. A land use incompatibility analysis explores possible adverse effects of different types of land uses imposing on each other. Then this chapters proceeds with common indicators for land use patterns, including density, diversity, design, and accessibility measures. A land use suitability analysis is necessary in cities because of the heterogeneous nature of urban land and land use, as well as the dynamic relation among them. This chapter presents the process of a land use suitability analysis for identifying the most suitable sites for a specific land use. At last, we introduce a handy method that coverts population and employment projections into land use projections.

Chapter 13 covers the data sources and methods of transportation planning analysis. The chapter will explain the process of transportation planning and how it relates to previous chapters of the book. The key methods in transportation planning are to construct and assess the Urban Transportation Modelling System (UTMS). UTMS involves four distinctive steps: Trip Generation, Trip Distribution, Mode Choice, and Trip Assignment. The chapter introduces the UTMS modeling, parking analysis, transit analysis, long-range transportation plans, and capital improvement planning. The future of mobility and transportation will be briefly discussed as well. Although this chapter is introduced separately from housing analysis and environmental analysis, transportation planning should consider integration with housing and environmental planning.

Chapter 14 introduces basic environmental elements, revisits suitability analysis for analyzing environmental sensitivity, and addresses various other environmental analysis techniques. As for basic environmental elements, this chapter covers topography and soil, watershed, climate, microclimate, watershed, air, and biodiversity. This chapter then proceeds with the universal soil loss equation and the air quality index. The universal soil loss equation provides an empirical technology to estimate soil loss by rainfall impact and surface runoff. The air quality index is an indicator for the pollution level in the air. At last, this chapter introduces several environmental analysis techniques that have been commonly used in the field, including (1) the environmental inventory analysis that can help planners to understand the existing physical opportunities and limitations imposed by the environment, (2) the environmentally sensitive area analysis that uses the suitability analysis in environmental issues and identify areas with special environmental attributes worthy of protection, (3) the carrying capacity analysis that identifies the growth limit of an environment and the maximum level of development without weakening the productivity and the functional integrity of the environment, (4) the ecological footprint analysis that studies environment-growth relation from the perspective of the environment, and (5) the environment impact assessment that is to evaluate the potential impacts of a proposed development on the environment.

Chapter 15 delves into one of the most important sectors in an economy and in planning, housing. Housing has significant impact on the household, economy, and the environment. Housing analysis centers around housing market condition analysis, housing finance analysis, development budget and cash flow analysis, housing affordability analysis, housing demand and supply analysis, particularly the demand and supply of affordable or workforce housing, and low income housing analysis. The chapter introduces different methods to evaluate affordable housing needs and gaps in local communities.

References

American Planning Association (Ed.). (2006). *Planning and urban design standards*. Wiley.
Baum, S. (2021). Why urban analysis. In S. Baum (Ed.), *Methods in urban analysis*. Springer Nature Singapre Pte Ltd..
Constant, C. K., & Forkenbrock, D. J. (1986). Planning methods: An analysis of supply and demand. *Journal of Planning Education and Research, 6*(1), 10–21.
Feldt, A. G. (1986). Observations on planning methods. *Journal of Planning Education and Research, 6*(1), 37.
Greenlee, A. J., Edwards, M., & Anthony, J. (2015). Planning skills: An examination of supply and local government demand. *Journal of Planning Education and Research, 35*(2), 161–173.
Guzzetta, J. D., & Bollens, S. A. (2003). Urban planners' skills and competencies: Are we different from other professions? Does context matter? Do we evolve? *Journal of Planning Education and Research, 23*(1), 96–106.
Kaufman, S., & Simons, R. (1992). Quantitative and research methods in planning: Are schools teaching what practitioners practice? *Journal of Planning Education and Research, 15*(1), 17–33.
Miller, E. V. (2019). Assessing the preparation of undergraduate planners for the demands of entry-level planning positions. *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X19873129
Morris, A. E. J. (2013). *History of urban form before the industrial revolution*. Routledge.
Ozawa, C. P., & Selzer, E. P. (1999). Taking our bearings: Mapping a relationship among planning practice, theory, and education. *Journal of Planning Education and Research, 18*(3), 257–266.
Patton, C. V., Sawicki, D. S., & Clark, J. J. (2013). *Basic methods of policy analysis and planning* (3rd ed.). Pearson.
Sawicki, D. (1989). Demographic analysis in planning: A graduate course and an alternative paradigm. *Journal of Planning Education and Research, 9*(1), 45–59.
Schon, D. A., Cremer, N. S., Osterman, P., & Perry, C. (1976). Planning in transition: Report on a survey of alumni of M.I.T.'s Department of Urban Studies, 1960-71. *Journal of the American Institute of Planners, 42*(2), 193–202.
Seltzer, E. P., & Ozawa, C. P. (2002). Clear signals: Moving on to planning's promise. *Journal of Planning Education and Research, 22*(1), 77–86.
Wachs, M. (1986). Discussion of three papers on planning methods. *Journal of Planning Education and Research, 6*(1), 38–39.

Web Resources

Association of Collegiate Schools of Planning: https://www.acsp.org/
American Planning Association: https://www.planning.org/
American Institute of Certified Planners: https://www.planning.org/aicp/

Chapter 2
Identifying the Research Problem

When we collect data and conduct data analysis, we may be doing it for fun, but mostly we do it to probe a question or multiple questions, or to find explanations and answers to an issue or a variety of issues we are concerned about. So what is a ***research problem***? Research problems are the focus of your research, whether it is a phenomenon, a relationship, an issue, a question, or any other forms of focus. When the research problem is identified, converting the problem to a question format will help focus the research efforts in finding the best solutions to the initial research problem. The entire process of your research from conception to completion should closely tie to the research problem. Research problems can be pragmatic, individual, widespread, or societal (Patton et al., 2013). Regardless of the nature of the problem, if it is deemed important by local officials, by the community residents, or by the planners, it may become a researchable problem. For example, as a transportation planner, you are always curious why in certain places, particularly lower income areas, there tend to be more traffic accidents. This concern was raised based on your close monitoring of the traffic accident data in your community. After identifying the research problem, you read some reports and research papers to see if this is a national trend, or international phenomenon, and whether other people have conducted research to probe the potential reasons. Quickly you learned that apparently this is the case for other places as well; lower income areas, oftentimes identified as disadvantaged areas, or economically deprived areas, are often associated with higher incidences of traffic crashes. There have been a few studies attempting to explain why, by exploring all the possible determining factors of traffic crashes, such as demographic characteristics and the built environment factors, including income, racial background, educational levels, occupation, age, gender, population density, employment density, land uses, road capacity, vehicle miles traveled (total annual traffic counts per mile), people's commuting modes (driving, transit, biking, walking, etc.), road design, and other factors. Built

---

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_2].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_2

environment indicates the overall physical environment associated with buildings and infrastructure. All these factors were proven by extant research to be significantly related to traffic accidents.

After identifying the factors that might contribute to traffic accidents, you begin to collect data about these factors in lower income areas, using higher income areas as a comparison group. Remember that income levels regarding high versus low is area specific. For example, a lower income in London might be a high income in Bangkok. Researchers often use government guidelines to define income levels, since many of the subsidy programs funded by governments have established guidelines about poverty line, and the threshold of low income. For example, in the United States, low income is often defined as less than 60% of the Area Median Income (AMI). AMI is the middle value of income in an area as defined by the U.S. Department of Housing and Urban Development (HUD). Median income indicates that half of the households in the area have the income less than the AMI, and half have the income higher than the AMI. Imagine that in your metropolitan statistical area, there are 90 lower income neighborhoods (e.g. less than $40,000 a year) and 140 higher income neighborhoods (e.g. higher than $60,000 a year). Metropolitan statistical area is a geographic unit that often encompasses an entire urban cluster in a region, regardless of the number of administrative units in the area. The concept of neighborhoods may be official, or may be ambiguous. Neighborhoods are often development clusters, subdivisions, street blocks, census geographies (a hierarchy of geographic units solely for the purpose of the census. Census is to collect data about each subject of your study units, while sampling is to collect data about a subset of the subjects in your study unit). Neighborhoods may be also your own mental reflection, often called mental maps in planning and urban design, of the boundaries based on your personal experience and perception.

You collected all the proven demographic and built environment factors for each neighborhood; here you define the neighborhoods based on official subdivisions from the planning and zoning departments of the metropolitan area. Then you run some statistical analysis, descriptive and inferential, to identify the factors that significantly contribute to traffic crashes based on income, in your metropolitan statistical area. Descriptive statistics indicate using basic statistics to describe the data. These basic statistics include statistics measuring central tendency and dispersion of the data. Inferential statistics use hypothesis testing to draw generalized conclusions about the relationships among different factors from your metropolitan statistical area. Methods of conducting descriptive and inferential statistics will be discussed in detail in Chap. 7. You also used some mapping techniques including an array of GIS (Geographic Information System) techniques to delineate traffic accident counts and the values of the determining factors. To make the research findings more meaningful, you conducted some field trips and interviews in a few low-income areas with high incidences of traffic crashes. The quantitative (e.g. numerical data collected from data agencies), and qualitative analysis (e.g. interview and/or document analysis) imply a myriad of reasons why lower income areas tend to have higher incidences of traffic crashes, helping answer your research questions associated with the research problem you identified. After the analysis, you wrote a staff report, not a complete research report, and disseminated the report within your agency. You also proposed some policy recommendations to correct problems presented in the determining factors of higher incidences of crashes in these areas.

Based on this example, it is clear that research problems are the foundations of the entire process of research from inception to completion. They are the guiding compass of your research and therefore make your research meaningful with a purpose. So, how do you identify research problems? Some research problems can be identified fairly easily; while others may need some research. The following sections will guide you to identify research problems from different angles and sources.

Research Problems Identified from Sub-disciplines of Planning

The research example mentioned in the introduction of this chapter is an example of research problems identified from planning practice. ***Planning fields*** are multidisciplinary, comprising of a variety of subjects. The Association of Collegiate Schools of Planning (ACSP)'s annual conference includes the following sixteen (16) tracks:

- Analytical Methods, Technology & Society
- Community Development
- Economic Development
- Environmental Planning & Resource Management
- Gender & Diversity in Planning
- Housing
- International Development Planning
- Land Use Policy & Governance
- Food Systems, Community Health, Safety
- Planning Education & Pedagogy
- Planning History
- Planning Process, Administration, Law & Dispute Resolution
- Planning Theory
- Regional Planning
- Transportation & Infrastructure Planning
- Urban Design

Some of the tracks are not the functional departments of a planning agency, such as gender and diversity, planning theory, planning history, and planning education. Nearly all other tracks are often either in the same agencies, or are separate agencies, depending on the geographic scale of an administrative unit. Administrative units of places are often the units within incorporated government entities, such as cities, towns, counties, villages, and states, etc. Census units are the geographic units solely defined out of the purpose of conducting statistical sampling or census.

Meanwhile, American Institute of Certified Planners (AICP) places much more emphasis on specific areas of practice (APA, 2020). Some of the practice areas overlap with the research tracks defined by the ACSP, while others extend or diverge from the tracks.

- Comprehensive and sectoral planning (e.g., comprehensive plans, long range planning, regional planning, and capital improvement planning, etc.)
- Community, subarea, corridor, and neighborhood planning
- Current planning (e.g., codes and regulations, and zoning, etc.)
- Sustainability planning
- Transportation planning
- Infrastructure planning
- Hazards, mitigation, and resiliency planning
- Environmental and natural resources planning
- Economic development planning
- Urban design
- Housing planning
- Parks, recreation, and open space
- Historic and cultural resources planning
- Institutional planning and siting
- Food planning
- Health planning
- Rural and small town planning

As a practicing planner, a researcher, or an educator in planning education, research problems might arise from the day-to-day practice, or from community input. Given the complexity of the planning disciplines and sub-disciplines, research problems and questions are numerous and nearly impossible to be exhausted in this text. Table 2.1 only demonstrates some of the typical research problems using the 16 tracks defined by ACSP.

In addition to the examples of research questions mentioned above in Table 2.1, numerous research question arise in response to unexpected special events or incidences, such as the Great Recession during 2017-2019, natural disasters (for example, earthquakes, wildfires, and hurricanes), the COVID-19 pandemic, and many other events. The most recent event, the COVID-19 pandemic, has drawn vast attention from researchers and practitioners about the impacts and lessons learned from the pandemic. The majority of the literature focuses on environmental quality (air quality, environmental factors, and urban water cycle), social impact, economic impact, governance, smart cities, and transportation and urban design (Sharifi & Khavarian-Garmsir, 2020). Studies on these unexpected events or incidences will help communities better prepare for the future in emergency management planning.

A researcher in planning may have a much more theoretical view of research problems than practitioners. Individual researchers tend to focus on one or a few related planning disciplines and are more likely to only engage in a narrower range of topics, yet with deeper exploration and probing. For example, as a transportation planning researcher, Joe might be only interested in autonomous vehicles and their policy and planning impact, while Samantha is interested in walkability measures and policies to promote pedestrian safety.

Table 2.1 Typical research problems by functional areas of planning

Planning disciplines	Examples of typical research problems for practitioners
Analytical Methods, Technology & Society	How can GIS be used in planning? Can we create a parking app for our municipal garages? How can remote sensing be used in traffic safety management?
Community Development	How can a federal grant, such as CDBG (Community Development Block Grant), be used to promote quality of life in a community? What initiatives can be used to promote a sense of community? What can be done to promote school quality in a community?
Economic Development	How to attract high-tech companies to a local community? Which business incentives can be offered to women-led small businesses in the area? Will a Walmart super center provide the types of jobs that a community needs?
Environmental Planning & Resource Management	How to improve the air quality in an area? If drinking water quality is compromised in a local community, how can we restore the water quality? What are the best strategies to prevent beach erosion?
Gender & Diversity in Planning	What are the best practices of increasing diversity in planning?
Housing	What strategies can be used to provide more affordable housing? Can a local community establish an inclusionary zoning ordinance to promote affordable housing? How to achieve job-house balance in a metropolitan statistical area?
International Development Planning	How can we learn from the social housing policies in European countries? How can natural resources be preserved among different countries? Which products should we export more?
Land Use Policy & Governance	Should a local community adopt form-based code to improve the physical environment and land use efficiency? Should a local community adopt inclusionary zoning policies? How effective are those policies in providing affordable housing? How can transportation and land use planning be integrated to promote job-house balance?
Food Systems, Community Health, Safety	How does urban design influence health outcomes of residents? Which places have food deserts and how can planners help creating an equitable food distribution plan? What is the impact of the COVID-19 pandemic on urban planning, particularly in urban design, urban density, transportation, and housing choices?

(continued)

Table 2.1 (continued)

Planning disciplines	Examples of typical research problems for practitioners
Planning Education & Pedagogy	How can planning education provide the skills required in the practice? What are the best planning curricula? What content should be covered in each course so that planning education will better prepare students for careers in planning?
Planning History	What is the implication of certain historic movement, e.g., city beautiful movement, in planning? What are the historical patterns of population centers in the U.S.? What factors influenced the shifting population centers? What does the historical climate data tell us about any distinctive climate patterns?
Planning Process, Administration & Law and Dispute Resolution	What are the best methods to increase public participation in the planning process? Which constitutional amendment dictates due process? How does it relate to urban planning? Does planning need consensus among key stakeholders? If so, what are the best strategies to achieve consensus?
Planning Theory	Which planning theory guides the comprehensive planning process? How can we use communicative planning theory in planning practice?
Regional Planning	Is revenue sharing a feasible model to achieve equity in fiscal planning? How can we link transportation planning with regional housing planning to achieve job-house balance? What is the best transportation model to efficiently transport goods and products in a region?
Transportation & Infrastructure Planning	What are the priorities of local transportation improvements? Which incentives are effective in promoting alternative modes (e.g. walking, biking, and/or public transit) of transportation? How to increase transit ridership?
Urban Design	How to design public space to ensure equal access by all residents? What benefits does mixed-use have on urban form? What should be changed in urban design to respond to a pandemic better?

Many planning educators are also planning researchers. Planning educators are interested in educational effectiveness, classroom engagement, community engagement, and how to convey both the theoretical knowledge and practical knowledge to the students.

Due to the differences in the nature of research problems, and the audience and beneficiaries from verifying and solving these problems, planning practitioners and researchers often have different approaches in detailing a research problem. Planning practitioners mostly focus on finding effective and efficient solutions to the research problem, while planning researchers stress dissemination of knowledge, sometimes hoping the knowledge would yield practical policy implications. Therefore, although identifying and defining the problem requires similar approaches regardless of practice or research, detailing, verifying, and inferencing the problems merit different approaches.

Box 2.1 AICP Certification Exam Outline

AICP (American Institute of Certified Planners) is an institute certifying planners who met experience qualifications and passed an exam created by the institute. It is affiliated with the American Planning Association (APA). The certification exam covers: Fundamental Planning Knowledge (25%); Plan Making and Implementation (30%); Area of Practice (30%); Leadership, Administration and Management (5%), and AICP Code of Ethics and Professional Conduct (10%) (APA, 2020). Research problems and questions arise from all these aspects of the exam areas; however, most of the practical research problems can be found from plan making and implementation, and functional areas of planning practice. The following are the key topic areas of the plan making and implementation process:

- Conducting research and acquiring knowledge
- Spatial analysis
- Public engagement
- Communication
- Preparing to plan
- Formulating plans and policies
- Plan implementation
- Monitoring and assessment
- Project or program management
- Social justice

Literature Search for Research Questions

Planning researchers often use literature search and synthetization to derive research problems. To make a study more meaningful, researchers often use the criteria such as whether the research filled a gap to judge whether the study has its merit. What is a research gap? A research gap means to focus on something new and has not been done about a research topic that has been studied numerous times by different scholars. Therefore, is there anything new about your research compared to all the prior studies? If it is new, will this newness be significantly contributing to the general knowledge in the topic area? Will it relate to any significant policy recommendations and implications? When conducting research, practicing planners often use different methods from planning researchers, and the research findings may not be required to be new or significantly contributing to general planning knowledge. ***Literature search*** often starts from online search using a search engine. A commonly used search engine, such as Google and Yahoo, should be sufficient when searching for relevant literature or documents. A quick browse through the literature may help one identify a researchable problem. If within an academic institution, one can also search the library sources to find articles or documents. Writing academic papers also needs to synthetize literature into a format called literature review. How to write an effective literature review will be covered in Chap. 3 Research Design. Planning practitioners often search for literature to find best-case examples and guiding principles in planning. The literature search and review methods covered in this book can be used for both the practitioners and the researchers alike.

Research Questions Identified from Data

Data are often the best source to identify researchable questions, particularly in the case of academic and scholarly research. For example, the *American Housing Survey* (AHS) has different categories of data, such as housing attributes, housing finance, doubled-up (multiple households living together), neighborhood characteristics, neighborhood satisfaction, and housing rehabilitation. Researchers will be able to find research questions from the variables included in the survey, then narrow down the research questions after quick exploration of literature. One research question might be how renters and owners are different regarding neighborhood satisfaction, after controlling for other household-level variables, such as income, educational background, length of residency, and ethnic and racial background. Each dataset has its own limitations, and these limitations should be considered while searching for the best research questions. For example, compared to *American Community Survey* (ACS), American Housing Survey (AHS) is based on housing units, not households. Any research questions exploring household level perception and activities need to be mindful about these limitations.

Research Questions Identified Through Other Sources

Research questions may be also identified through personal experiences, social media, news articles, public meetings, focus groups, and other types of outlets, particularly for a practicing planner. These questions sometimes are anecdotal, and therefore need verification and detailing. A quick analysis, such as back-of-the-envelope analysis or eyeball analysis, of basic data or evidence might help verify the validity of the research questions. For example, in Chap. 1, the traveler's personal experience of traffic issues in a city may manifest as a researchable problem, if the traffic issue is inherent to the city instead of being caused by a major event.

References

APA (American Planning Association). (2020). AICP Certification Exam Outline. Retrieved on 22 Sept 2020 from https://planning.org/certification/examprep/subjectmatter/

Patton, C. V., Sawicki, D. S., & Clark, J. J. (2013). *Basic methods of policy analysis and planning* (3rd ed.). Pearson. New York City.

Sharifi, A., & Khavarian-Garmsir, A. Z. (2020). The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. *Science of the Total Environment, 749*(2020).

Web Resources

American Planning Association: https://www.planning.org/
Cyburbia, Urban Planning Community: https://www.cyburbia.org/
Planetizen: Planning News, Jobs, and Education: https://www.planetizen.com/
Bloomberg City Lab: https://www.bloomberg.com/citylab

Chapter 3
Research Design

Previous chapters of the book explained the relationship between research and the planning process (Chap. 1) and how to identify a researchable problem in urban and regional planning (Chap. 2). Once a research problem is identified, it is imperative to identify the key concepts which define the research problem. Then a thorough literature search should be conducted to explore how these key concepts are defined and whether these concepts have any significant relationships among them. Many planning practitioners or students often ask whether this step is necessary when writing a research report. Some of us may have already known that research reports and academic papers may have different structures. For example, a research paper which is published by a journal or book, or presented at a conference, will have a distinctive structure compared to a research report that a planning consultant writes for a planning agency. Regardless of dissemination methods, research papers tend to have a standard format across nearly all academic disciplines (as shown in Fig. 3.1). It is onion-shaped, meaning that the introduction and the conclusion sections need to fit to a broader audience, while the other sections, such as literature review, data, methods, and analysis are often tailored to a more specialized audience (Fig. 3.1).

A research report, on the other hand, does not have a standard structure but benefits from a literature search and review. A report or a plan usually starts with an executive summary highlighting the report, followed by a table of contents (sometimes a list of tables/figures as well), introduction, methods, analysis, and results or plans. Literature review is often embedded in the content and not explicitly mentioned as a review of extant studies. Sometimes the literature review is defined as a review of guiding planning principles or best-case planning studies. Citations are not straightforward and often formatted as end or foot notes.

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_3].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_3

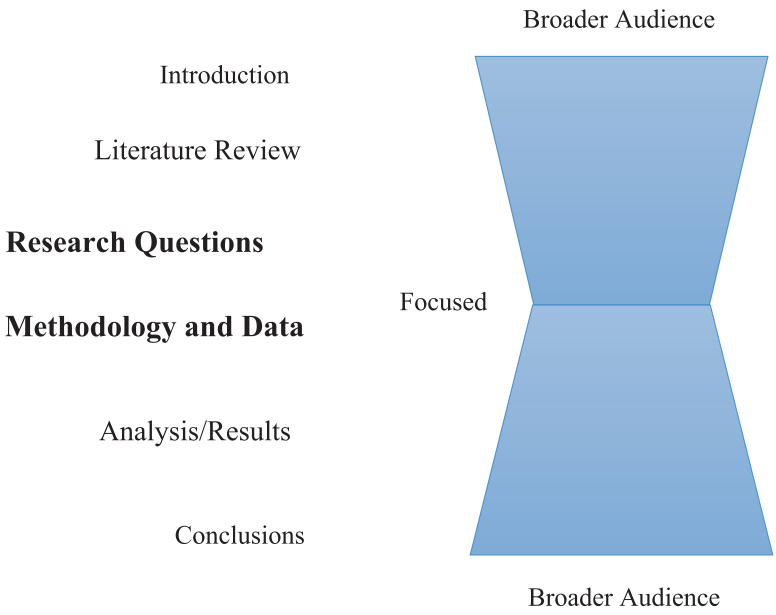


Fig. 3.1 The structure of a research paper

Regardless whether the final delivery format is a report, a portfolio, or a paper, the key to successfully completing any study is to carefully design the research, and to identify any data sources needed in order to answer the research questions.

Research design is the process to deliberately plan for your research. As mentioned before, identifying the key concepts related to the research problems is the initial step to success. Once the key concepts are identified and mapped to highlight the potential relationships, a literature review is imperative. Then from the literature review and exploring the current data options, data sources of these key concepts should be identified. Literature review and knowledge in data analysis will then determine the appropriate analytical methods to be used to analyze these concepts and their relationships. Specified analytical methods will be covered from Chaps. 5, 7, 8, 9, 10, 11, 12, 13, 14 and 15 and therefore this chapter will focus on literature review, theorizing, and citation format. Due to their complexity, data collection methods will be explained independently from the research design section. In a nutshell, the following is the procedure of research design:

1. Define the purpose of your project. Determine whether it will be exploratory, descriptive, or explanatory.
2. Specify the meanings of each concept you want to study.
3. Select a research method.
4. Determine how you will measure the results.
5. Determine whom or what you will study.
6. Collect empirical data.

7. Process the data.
8. Analyze the data.
9. Report your findings.

Project Statement

Chapter 2 discusses the process of identifying researchable problems through various methods, such as literature search, quick decision analysis, and field observations. During and after identifying these researchable problems it is imperative to theorize the nature of the problem, why it is important, and the scope and dimensions of the project. The project statement needs to identify the major and minor issues relating to the project, the types of activities to address these issues meaningfully, the relevant planning precedents that inform these issues, and the implications and benefits of addressing the research questions and all the issues associated with or emerged from the project.

For example, when conducting research about redeveloping a dilapidated and mostly vacated shopping mall in a suburban city, a well-defined project statement should include at least the following elements:

Project statement structure	Content
Introduction	Location, brief history, and socioeconomic conditions, and identify potential issues
Significance	Why is it important to redevelop the mall? If the city does nothing, what will happen? What is the tradeoff between doing nothing versus redevelopment?
Assessing community needs	What data are needed to assess the community needs regarding the reuse? If there is a need to redevelop the mall, what should be the proposed uses? What are needed by the community? These are the most important questions that the project needs to ask. Community needs assessment often starts with demographic analysis to find out demographic segments, and consumption and expenditure patterns. Demographic census and ESRI business analyst provide such information in the U.S. Other countries may have similar datasets available as well. Current economic and business conditions can also help pinpoint industries and businesses that have a surplus or a leakage (insufficient supplies). The needs assessment should also be from primary data, such as surveys or interviews of the residents or key informants. If from residents, how to select the samples and how to collect the data? If key informants, who should be included?
Proposed plan and guiding principles/ aspirations	After identifying the current community needs, which future land uses should be proposed? What is the cost (socioeconomic, financial, fiscal, or environmental) versus benefits (socioeconomic, financial, fiscal, or environmental) of having such developments? If the benefits are larger than the costs, which planning principles should the redevelopment follow? What will be the new brand of the area? Planning principles and branding will become the guiding foundation of site plans.

Project statement structure	Content
Implementation and monitoring	Once the master plan is developed, how to implement the plan? Who will be the project managers, what are the funding sources, and how to ensure the successful development of the project? How can the project be monitored during and after the development? What will be the lifecycle term for the project when it becomes obsolete?

Literature Review

In basic academic research, a literature review helps identify the research questions, provides theoretical foundations to address the questions, and to amplify the importance of these research questions. In applied research, a literature review will help reveal the best case studies, the guiding principles of a planning project, and the importance of the research questions.

There are different approaches to research. ***Inductive research*** is to derive theories from empirical data. ***Empirical data*** are those data generated from the real world or in a lab or experimental setting. ***Deductive research*** is to use data to test or prove a theory (Fig. 3.2). Whether a study is deductive or inductive, or using combined methods, each researcher needs to identify the research problems that he or she intends to tackle.

Identifying research problems can be from quick analyses, reading previous studies, or be generated from analyzing certain data. For example, a student read local newspaper and heard about water contamination in a local community. Then she was curious if her own community and other communities have the problem, how it affects people's health, and what the overall solutions are. First of all, she needs to learn about different contaminants, and the legal levels of these contaminants as set up by regulating agencies. Then she needs to read more news articles and published papers or books to learn more about this issue. She may also need to learn about ways to help reduce or eliminate these pollutants. As illustrated in this simple example, finding a research problem and exploring all the key concepts and their relations need extensive search for literature covering these topics. Broadly speaking, literature can be in any format; although the preferred format is formally published journal articles, books, and reports. Certain online sources, such as Wikipedia, should be viewed with caution since some content may not be reviewed by experts in the field.

Sources of Literature

Digitization and automation has progressed greatly over the recent years. Searching for literature no longer requires heading to a library and browsing through all related collections. Online search is the primary means to find relevant literature. The

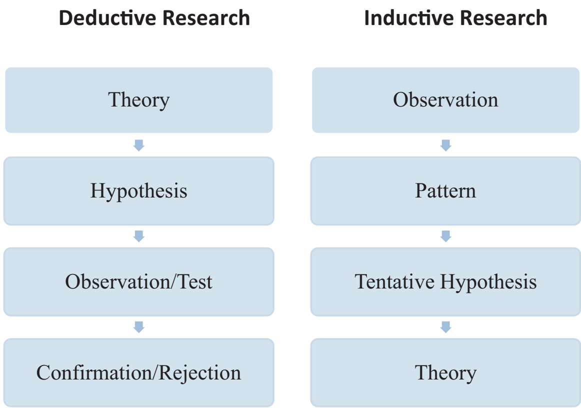


Fig. 3.2 Deductive research versus inductive research

starting point is usually from university libraries and search engines like Google. *University libraries* routinely subscribe to various journals and acquire books for uses by their employees and students. Key words, or author search, through the electronic databases these libraries subscribe to, will be able to generate a large amount of related literature from various databases such as the Web of Science and ProQuest. However, one limitation is that whoever not affiliated with the institution will not be able to have full access to all the subscribed resources. Another useful starting point is Google Scholar (http://scholar.google.com). On *Google Scholar*, one can use key word search to find all published literature. However, except open source information it may still need a subscription to gain access to the articles. Abstracts of journal articles, however, will be fully displayed in any of the search engines or journal databases regardless of subscription status. For planning reports and other types of nonacademic publications a search in Google or Yahoo will generate a list of relevant literature. Although literature search for applied planning research often relates to planning reports, plans, land use and zoning documents, local ordinances, and case studies, theoretical literature regarding planning principles, foundations, which principles are better, etc., may require extensive search of academic research articles or books. When using key word search on these library or external search engines or databases, exact match can be generated when using double quotation marks to enclose the key words. For example, if you are interested in any literature in transit-oriented development, you can use "transit-oriented development" as the search key words (including the quotation marks) so that the phrase will not be fragmented during the search process. One tip to find literature is to find the first relevant article, then go over the reference list and find more. Overall, literature is usually found through the following sources:

- University or public libraries, especially the electronic databases.
- Online search engines such as Google or Yahoo. A very helpful tool is Google Scholar.
- Once you found one paper that is highly relevant to your research topic, it is easier to find more literature from the reference list of that paper.

Systematic Quantitative Literature Review

Systematic Quantitative Literature Review (SQLR) is a systematic review of literature discovering the interrelationships between key concepts and terms. It is systematic, highly visual, and complex, when being compared with the traditional narrative and meta-analysis reviews (Pickering, Johnson, & Bryne, 2021). Traditional narrative reviews usually summarize study findings while systematic literature review may also include the methods and criteria of finding relevant literature, and the quantified analysis of keywords and key themes of these studies. Meta-analysis of literature is often used in health and educational studies, where the reviewer takes several existing quantitative studies and generate composite statistical results from these quantitative studies, using methods such as weighted averages to provide more comprehensive analysis of the existing literature. On the other hand, SQLR is a step-by-step analysis of the literature for the research topic using analytical software and is able to analyze thousands of literature simultaneously.

SQLR is highly visual, using some of the newer graphic techniques, such as ***Word Clouds*** (Fig. 3.3). In a word cloud, words that appear more frequently appear more prominent. Various online applications can help create a word cloud. Other graphics include spider diagrams, tree maps, stacked bar charts, Venn diagrams, geographical heat maps, etc. Statistical procedures such as the Chi-square analysis and multivariate analyses are used to compare multiple themes to identify more complex patterns among these themes.

Key word network analysis is another approach and representation in SQLR methods. The website, www.connectedpapers.com, offers the linkage of published literature based on key words. For example, when searching for "climate gentrification", all the relevant articles appear in the search results. One can choose any of these to map the linkages to other publications. A result from the search is presented

Fig. 3.3 "Congressional Tweet Word Cloud (research in progress)" by justgrimes is licensed under CC BY-SA 2.0

in Fig. 3.4, where larger bubbles indicate stronger connections while smaller bubbles indicate weaker connections.

To conduct a SQLR analysis, the researcher should start with determining the research topic, formulating research questions, and selecting the keywords highly relevant to the research questions (Pickering, Johnson, & Bryne, 2021). Once the key words are determined, a web or library search (e.g. Google, Google Scholar, Scopus, Web of Science, Science Direct, and ProQuest Central) will help generate a list of publications centering on these key words. To prepare for the quantitative literature analysis, a data in spreadsheet format should be then generated. Table 3.1 provides an example about the variables included in the pool of literature. The selection of variable included in the data is dictated by research questions and therefore should not always be the same as Table 3.1.

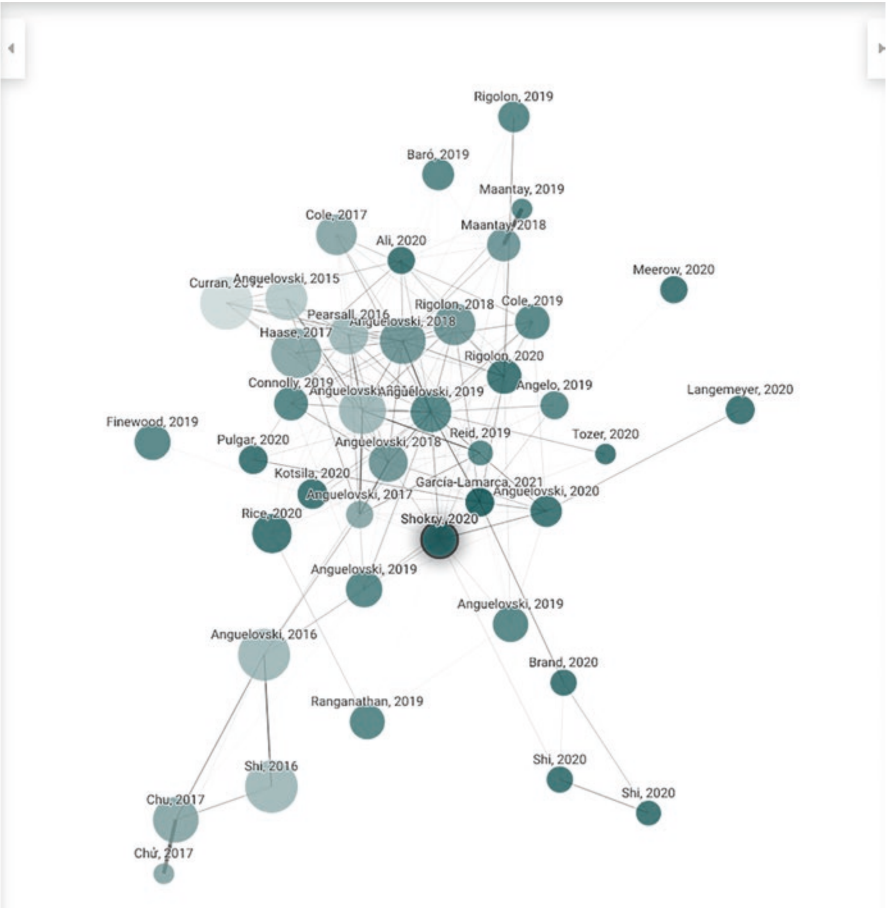


Fig. 3.4 Literature connection network map. (Created by author from https://www.connectedpapers.com, using the key word "Climate Gentrification")

Table 3.1 Database structure of the literature

Author	Date of Publication	Title	Journal Name	Data	Methodology	Key Findings	Study Area	Key Words

Adapted from Pickering, Johnson, and Bryne (2021)

During the literature search process SQLR usually requires well-defined and precise criteria for the keywords. This means that for Table 3.1 each keyword needs to add more specific columns such as "defined, discussed, or demonstrated" (Pickering, Johnson, & Bryne, 2021). Overall, the SQLR procedure should go through the following 15 steps (Pickering, Johnson, & Bryne, 2021):

Step 1: Define topic.
Step 2: Formulating research questions.
Step 3: Identifying keywords.
Step 4: Identify and search literature databases.
Step 5: Read and assess publications.
Step 6: Establish a structure database.
Step 7: Enter the first 10% papers into the structure database.
Step 8: Test and revise categories.
Step 9: Enter bulk of papers.
Step 10: Produce and review summary tables.
Step 11: Draft methods.
Step 12: Evaluate key results and conclusions.
Step 13: Draft results and discussion.
Step 14: Draft introduction, abstract, and references.
Step 15: Revise paper.

Writing a Narrative Literature Review

Once the literature search is complete there are various ways to organize the literature. Reference applications such as EndNote and RefWorks play important roles in organizing the literature. When writing about the literature citation index stored in these types of reference software can be directly cited in Microsoft Word. However, the traditional approach is to store all the literature in the computer, read them, make notes, and then embed and cite them in the literature review. There are two ways to write a literature review. The first is annotated literature review. Annotated literature review is a list of literature organized by title, summary, and contribution. Annotated literature review is usually the first formal step before writing a synthesized review. However, annotated literature review only has very limited applications. Only journals or periodicals specializing in literature review will publish annotated literature review. In most cases, when writing a research paper or report a synthesized review, instead of an annotated review, should be produced. Writing a synthesized review is challenging to many, including established scholars. To synthesize the review, reading and categorizing the literature is imperative. The writer needs to understand when to cite the source and then pull the source out easily to cite them. One approach to handle this is to read the literature first, then think about how to organize the literature review. Once you have the structure of the review, you start drafting the frame of the review as if you are writing an essay. When writing the essay, keep in mind of the literature that you are familiar with and use them to support your writing. Keep the thoughts flowing while writing and constantly cite literature to support them. Using this approach will avoid writing a fragmented literature review which is solely a summary of the literature instead of a cohesive essay of the literature review. Read the following example and think about how the author wrote the review (Li & Walter, 2013).

Subprime lending represents 51% of home loans in black neighborhoods comparedwith 9% in white areas (U.S. Department of Housing and Urban Development, 2000). When comparing high-income black areas against low-income white neighborhoods, high-income black neighborhoods have twice the rate of subprime loans compared with low-income white neighborhoods.

Subprime lending often leads to foreclosure. Therefore, foreclosures are spatially concentrated. Foreclosures are disproportionately located in low-income minority neighborhoods (Belsky & Drew, 2007; Bocian et al., 2008; Bocian, Li, Reid, & Quercia, 2011; Courchane et al., 2004; Goldstein, 2004; Nichols et al., 2005). About 20% of the mortgages in low-income neighborhoods and 25% of the mortgages in high minority neighborhoods experienced foreclosures or were delinquent (Bocian et al., 2011). Other research indicates that high-income black borrowers also tend to be targeted by subprime loans and thus are more likely than other racial and ethnic groups to foreclose on their properties after controlling for the effect of income (Anacker, Carr, & Pradhan, 2012). Among the Asian and Hispanic populations, the higher the income, the more likely they will experience foreclosures (Bocian et al., 2011).

Reading extensively will help tremendously in writing. When reading, critically analyze how the author organizes the article or report. Learn how the literature review is written. Yet, writing a high quality review needs a large amount of practice. In this case, "practice makes perfection" stands its truth.

Theorizing the Key Concepts

The purpose of writing a review of literature is to help theorizing the key concepts in your research questions. The review will reveal what has been done to explore these key concepts and their relationships. The review will also reveal what has not been done but it is important for your research to fill the void. There are two purposes of your research: one is to fill the gap of an issue or question that the current literature has not addressed, and another purpose is to testify whether the findings from extant literature hold when parameters or geographies change. Therefore, literature review helps to derive your research questions and hypotheses. Only after finalizing the research questions and hypotheses can research methods be formulated. *Hypotheses* are testable assertions. In statistical terms strict hypothesis testing needs to have a pair of hypotheses: null hypothesis and alternate hypothesis. *Null hypothesis* is the assertion assumed true, while *alternate hypothesis* is usually what you believe true based on the empirical data. Null hypothesis and alternate hypothesis are mutually exclusive yet cover all possible propositions about your research questions. In order to prove that your hypothesis holds, you need to reject the null hypothesis, which is the assumed truth, since rejection is usually easier than proving the assertion true. No amount of evidence will be enough to prove somebody is innocent; however, one account of wrong-doing will indicate the person is guilty. The following is the summary of a literature review and how it relates to the study the article intends to focus on (Li & Walter, 2013):

> These studies indicate that foreclosures tend to be sold at a discount vis-a-vis comparable market-rate properties and consequently take a shorter time to sell. Regardless of the quick turnover of foreclosed properties, the time between mortgage delinquency and foreclosure sale is often lengthy, particularly in states with judicial foreclosure procedures (Pennington-Cross, 2003). Before the recession, it took an average of 28 months before a foreclosed property became REO and another six months to sell as an REO (Pennington-Cross, 2003). It is not clear which foreclosures sell faster and which factors determine the market duration of different types of foreclosures. Studying the market duration of REO properties will help investors and policymakers understand the foreclosure resale market when making investment and policy decisions.

> Immergluck (2012) analyzed the lower price REO markets in Atlanta and found that many low-value foreclosure properties are often "dumped" to smaller investors. Lee and Immergluck (2012) noticed that the pace of REO disposition, particularly low-value REO properties, accelerated between 2007 and 2009. Lender type, neighborhood characteristics, and submarket location all determined the rate of acceleration of REO resale. This supports our hypotheses that REO durations rely on multiple factors. Although there are many unanswered questions about foreclosure resale, our research will focus on the market duration of REO properties before they are resold. We hypothesize that market duration for each housing submarket is significantly different and the duration is determined by housing attributes and neighborhood characteristics. The difference between our research and Lee and Immergluck (2012) is that we did not distinguish among different years and we used a mixed methods approach. GIS techniques, such as hot-spot and nearest neighborhood analyses, and a shared frailty model by controlling for the classification of REO submarkets by property values were employed. Therefore, our research has policy implications and methodological merits.

Note that the hypothesis indicates "We hypothesize that market duration for each housing submarket is significantly different and the duration is determined by housing attributes and neighborhood characteristics." This is the alternate hypothesis. Therefore, the null hypothesis is "market duration for each housing submarket is NOT significantly different and the duration is NOT determined by housing attributes and neighborhood characteristics". Note the mutually exclusive nature of the pair of the hypotheses. Detailed introduction of hypothesis testing will be in Chap. 7 Statistical Analysis.

Citation Formats

When reading the two examples of literature review in the previous sections you may have noticed that literature citation was embedded into the main text. In-text citation is usually easier and it almost always includes the author (usually the last name only if the author is individuals) and publication year. Whether to use a comma between the author's name and publication year is up to specific journals and citation format. In-text citation is not enough to cite the sources since readers will wonder what the in-text citation entails. Therefore, a list of references will accompany the main text, either in a form of end notes, foot notes, or references following the main text. The list of references needs to follow certain citation styles. Citation format mostly defines how you write the list of references.

There are different types of citation format. ***APA*** (American Psychological Association) is one of the most commonly used citation styles. The following examples are some of the uses of the style in a reference list.

- *Citing a paper published in a journal found online*

Cervero, R., & Duncan, M. (2006). Which reduces vehicle travel more: Jobs-housing balance or retail-housing mixing? *Journal of the American Planning Association*, 72(4), 475–490. doi: https://doi.org/10.1080/01944360608976767

Note: Eliminate the DOI (Digital Object Identifier) serial number if the paper is found from a printed journal or there is no DOI assigned.

- *Citing a printed book*

Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House

- *Citing a chapter in an edited book:*

Florida, R. (2011). The creative class. In LeGates, R.T. & Stout, F. (Eds.), *The City Reader* (pp.163–170). New York: Routledge.

- *Citing a newspaper article found online:*

Brasuell, J. (2017, June 15). Critiquing the proposal to locate a bike lane on an interstate. *Planetizen*. Retrieved from: https://www.planetizen.com/node/93252/critiquing-proposal-locate-bike-lane-interstate.

- *Citing an online report:*

Washington D.C. Office of Planning. (2011). *Comprehensive Plan*. Retrieved from: https://planning.dc.gov/page/comprehensive-plan

For in text citations the format is much easier to grasp. For example:

Li (2013) found that residential foreclosures concentrate in low income minority neighborhoods…

Density decrease of 10% requires a significantly larger amount of investment in infrastructure than when density increases by 10% (Smith, 2007; White & Wenning, 2015).

If the source has two authors, always include both names in each in-text citation. If the source has three, four, or five authors, include all names along with the date when the citation first appeared in the main text. In the following in-text citations, only include the first author's name and follow it with et al. For example: 1st in-text citation: (Wang, Li, & Smith, 2015), 2nd and any other subsequent citations: (Wang, et al., 2015). If the source has six or more authors, only include the first author's name in the first citation and follow it with et al. For example: (Zhang et al., 2017). Include the page numbers following the publication year, if there is a direct quote.

Detailed and comprehensive conventions about the APA citation format can be found online or from a library. Purdue University Online Writing Lab has a detailed section explaining the usage of the APA style and the URL (Uniform Resource Locator) link can be easily found through an online search.

Research Methodology

Chap. 4 through Chap. 15 will explain data and different types of methodology in detail. Chapter 4 summarizes the types of data and the basic principles of choosing the best methodology to collect and analyze the data. Research methodology is the process or steps of conducting the research from inception to completion, focusing on the kinds of research tools and procedures needed to obtain representative data and analyze the data (Opoku, Ahmed, & Akotia, 2016).

The selection of different research strategies, such as survey, case study, use of secondary data sources, or ethnography, depends on the research question(s), research objectives, existing literature, available time and resources, and other factors. Methods of analyzing the collected data are part of the methodological framework of qualitative, quantitative, and mixed methods.

A ***qualitative research method*** focuses on methodology addressing research questions that cannot be answered by way of quantification. For example, reasoning of behavioral issues, residential preferences in land use decisions, consumer biases towards certain products, etc., may benefit from using a qualitative research methods. Qualitative method usually analyses interview transcripts, documents, participatory behavior, artifacts, and other non-quantified data or subjects.

*A **quantitative research method***, on the other hand, stresses data objectivity, reliability, and validity by analyzing quantitative data using statistical methods. Quantitative research tends to have a statistically testable hypothesis and then chooses appropriate methods to reject or do not reject the hypothesis. Quantitative data often requires high levels of representativeness, and therefore, randomness and the power of representativeness are required in sampling techniques. For example, qualitative research may use interview data from a small number of people to answer a research questions; while in quantitative research the small number of people need to be randomly selected with the right sample size, and represent the characteristics of the population that the sample is from. Otherwise, quantitative analysis will be meaningless.

A ***mixed method*** approach allows the researchers to answer both the qualitative and quantitative research questions simultaneously, by combining qualitative and quantitative methods in a single study, being concurrently or sequentially carried out (Borrego, Douglas, & Amelink, 2009; Creswell, Plano Clark, & Garrett, 2008).

Once the research methodology is determined data collection is often the process that follows. There are various methods collecting qualitative and quantitative data, but there are fundamentally two approaches of data collection: *Primary data collection* and *secondary data collection*. Primary data collection focuses on collecting data first hand, without using any other sources. Secondary data collection uses data collected by other people or entities. Therefore, primary data collection tends to be more time consuming, more difficult, and has uncertainties about the quality of the data collected. Data collection methods will be discussed in Chap. 4 in detail.

Once the data is collected, describing the data is often the first analysis step, regardless whether it is qualitative or quantitative data. Further, different methods will be used to explore more complex relationships among variables. For example, in quantitative methods a regression model may be used to probe the statistical relationship between the independent and dependent variables. While in qualitative methods, thematic coding, concept mapping, and other layers of analysis will be added to basic data description. Chapters 5, 7, 8, 9, 10, 11, 12, 13, 14, and 15 will explain detailed methods that can be used to analyze the data for the research project.

After data analysis, research findings need to be communicated to the intended audience. Research reports, research papers, staff reports, etc., are often the common approaches communicating the results. Chapter 6 will cover the details about communicating the research findings. The continuum of research methodology is to reiterate the procedure when needed to further test, challenge, or fill a void as presented in previous research findings. The process of research, therefore, becomes more meaningful to help advance knowledge.

Research Ethics

As briefly mentioned before, when a study is about human or animal subjects the researcher should follow strict ethical principles. Research institutions and higher education institutions often routinely require a researcher to go through the ethics training and submit the research design for approval. The training is mandatory prior to submitting the research proposal for review. Detailed training content in research ethics would cover all relevant information; therefore, this section will only mention the key elements of research ethnics.

The first ethical principle is ***no risk or minimal risks*** to the research subjects. Regardless whether there may be no risk or some minimal risks involved in participating in the research an ***informed consent*** is often obtained from the participants. The participants have the right to maintain ***anonymity*** and have the right to withdraw from the research any time (***voluntary participation***). ***Vulnerable/sensitive population***, such as the homeless, the LGBTQ+ (lesbians, gays, bi-sexual, transgender, and queer or questioning) community, the prisoners, children, people with mental and physical illnesses, are especially protected. Strict and separate rules are often used on research targeting children as participants in the study. Video or audio recording of the interviews or interactions need consent from the participants. The data files and all relevant agreements related to the research need to be kept for an extended time period, and some may need to be encrypted and stored without identification information easily accessible.

When there is no risk or minimal risks associated with the research participants, *deception* may be allowed to conduct double blind or blind observations or experiments. *Double blind observations or experiments* mean that both the research participants and the facilitators are not aware of the true nature of the research, which was innocent deception to minimize the negative impact of reactivity on research results. Blind observations or experiments simply mean that only the research participants are not aware of the real intent of the research. If deception is used a *briefing* is often conducted after the collection of data.

For detailed guidelines about basic ethical principles, please refer to the Belmont Report published in 1976, authorized by the National Research Act of 1974 in the United States. The Belmont Report was written by the National Commission for the Protection of Human Subjects of Biomedical and Behavior Research.

References

Borrego, M., Douglas, E. P., & Amelink, C. T. (2009). Quantitative, qualitative, and mixed research methods in engineering education. *Journal of Engineering Education, 98*(1), 53–66.

Creswell, J. W., Plano Clark, V. L., & Garrett, A. L. (2008). Methodological issues in conducting mixed methods research design. In M. M. Bergman (Ed.), *Advances in mixed methods research: Theories and application* (pp. 66–83). Sage.

Li, Y., & Walter, R. (2013). Single-family housing market segmentation, post-foreclosure resale duration, and neighborhood attributes. *Housing Policy Debate, 23*(4), 643–665. https://doi.org/10.1080/10511482.2013.835331

Opoku, A., Ahmed, V., & Akotia, J. (2016). Choosing an appropriate research methodology and method. In V. Ahmed, A. Opoku, & Z. Aziz (Eds.), *Research methodology in the built environment: A selection of case studies*. Routledge.

Pickering, C., Johnson, M., & Byrne, J. (2021). Using systematic quantitative literature reviews for urban analysis. In S. Baum (Ed.). *Methods in Urban Analysis (Cities Research Series) (pp. 29–49)*. Singapore: Springer.

Chapter 4
Data Collection

The purpose of collecting data is to analyze them and answer the research questions. Therefore, data should reflect the key concepts in the research questions. For example, someone's research question is whether housing units within a certain distance of a light rail line would have a significantly higher appreciation rate than those that are not within the distance. The hypothesis (alternate hypothesis) is housing units within a certain distance of a light rail line would have a significantly higher appreciation rate than those that are not within the distance. The key concepts in this proposition are housing units within a certain distance of a light rail line, price changes of these housing units, and housing units beyond the distance. The first step to find data is to define the case study areas. You can choose multiple cities, regions, or only focus on one region. Geographic limitations should be mentioned in the paper. Rationale of choosing that specific region, for example, Chicago, may be because the region has a mature light rail system. Once the geographic area is chosen it is time to determine the study units. Study units are also called ***units of analysis***. Unit of analysis is what the data collection is based on and whether the units are individuals, groups, one region, aggregated regions, organizations, social interaction, or social artifacts. In this housing price appreciation example, the unit of analysis is each individual housing unit. Studying the attributes of these housing units is one type of ***cross-sectional study***. If we also explore how price changes over time it will be called a ***longitudinal study***. Based on the research questions and hypothesis this study will be both cross-sectional and longitudinal. The next definition to consider is within a certain distance. This needs some search for literature to check how other people have defined the distance. In planning literature, walking distance is usually defined as quarter mile (.25 mile) to half mile (.5 mile) but property value impact is usually defined as somewhere between half mile to one mile. If we are interested in the spline effect, we could use a sensitivity analysis to explore the impact within half mile and one mile distance. Once we choose a region, we need to

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_4].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_4

define the light rail line, and define the distance. We also need to decide whether to focus on single family or multi-family housing units. If we collect all residential housing data but curious about how different types of housing react to the impact, we can simply add a few dichotomy/dummy variables (with attributes of 0 and 1) to describe whether the housing unit is single family, duplex, or other types of housing.

Multiple factors influence price change of housing; the most predominant is location of the property and housing attributes such as size, number of bedrooms, and other features. Light rail may be one location factor. To control for other factors which might influence housing price change, certain data, such as the size of the housing, number of bedrooms, number of bathrooms, type of housing, distance to Central Business District, whether the unit is within certain distance of water, and whether the housing unit is in proximity of locally undesirable land uses (LULU), need to be collected as well. More factors to control for, more reliable the results will be. Reliability indicates the replicability of the research findings. Another key question to ask is how to determine the time frame of the housing price change. If the light rail line is new, it makes sense to collect the data after the establishment of the light rail. It might also make sense to use quasi-experimental methods to gather data prior to and after the construction of the light rail (pre and post-tests), comparing housing units within a certain distance (experimental group) and beyond that distance (control group).

Once the data needs are identified, the next step is to identify the data sources and the number of housing units to gather, which is called observations. Housing units beyond the defined study area, which might be five miles from a light rail, may not be necessary to be included in the research. Once all these are determined we find out that many county property appraisers and the GIS department have all relevant data. A few simple GIS procedures on ArcGIS will then create the data needed to complete your research.

Data and Variables

Data is the collection of variables and observations explaining the features of the study subjects. Data can be quantitative or qualitative and oftentimes a dataset includes both quantitative and qualitative data. *Datasets* are collections of statistical values or information for all the variables and observations/cases under investigation. ***Quantitative data*** are data in numerical format. ***Qualitative data*** are data in a form of texts, narratives, visual materials, etc., to describe the feature of a variable. The following is an example of a part of the dataset that you may need to complete the research project mentioned earlier.

In Table 4.1. "Folio" is the identification number of each housing unit. This variable, similar to many numeric identification numbers such as banking accounts, social security number, drivers' license number, is numeric but does not have mathematical meanings. "Distance1" is the actual distance as miles from the housing unit to the light rail line. "Distance2" is the converted distance indicated as a

Table 4.1 An example of a dataset

Folio	Distance1	Distance2	LotSize	BuildSize	Bedroom	Bathroom	Price1	Price2
0011021	0.6	1	4200	1400	3	2	150,000	160,000
0011022	1.2	0	3500	1425	3	2.5	152,320	161,700
0011023	2.3	0	4225	1500	3	2	160,000	173,000
…	…		…	…	…	…	…	…
1001101	0.5	1	4000	1400	3	2	153,000	160,000
1001102	2.7	0	5000	1650	3	2.5	171,000	182,000

dichotomy variable of 1 and 0, often called dummy variables, where 1 means the distance is within one mile and 0 indicates the distance is beyond one mile. "Lotsize" is the size of the lot and "BuildSize" is the size of the housing unit. Both are measured as square footage. "Bedroom" indicates the number of bedrooms and "Bathroom" indicates the number of bathrooms. Fraction of the bathrooms as .5 indicates half bathrooms where it may only have a lavatory. "Price1" is the baseline price and "Price2" is the new price. Both are expressed as dollars. The explanation of the variables and their attributes are often called a ***data dictionary***. When adding other pertinent data, such as the number of observations, sources of data, and relevant GIS parameters, we call the document a ***metadata***. The center of the inquiry now is to calculate a change in price (from Price1 to Price2), after controlling for consumer price index, and how this change is related to all the other housing and locational variables.

Once data are collected, regardless of primary or secondary data, it often requires data cleaning, processing, and manipulation. If the dataset is small, with a limited number of variables and observations, Microsoft Excel will be a great tool to conduct basic data merging, new variable creation, basic queries, elimination of unwanted data, and even chart creation for descriptive statistical analysis.

Box 4.1 Introduction to Microsoft Excel, SPSS, ArcGIS, and SAS

Depending on personal preferences and the capacity of a software there are various statistical and data processing software that researchers can use. The most used are Microsoft Excel, SPSS, SAS, and ArcGIS, among which Microsoft Excel and ArcGIS are the most used for planning practitioners. Microsoft Excel is part of the Microsoft Office bundle and can read data files in various formats. It is comprised of spreadsheets and workbooks and data can be referenced among different spreadsheets. Charts can be easily created and formatted in Excel. Excel also provides basic statistical analysis by enabling the data analysis add-in option. Advanced users are able to generate applications (apps) using Visual Basic programming and the forms function in Excel. The interface of Excel is self-explanatory and easy to use. However, Excel can only read and process data files with a limited number of variables (16,384) and observations (1,048,576) for its spreadsheets. Therefore, if one has particularly large files Excel will not be able to read and process the data properly. Another limitation of Excel is that its statistical functions are very limited. SPSS and SAS should be used when processing large datasets requiring complicated and advanced statistical analysis.

SPSS is menu-based, although prompts on commanding language will expedite the coding process. Most researchers opt to use the menus in data processing and analysis. SPSS has a standard window format as in Microsoft Office, with different menu names and keyboard short-cuts. SPSS has a data input window and an output window. The input window shows the data and variables. The output window shows the results of an analysis. SPSS is also able to read various data files. SPSS can also generate and format charts easily. The output tables and charts can be directly copied and pasted into Microsoft Word for further editing. SPSS is one of the most used statistical analysis software among researchers.

However, some others prefer using SAS, a statistical software requiring programming language, although the programming language is usually much simpler compared to Visual Basic, C++ and other types of languages. SAS is one of the most powerful statistical software and offers an array of functions not only for researchers but also for database developers. The SAS Institute offers various certifications using the software. Other commonly used statistical analysis software and programs are STATA and R.

Comparable to the comprehensiveness of SAS and other statistical applications, ArcGIS is the most used software bundle when handling databases with geographic references. The combination of Microsoft Excel and ArcGIS usually covers most of the day-to-day data processing and analysis tasks in a planning department. ArcGIS can execute spatial statistics such as spatial regression analysis and auto correlation analysis. Many urban planning programs in the U.S. offer specialized courses in GIS, some of which offer certificates and other degree programs on GIS.

Variables

Variables are the logical groupings of attributes. *Attributes* are characteristics of people or things. In a dataset they are the columns in a spreadsheet, describing the attributes of the observations. There are nine variables in Table 4.1. All the variables are quantitative variables since the attributes are described in numerical format. However, some of the quantitative variables, such as “Folio”, do not have numerical meanings and cannot be used as numbers in mathematical formulas. In ***level of measurement***, meaning different levels measuring the variables, we call variables such as “Folio” as nominal variables. “Distance2”, “Bedroom”, and “Bathroom” are interval variables. “LotSize”, “BuildSize”, “Price1”, “Price2” are ratio variables

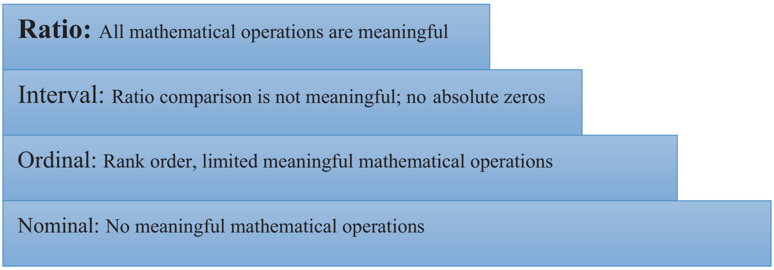


Fig. 4.1 Levels of measurement for variables and data

where numerical numbers have absolute zero and will be able to be used in any mathematical formulas. Figure 4.1 indicates that ratio variables are the ones with the most flexible mathematical manipulation, while nominal variables cannot be used in any mathematical operations unless being recoded to do so. Identifiers of individuals and places are nominal variables. Description of certain features of the study subject is nominal variables. Some nominal variables, such as parcel IDs, social security numbers, student ID numbers, membership ID numbers, are numerical; however, these numerical attributes cannot be used in mathematical operations. Ordinal variables have rank orders, such as the ranking of cities by population size, the ranking of scale from strongly disagree to strongly agree, and the ranking of cities by traffic fatalities. For ordinal variables the rank order matters but the difference between the rank orders is not meaningful. For example, New York City ranks No.1 in population size in the U.S. Chicago ranks No.3. The difference between No. 1 and No. 3 is 2 but does not have true measurement meanings. Ordinal variables can be used in simple statistical analysis, such as frequency analysis, or rank-order correlation analysis. Interval variables are numerical and can be used in more complex mathematical operations, but it does not have absolute zeros. The differences between values of interval variables are meaningful. For example, the difference between 90 Fahrenheit and 100 Fahrenheit has the same measurement of the difference between 70 Fahrenheit and 80 Fahrenheit. However, if the temperature doubles it does not necessarily mean that the hotness doubles; if the PH level of a liquid doubles it does not necessarily mean the liquid is twice less acidic. For ratio variables, for example, weight is a ratio variable because doubling the weight means twice as heavier. Therefore, the ratio comparison is not meaningful for interval variables but is meaningful for ratio variables. This is the fundamental difference between interval and ratio variables.

In certain statistical software, such as SPSS, level of measurement is only presented as “Scale” and “String”. “String” means nominal data, can be either text strings or numeric strings. “Scale” can be ordinal, interval, and ratio data, which have meaningful mathematical implications.

Observations

Observations are all the study subjects presented in a dataset. In Table 4.1 observations are each housing unit within the defined study area. The number of observations is the total number of housing units within the study area.

Census Data Versus Sample Data

The ***census data*** explores the attributes and characteristics of all the elements in a study population. On the other hand, the ***sample data*** only explores a subset of all the elements in the study population. Study population can be human beings, animals, places, or organizations. For example, there are 5000 urban planning agencies in the United States. If one is interested in knowing the staff characteristics of all the 5000 agencies, a census can be conducted where all the staff members in these 5000 agencies at a certain time point will be tallied for variables such as age, gender, educational attainment, professional certifications, ranking, salary, and other pertinent variables. It is tempting to conduct a census since we are often curious about all the elements in a study population. However, census is usually time consuming, expensive, and statistically unnecessary. Probability theories in statistics indicate that a statistically representative sample, a subset of the population elements, is often sufficient to be used to generalize about the population elements. The process of generalization is called inferential statistics, which uses hypothesis testing to test statistical significance of population parameters based on the sample data. For the planning staff example, if there are 50,000 staff members in these 5000 agencies, we can simply choose a sample size so that we can have 95% confidence that our sample statistics represent the population (all planning staff members in the 5000 agencies) statistics. Calculating the sample size has different methods and requires mastering the hypothesis testing concepts, which will be covered in later chapters. Many online portals offer calculation of sample size after inputting some simple parameters.

Once the ***sample size*** is determined we need to proportionally know the number of employees in each agency, by employee size, which will be in the sample. Later we will mention this is called quota sampling, which is not a probability sampling method, regardless that it strives to improve representativeness. We use quota sampling to determine the size of employees from each agency; then we randomly choose these employees from the agency. Only the characteristics of these randomly selected employees will be collected and aggregated to generalize about all the planning staff members in the 5000 agencies. In reality, we often use survey methods to collect the data, if secondary data is not available. We often send out the survey questionnaires to all or selected staff members, collect the responses, and then conduct a randomness test to verify if randomness is violated. Response rates of surveys depend, but vary from very low, for example, 5%, to very high, 90% or even 100%. The acceptable response rate is controversial. Some researchers indicate 50%, while others have lower or higher thresholds. If the collected data is randomly distributed, a lower response rate may be acceptable since it will not severely compromise the statistical analysis results. Survey techniques will be covered in a later section in this chapter.

Sampling error is the difference between sample statistics and population parameters. Population parameters are the same statistics inferred from the sample. For example, if a county has 250,000 households, among which 1000 were in the sample. The average household income for the 1000 sample households were $52,500 in 2016. Based on the census data we estimate that the average household income was $55,000 for all the households in the county in 2016. The difference of $2500 is therefore attributed to sampling error. The sampling techniques section will explain in detail that the sampling error is due to systematic and nonsystematic errors. Systematic errors are intrinsic errors and are difficult to reduce or eliminate; while nonsystematic errors may be possible to be reduced or eliminated.

Sampling Techniques

Collecting sampling data is science and needs careful planning. Fundamentally there are two approaches to sampling. One is the probability sampling technique, and another is the nonprobability sampling technique. If representativeness is important and if the sample data will be used to generalize the population characteristics, probability sampling techniques are strongly recommended. However, if the researcher only cares about exploring the characteristics of the sample and does not care about inferential statistics, nonprobability sampling techniques may be used to collect data. These two techniques are often used when gathering primary data; however, when obtaining secondary data sources, it is imperative to also check how the data were collected. The data collected should be reliable and valid. Reliability of the data means that the pattern of the data will hold even when other parameters have changed. Reliability often means replicability of the research and data collection in other circumstances. Validity indicates that the data collected are free of non-systematic errors, not fake, and valid.

Non-systematic errors are data collection and processing errors not related to the intrinsic nature of sampling and data collection. These errors, for example, data entry errors, can often be avoided if the data collector takes measures to reduce or avoid these types of errors. On the other hand, systematic errors are often intrinsic to the data collection process and cannot be avoided. For example, data precision errors cannot be fully avoided, particularly for ratio variables. Measuring distance, heights, or temperature, depending on the precision requirements, may yield different readings.

Probability Sampling

As it indicates, probability sampling is based on statistical probability. The foundation of probability is randomness, which means each element in the study population should have equal chance to be selected. If you are familiar with tossing a coin to have head or tail, or tossing a dice to have different values, you know that you have an equal chance to get a head or a tail and an equal chance to get each of the values on the dice. The more you toss the coin the more likely there will be a 50–50 chance of getting a head or tail. If there are six values on a die, the chance of getting each value will be 1/6. The Law of Large Numbers indicates that the more you toss the more likely the chances will become equal. Equal chance randomness is very important when conducting inferential statistical analysis based on hypothesis testing. There are four types of probability sampling methods: simple random sampling, systematic sampling, stratified sampling, and cluster sampling.

Simple random sampling is to randomly choose the study subjects without following any specific rules. For example, you uniformly fold the names of the students in a hat, toss the names, and randomly draw a few names out (sometimes even blindfold the eyes or do not look at the hat). Randomly shuffling the cards is also a way to randomize. These are the situations when the number of observations is small. When the study population has a large number of elements, computerized randomization can be used to choose the sample elements. Figure 4.2 shows an example of using a website to shuffle the numbers and randomly select 5 numbers out of 1-50.

Systematic sampling method is to choose the sample elements based on k^th^ intervals. For example, one needs to select several households from a few streets to conduct interviews and he decided to use the systematic sampling methods. In this case, he chose every 10th household based on the street number of the housing

Sets of numbers to generate:	2
Numbers per set:	5
Number range:	1-50
Set 1 of randomized unique numbers:	13, 39, 10, 48, 21
Set 2 of randomized unique numbers:	13, 8, 16, 4, 17

Fig. 4.2 Generating random numbers using a digital randomizer

addresses. For example, the following is a list of the street addresses which were systematically chosen by the researcher:

1001 3rd Ave.
1011 3rd Ave.
1021 3rd Ave.
1031 3rd Ave.
1041 3rd Ave.
…

The sampling interval is determined by the total sample frame (population size) divided by the sample size. For example, 500 households need to be selected from 2000 households. The sampling interval is thus 2000/500 = 4, which means the 4th household will be chosen to be in the sample.

Stratified sampling method usually deals with more complex situations where there are groups of elements which may share certain common features. These groups sometimes are also referred to as cohorts. There are between group differences, for example, the characteristics between mutually exclusive age groups (or age cohorts). In this case, age is the critical variable in the research question and the population is then divided to mutually exclusive age groups. Among these groups people are randomly selected to be in the sample. The sample size for each group is often determined by the proportion/quota of these groups in the population. For example, the age groups are classified as 0–4, 5–9, 10–14, 15–19, … 80–84, and 85 and over. If among the population 2% are between the age of 20–24, then population within this age range in the sample should be roughly around 2% as well.

Clustered sampling is like stratified sampling and is used in more complex situations. While stratified sampling technique stratifies all the groups in the study population and select among all the groups, clustered sampling only selects part of the groups randomly and then selects the elements in these chosen groups to be included in the sample. For example, there are 980 block groups in a county. A researcher randomly selected 200 block groups from these 980 block groups. Among the 200 block groups he then randomly selected some households to be included in the sample, based on certain stratification criteria such as age or gender.

An example of clustered (multi-stage) sampling can be explained in the following steps (Babbie, 2012):

Stage 1: Identify geographic units (e.g. census block groups) and select a sample. In Fig. 4.3 below, the selected block groups are shaded.
Stage 2: Go to each selected block groups and list all households in order. Figures 4.4 and 4.5 will use one block group as an example.
State 3: For each household list, select a sample of households (In the example, every fifth household is selected, starting from household #3 which is selected at random) (in Table 4.2, the highlighted households were selected from the example block group).

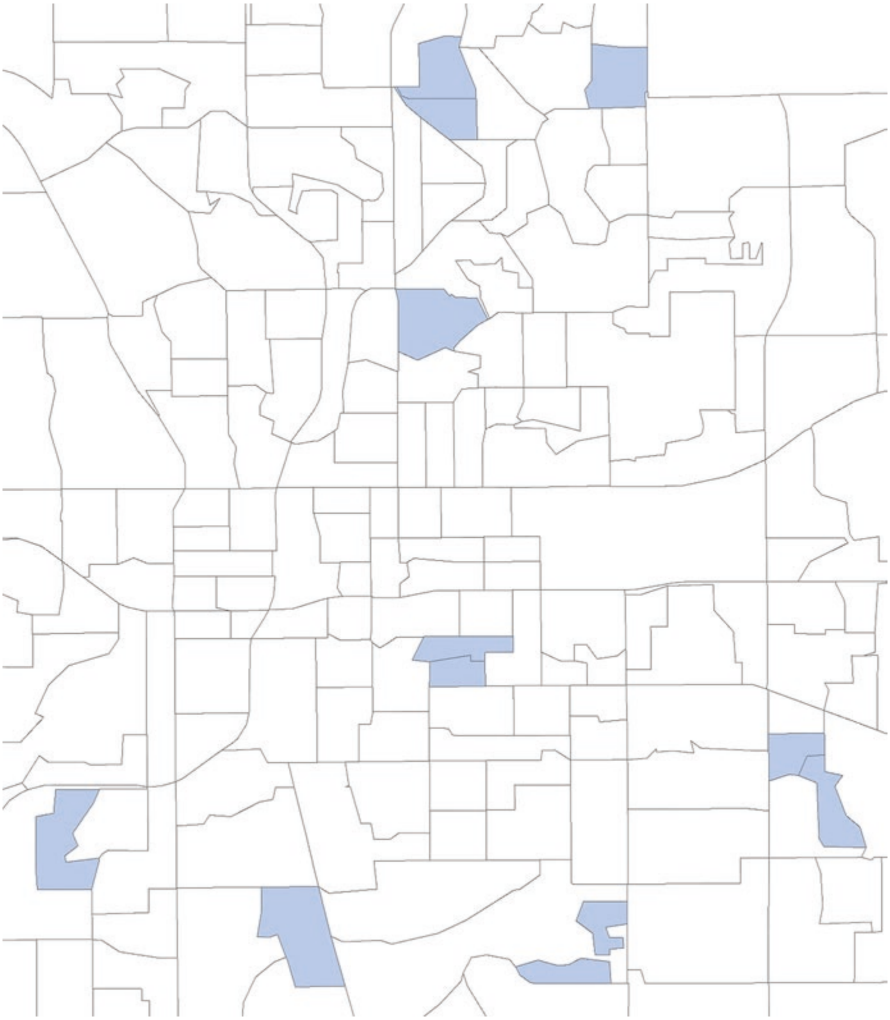


Fig. 4.3 Stage 1 of clustered multistage sampling

Non-probability Sampling

Non-probability sampling techniques are not specifically interested in sample representativeness or using the sample data to generalize about the population characteristics. Non-probability sampling techniques are often used in probing certain inquiries, such as people's behavior or attitude, in detail, and are often used on study subjects that are difficult to reach. Therefore, non-probability sampling is often used in qualitative research. The following are a few of the non-probability sampling techniques that researchers use.

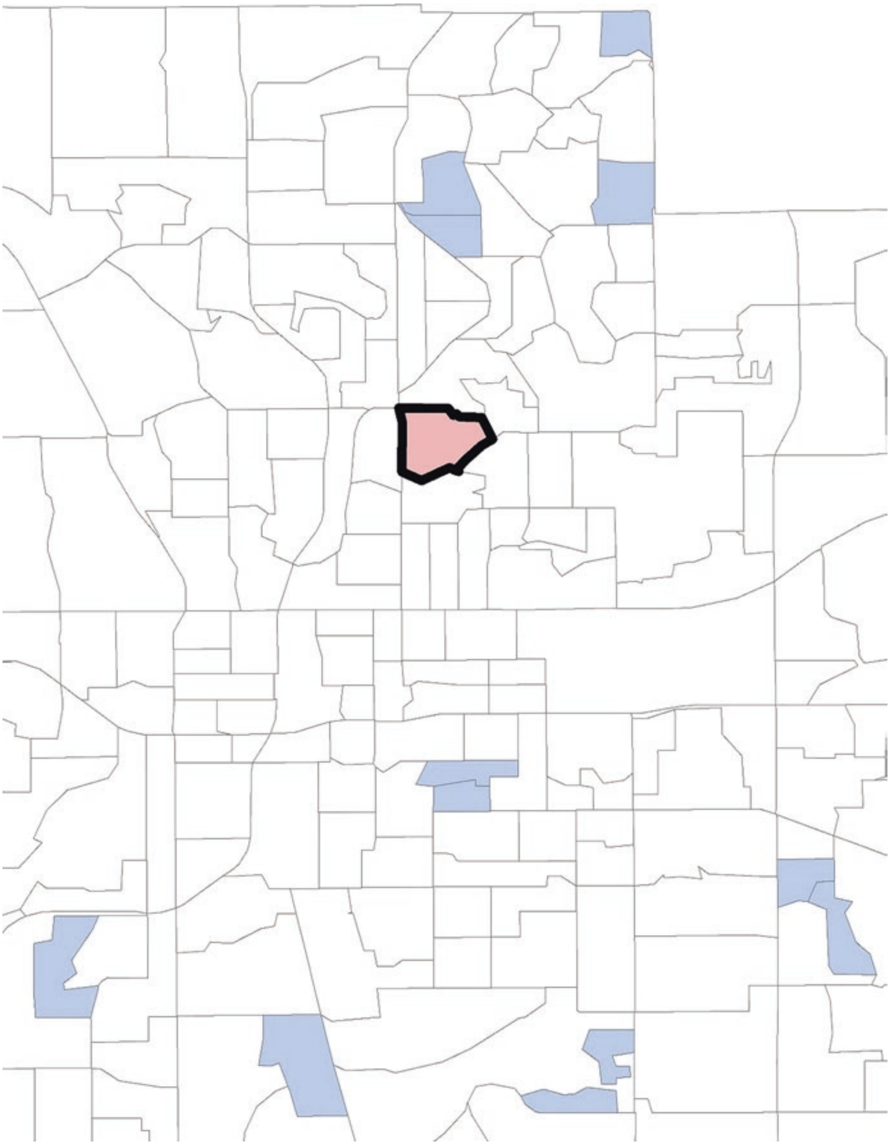


Fig. 4.4 The example block group (highlighted) for stages 2 and 3

Convenience sampling is to draw the sample based on whoever is available. Sampling a group of students in a classroom or sampling the passers-by on a street is a type of convenience sampling.

Snowball sampling technique is to start with a few research subjects to gather necessary data; then these subjects introduce more research subjects to the researcher. The sample size is therefore "rolled" larger and larger like a rolling snowball.

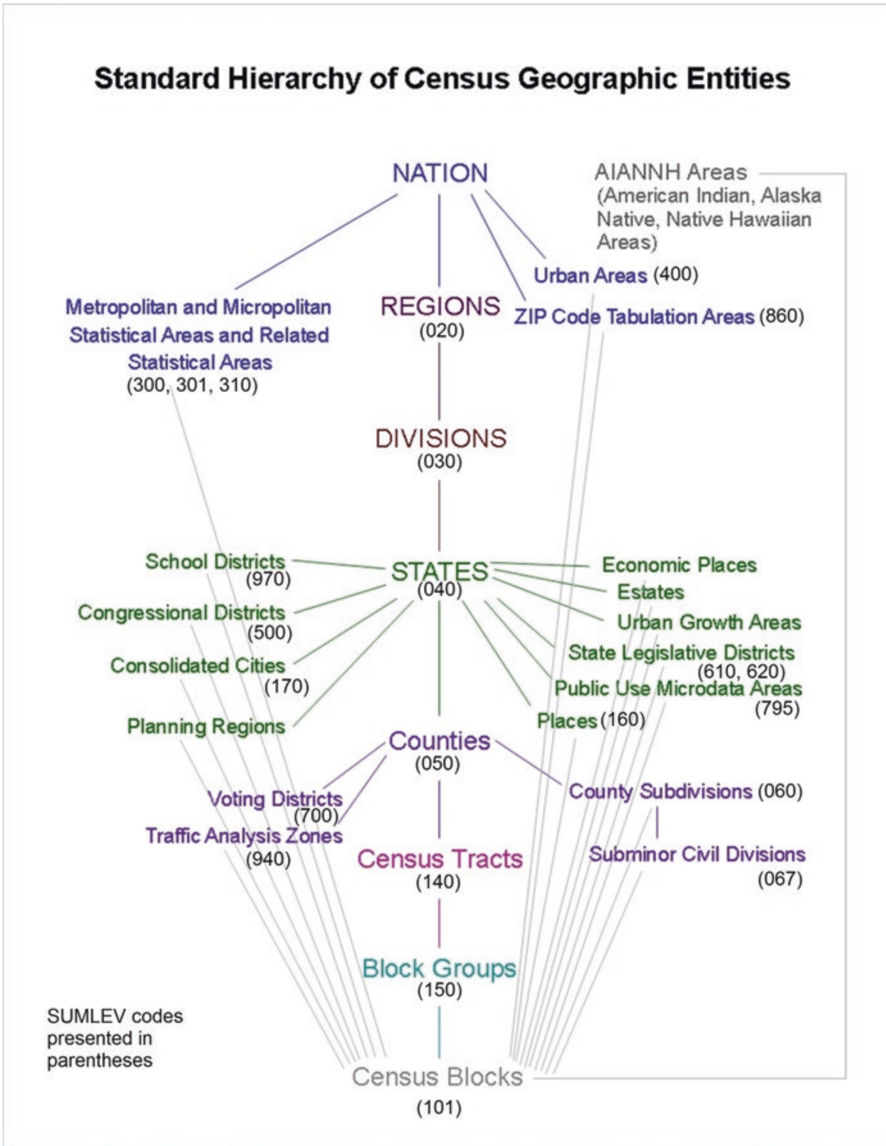


Fig. 4.5 Hierarchical census geographies. (Source: U.S. Census Bureau)

Purposive sampling is based on the judgment of the researcher per the research purposes. For example, the most vocal people at a public meeting were chosen as the research subjects.

Quota sampling technique is used when representativeness is somewhat important. In quota sampling the observations should reflect the structure of the population. For instance, if a city has 80% white, 15% black, and 5% other racial categories

Table 4.2 List of households for the example block group

Household	Address
1.	1782 KILLARNEY DR
2.	1400 LAKE SHORE DR
3.	1503 LAKE SHORE DR
4.	1228 MUNSTER ST
5.	1258 MUNSTER ST
6.	1205 DORCHESTER ST
7.	1245 MUNSTER ST
8.	16080 OLD ASH LOOP
9.	492 FLETCHER PL
10.	1200 MUNSTER ST
11.	1220 MUNSTER ST
12.	1234 MUNSTER ST
13.	1255 DORCHESTER ST
14.	1243 DORCHESTER ST
15.	1215 DORCHESTER ST
16.	1241 MUNSTER ST
17.	1225 MUNSTER ST
18.	1240 E LAKE COLONY DR
19.	2911 MIDDLESEX RD
20.	2901 MIDDLESEX RD
21.	1229 DORCHESTER ST
22.	1235 DORCHESTER ST
23.	2530 NORFOLK RD
24.	2518 NORFOLK RD
25.	2500 NORFOLK RD
26.	2420 NORFOLK RD
27.	2408 NORFOLK RD
28.	2404 NORFOLK RD
…	…
368.	1941 ENGLEWOOD RD

and if race is one of the most important factors determining your research questions, the sample should reflect the racial structure of the city. Therefore, in the sample white population should be around 80%, black 15%, and others 5%.

Key informants sampling technique is to only focus on a few key people or insiders who know much about the research questions or the study area. For example, a planner is curious to know the community's reaction of a planning project. She decided to talk to the community civic leaders, Homeowners' Association (HOA) board members, business leaders, and other key people representing the community. The technique that she uses is called key informants sampling technique. This technique is useful when gathering data from a large amount of people is not possible, too time consuming, or too expensive. When choosing the key informants, it is imperative to consider the groups they represent and make sure there are no groups left behind.

Data Sources

Data are gathered by various agencies and by the researchers based on the sampling techniques mentioned above. If somebody gathers data themselves and use the data in research, these data are primary data for the individual or organization; however, if these data are then used by third party individuals or organizations the data are secondary data for these third-party entities. When conducting research, finding free, low cost, and reliable secondary data should be the first attempt in data collection, since gathering large-scale primary data takes longer time, consumes more financial resources, and needs more personnel support (Baum, 2021). Depending on the scope of the research project, sometimes gathering primary data is the best approach to help answer the research questions. For example, for a planner working for a municipal government, secondary data on residents' needs and perceptions are usually lacking. Under such circumstances, gathering primary data using methods such as public forums, workshops, focus groups, internet surveys, and interviews are mostly doable.

With the advancement of technology, finding free secondary data is much easier than before. A large amount of secondary data for each of the planning subdisciplines, such as housing, community development, transportation, environmental planning, etc., are freely available across the globe, particularly from prominent international organizations and developed countries. For example, the U.S. Census data (https://www.census.gov/data.html), the United Kingdom Household Longitudinal Survey (https://www.understandingsociety.ac.uk/), the United Nations databases (https://data.un.org/), the Human Development Data Center from the United Nations Development Programme (http://hdr.undp.org/en/data), and many local open GIS datasets (e.g. Palm Beach County, Florida Information System Services at https://discover.pbcgov.org/iss/cwgis/pages/default.aspx) can be easily accessed across the world when there are no network or political restrictions. However, sensitive government, demographic, or business data linked to national security are usually classified and not readily available to the public. Much of the secondary data allow not only cross-sectional studies but also longitudinal studies. For example, the census, households, and housing data maintained by many national governments include geographical identifiers and often gathered each year or every few years. American Housing Survey is a great example of **panel data**, which include the same survey subjects longitudinally by surveying the same households every 2 years. Periodically new households will be added to the panel. Longitudinal panel data such as these provide a wealth of information to look for trend and patterns for the same geographical areas and the same households, or longitudinal comparison between different geographical areas and households.

Some secondary data can be costly without necessary funding support. Some of the expensive data sources include data associated with ESRI ArcGIS such as Business Analyst since subscription of ArcGIS is also required to use Business Analyst. Certain real estate data, such as mortgage delinquency data managed by CoreLogic Inc., are also costly to acquire.

While using secondary data are proven beneficial and cost effective, these data do have inherited disadvantages when looking for individual characteristics or more detailed information. Sometimes missing data and inconsistent geographical boundaries are challenging for researchers. These secondary data are often not tailored to the research needs of each individual research question and using them would yield invalid or unreliable findings. To overcome the limitations of secondary data sources, primary data collection, which relies on field observations, focus groups, surveys, interviews, and other collection methods, is imperative.

This chapter will mostly focus on free or low-cost secondary datasets that can be easily downloaded from the internet. After exploring ***secondary data sources, primary data sources*** will be introduced in case the researchers need to gather their own data.

Secondary Data Sources

As mentioned previously, there are abundant secondary data available, mostly on the internet in the form of data repository or clearing houses, ranging from the census demographic data, household data, housing data, transportation planning packages, geological survey, remote sensing and satellite imagery, property taxation and attribute data, mobility data, meteorology and climatology data, and other types of data. Many of the databases are geo-databases, including geocoded features to make data retrieval and comparison possible.

The most prominent and popular secondary data in the United States is the ***U.S. census data*** from the Census Bureau. The Census Bureau not only maintains and produces the decennial census, but also American Community Survey (ACS), Census Transportation Planning Package (CTPP), County Business Patterns (CBP), Economic Census, street and boundary GIS data, and other data sets. The decennial census and American Community Survey are used the most among the urban planning profession (Williamson, 2008).

The decennial census started in 1870 in the U.S. and includes both the long-form and short-form versions until 2000. Starting 2010 only short-form census was collected. If one needs to probe more detailed information yearly American Community Survey will be the source to rely on. Short-form census only includes seven key questions about the basic characteristics of each household. The seven key questions are the following:

- Name
- Sex
- Age
- Relationship
- Hispanic Origin
- Race
- Owner/Renter Status

Long-form census includes more questions, such as detailed race and ethnicity categories, marital status, citizenship status, educational and income information, housing attributes, mortgage information, economic status and occupation, commuting information, and many other variables.

To understand the census data, one needs to understand census geographies. Figures 4.5, 4.6, 4.7, and 4.8 illustrate the census geographies. The census geography starts from a simple block, which is bounded by four streets or other boundaries, to a few blocks (block groups). A few block groups then form a census tract. Multiple census tracts then become a politically formed county or parish. Counties/parishes become states and all the 50 states and territories then become the U.S. This linear line of geography is inclusive of geographic units at a lower level. The non-inclusive line includes other geographies such as the school districts, places (cities/villages/towns), census designated places, congressional districts, and others. These geographies are usually bounded by two major boundaries, administrative/political boundary and census statistical boundary. Counties/parishes and incorporated cities/villages/towns are administrative boundaries while census designated places, Metropolitan Statistical Areas, and others are for statistical purposes only.

Note: SUMLEV is used to aggregate data by different geographical levels.

Census blocks are the smallest geographic unit in census and are often bounded by visible or nonvisible boundaries such as streets, railroads, rivers, property lines, county lines, city boundaries, etc.

Block groups usually have a population between 600 and 1300 and are groups of blocks.

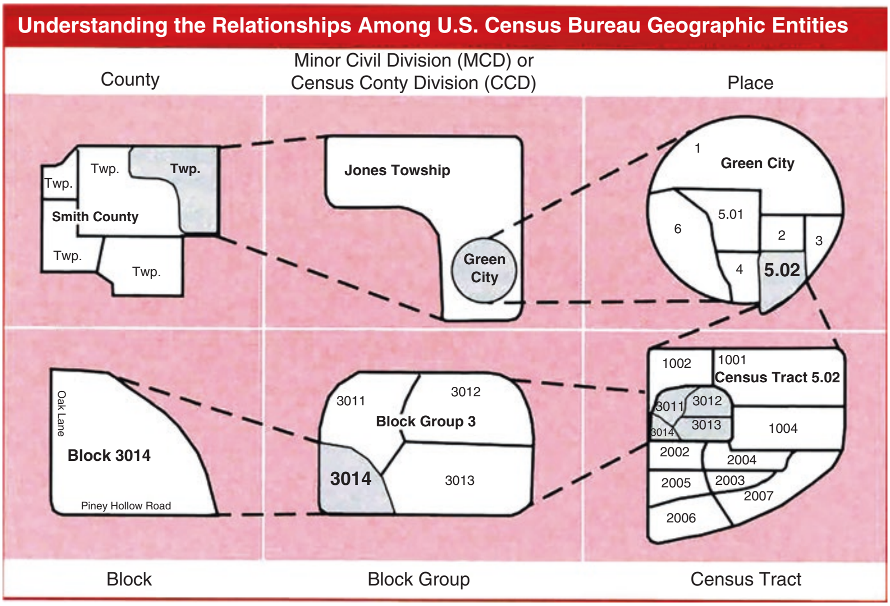


Fig. 4.6 Small area census geographies. (Source: U.S. Census Bureau)

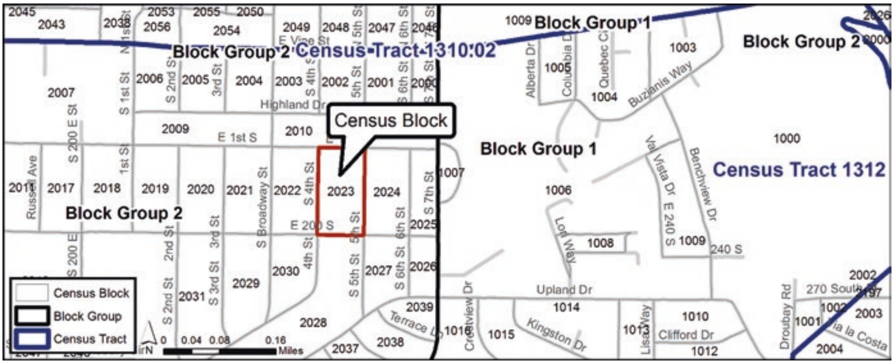


Fig. 4.7 Census blocks. (Source: U.S. Census Bureau)

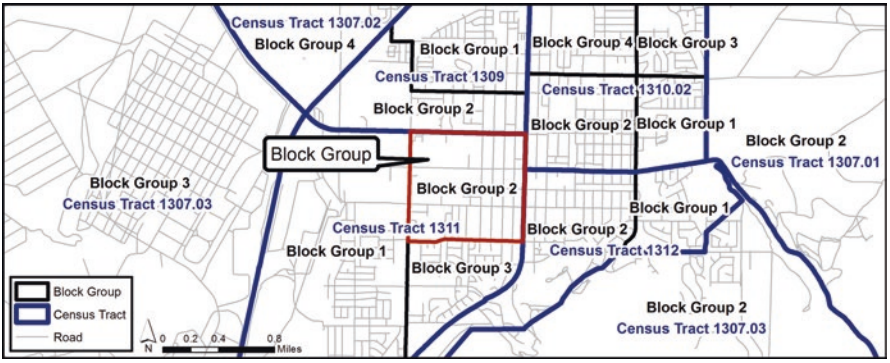


Fig. 4.8 Census block groups. (Source: U.S. Census Bureau)

There are multiple ways to retrieve the census data (The U.S. Census Bureau, n.d.). The simplest is to use its data tool on the official website (http://www.census.gov), such as American FactFinder, to conduct a few queries, and then download the data in Excel or text format. The most used text format is called comma delineated format with the file type of .csv. Similar to the Excel files comma delineated files can be opened in any statistical analysis software. A detailed description of retrieving the census data will be in Chap. 9 Demographic Data Analysis.

Other commonly used secondary data include data from federal agencies, such as the U.S. Department of Housing and Urban Development (HUD), U.S. Environmental Protection Agency (EPA), U.S. Department of Agriculture (USDA), U.S. Geological Survey (USGS), Bureau of Economic Analysis (BEA), Bureau of Labor Statistics (BLS), etc. Based on interests researchers should explore the websites associated with these agencies to find out which data are available from them. At the state and local levels there are also multiple arrays of data available. States usually work with nonprofit or higher education institutions to maintain their data centers. At the local level, the most widely available data are the GIS (Geographic Information System) data and the county property appraiser data. The

GIS and property appraiser data are mostly public data, although in some states they are not publicly available. Regarding the GIS data the Census Bureau provides boundary data at each level of the census geographies. The Census Bureau also provides other GIS boundary data, such as roads, waters, address information and demographic data associated with the census boundaries. These GIS data provided by the Census Bureau are referred to as TIGER (**T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing) data.

In planning analysis, sometimes the secondary data are qualitative (Denzin & Lincoln, 1994; Gaber & Gaber, 2007). For example, planners and researchers have analyzed the quality of various plans, such as comprehensive plans and hazard mitigation plans. The data used in these plans are often reports and documents. Such type of analysis is called content analysis and is part of the unobtrusive measures used in gathering primary or secondary data. Details of content analysis will be discussed in Chap. 6 Qualitative Data Analysis.

Some other secondary data sources are commercialized and often require paid purchase or subscription. Mortgage delinquency data, real estate multiple listing data, and certain business and market data are among those paid services. CoreLogic, CoStar, ESRI, and ESRI Business Analyst are among those commercial data vendors in planning and related disciplines. Some commercial data carriers, for example, PCensus and Geolytics, also create easier platforms to retrieve and map the Census demographic data. Specialized software, such as IMPLAN, deals with analyzing economic impact of planning projects and provides detailed input-output data of economic activities.

Primary Data Sources

In situations where a researcher needs to collect data first-hand, there are multiple methods to achieve the goal. The methods used most often in urban planning are surveys and interviews, field observations, experimental design, and public forums and focus groups. Collecting primary data needs deliberate planning and should consider the timing, costs, sample size, reliability, and validity of the data collected. If statistical inferential analysis is the goal of the research a sample needs to be random and has the necessary size for making such analyses. Primary data collection should be only used when there are no reliable secondary data readily available. However, primary data oftentimes reveals much deeper relationships among variables and the collection process often helps the researchers to deeply connect to the research project. This section will focus on the commonly used data collection methods in planning and explains how to use these methods efficiently and effectively to gather intended primary data.

Surveys

Surveys, and sometimes interviews, are the most used data collection methods in social science research. Many of the secondary datasets that planners use frequently, for example, the decennial census, American Community Survey, and American Housing Survey, are based on the survey data. Survey administrators send questionnaires directly to the research subject. The research subject answers the questionnaire, and the researchers use the data to conduct analyses answering various research questions. Large datasets are usually able to help researchers derive many more research inquiries and questions than smaller and more purposeful datasets. Based on the space and time horizons of the data collection process, surveys can be classified into cross-sectional and longitudinal. Cross-sectional surveys study the survey subjects at one time spot and longitudinal data follow up the subjects for multiple time periods or a long time. Before conducting surveys to gather data research questions need to be defined and operationalized so that the key concepts are appropriately mapped. The key concept map will then identify the variables needed to answer the intended research questions. Once the variables and sample size are determined the researcher needs to design the survey questions.

There are two types of survey questions, ***open-ended questions*** and ***closed-ended questions***. Open-ended questions are questions where the choice of answers is not given. For example, the survey subject was given three site design maps of a subdivision and then the following question was asked:

Which of the subdivisions do you like to live in? Please explain why.

Closed-ended questions offer answers so that the survey subject can choose from the given answers. Closed-ended questions need to provide answers that are mutually exclusive. Therefore, if a researcher is not able to list all the possible choices of answers a category of “other” needs to be added to the choices. Some examples of closed-ended questions are as follows:

What is your racial background?

☐ *White*
☐ *Black or African American*
☐ *Asian*
☐ *Pacific Islander*
☐ *Alaska or Native American*
☐ *Mixed Race*
☐ *Decline to answerWhat is the primary transportation mode for you to commute to work?*

☐ *Driving alone*
☐ *Taking a train or light rail*
☐ *Taking a bus*
☐ *Riding a bike*
☐ *Walking*
☐ *Car sharing*
☐ *Other, please specify:*

Box 4.2. Racial and Ethnic Background

Many confuse racial and ethnic backgrounds when designing the survey or analyzing the data. Racial background reveals the racial and heritage background of the survey subject, which is often manifested by physical features such as skin color, while ethnic background reveals the origin of the survey subject. The official racial categories defined by the Census Bureau are reflected in the above example of survey questions. Ethnicity is often expressed as, for example, Hispanic/non-Hispanic, Japanese, Polish, Mexican, etc. Distinguishing between racial background and ethnicity is critical since the two categories overlap, for example, white Hispanic.

There is a type of survey questions which are called ***contingency questions***. Contingency questions are a set of questions that later questions are contingent upon the answers from previous questions. For example,

Do you believe that human activity is one of the major factors contributing to global climate change?

☐ Yes
☐ *No*
☐ *Not sureIf yes, explain what you think is the most effective public policy to mitigate the negative impact?*

In designing survey questions there are a few scales that a research needs to be aware of. ***Likert Scale*** is the most often used scale. Answers in Likert scale are ranked in ordinal sequences so that each scale is either stronger/larger or weaker/smaller than the next scale. The scale of agreement, from strongly disagree to strongly agree, is a type of Likert Scale. There are other types of scales, such as Semantic differential scale, Guttman scale, Thurstone scale, and social distance scale, but they are either less commonly used or used but not theorized in planning and related research. For further references, Babbie (2012) and other books on research methodology has detailed description of these different scales.

Some other types of questions include the matrix questions, which are collections of closed-ended questions with a similar structure of choices. Matrix questions are used to better structure the survey questionnaire. For example, the following questions are matrix questions:

Regarding the neighborhood that you live in please check all that apply.

	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	*Not Applicable*
1. *Public transportation is easily accessible*	☐	☐	☐	☐	☐	☐
2. *Public roads are well maintained*	☐	☐	☐	☐	☐	☐

	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	*Not Applicable*
3. *Recreational facilities are easily accessible*	☐	☐	☐	☐	☐	☐
4. *Grocery stores are easily accessible*	☐	☐	☐	☐	☐	☐

When asking survey questions, it is imperative to choose appropriate question forms. The wording of the questions and choices should be short, clear, and concise. Avoid double barreled questions, such as "Do you prefer having a park or a school on this site?", or "Do you drive and ride a bike to work?" Another point is that the respondents should be competent to answer the questions. Asking a child about a mortgage associated with her house may not yield reliable answers since the child may not have the necessary knowledge about the mortgage. Meanwhile, qualified respondents should be willing to answer the questionnaires. To get reliable and valid answers the questions should be relevant to the respondents. Avoid negative questions such as "Don't you like building a light rail in your city?" As in other field research, avoid biased terms and avoid offensive wording and languages.

When structuring the questionnaires, it is best following the rules below:

- Questionnaires should be spread out appropriately and uncluttered.
- One question should be about one line.
- Use contingency questions when necessary.
- Format matrix questions to add readability of the questionnaire.
- Place interesting items in the beginning (opposite to interviews).
- Place non-threatening questions first while sensitive questions, e.g. personal information and sensitive personal attitude/behavioral questions, near the end of the survey.
- Include instructions to the questions.
- Include a disclaimer at the beginning of the questionnaire.
- Offer to share the survey results with the respondents.
- Pretest all or part of the questionnaire.

When conducting surveys there are different logistics involved. The first step is to determine the study area, the study population, and the sampling methods explained in previous sections of the chapter. Once the target audience and sampling methods are determined, distributing the survey is a critical step to ensure success. There are mail surveys, in-person surveys, and online surveys. Online surveys are the most often used nowadays, due to advances in technology and internet. Designing online surveys has become easier than ever. There are various surveying software and applications and some of the most used survey instruments are Survey Monkey, Google Forms, and Qualtrics, among others. Survey Monkey is easy to use, and it is free if the number of questions is limited to 10 or under. However, for most researchers conducting empirical research, a paid version will be very beneficial. Google Docs is free, but the design platform may not be as easy and convenient as other survey software or applications.

Once the questionnaire is designed and completed it is emailed to the respondents. All the survey software or applications can store the answers in an Excel spreadsheet from the respondents, which made data processing much easier than the traditional pencil and paper surveys. Qualtrics is a paid program and the collected data can be exported as an Excel file or a SPSS file. These software applications can also identify who have responded and who have not, and Qualtrics saves the IP addresses of the survey respondents. Follow-up emails can be sent to those who have not responded. However, these survey instruments do require acquisition of email addresses of the survey subjects. Another drawback of online surveys is that many recipients of the survey emails simply delete these emails, or the emails will directly go to the spam folder and thus may be deleted by accident. Some researchers use social media and apps, such as Facebook, Twitter, LinkedIn, and Instagram to distribute the survey questionnaire; however, privacy concerns, selection bias, and non-randomness are some of the major hurdles using social media in gathering data (Andrade, 2020). Data gathered from social media, therefore, often mislead, and cannot be generalized. Online questionnaires should avoid excessive scrolling and limit the time to answer the questions within about 15 minutes.

As with other primary data, collected survey data need to be cleaned first to eliminate or recode responses that are not needed, incomplete, or invalid. Then coding of the variables to make them easier to analyze should be conducted. For example, one can create dummy variables for categories with dichotomy, and assign missing data or data that are not applicable to missing values. Data analysis follows cleaning and coding the data. Detailed analytic methods are determined by the research questions and are covered in different chapters of this book.

For any primary data collection involving human beings or animals, if the researcher intends to formally publish the results, an Internal Review Board (IRB) needs to be informed if there is one in the researcher's organization. If there is no IRB board the researcher still needs to follow certain ethical standards when conducting the research. Research ethics was discussed in detail at the end of Chap. 3.

Box 4.3. Using Google Docs and Survey Monkey

Software and online applications change drastically over time; however, as of now Google Docs and Survey Monkey are among the most popular tools to be used in designing survey questionnaires for free (If fewer than 10 questions in Survey Monkey).

For Google Docs, one can log on to https://docs.google.com using either an institutional account (if the institution subscribes to Google products) or a Gmail account. Once logging on, click the three horizontal bars at the upper left corner of the page, choose "Forms" (there are four applications, "Docs", "Sheets", "Slides", and "Forms"). Then click "Start a New Form" either using the "blank" document or a template. Then once the survey interface appears format the title, the form description, color pallet and theme, background, and add different types of questions. The interface is self-explanatory and very simple to use. Feel free to click all the buttons and drop boxes to find out what each entails. After the questionnaire is created, preview it (the eye button at the top right corner of the window), share the survey questionnaires to the survey respondents through emails or links. Responses will be automatically saved as a Microsoft Excel file.

For Survey Monkey, create a username at https://www.surveymonkey.com and click "Create Survey". If you have used the website before to create survey questionnaires the website will ask if you desire to copy an existing survey or start from scratch. After you click "Start from scratch", name the survey and choose a category that the survey belongs to. Click "add new questions", then start creating different types of questions using dropdown boxes and the dialog windows. Change the themes and formats using the tools on the left. After creating all the questions click "save" and then email the survey to intended recipients. As in Google Docs the responses will be automatically saved as Microsoft Excel files. Additionally, Survey Monkey offers basic statistical analysis of these responses. The free version of Survey Monkey is limited to 10 survey questions and 40 responses. Paid subscription is required if going beyond these limits.

Field Observations

There are different approaches to field observations, site reconnaissance, windshield surveys, participant observation, physical accretion measures, or field interviews. As mentioned previously the researcher needs to strictly abide by ethical standards when conducting the research if the research subjects of the field observations are human beings or animals.

Site reconnaissance is similar to the due diligence site visit in real estate development. In this situation the researcher goes to the study site, observe, take notes, take pictures, or record any relevant data about the site. There are some modern devices helping with data collection on site. For example, GPS tracker, drones, or certain apps can be utilized to aid in data collection. Due to advances in technology, much of the data that were not easily collected before have become much easier to collect. For example, aerial footage of the study area can be either collected through a drone equipped with cameras or retrieved from Google Maps or Google Earth.

Windshield survey is usually done through driving through the site or area and recording relevant data. Windshield survey is often used when the study area is large or the study time (e.g. night time) is not permitting walking in the area due to concerns of safety and other issues. It is ideal to rely on devices, such as cameras (for example, smartphones or Go Pro), to record the data if driving safety is a concern. A passenger who will be focusing on data collection will be helpful so that the driver can focus on driving.

Participant observation is used when researchers observe the behavior of the study subjects. Due to the reactivity issue, which is defined as people will likely behave differently if they are aware they are being observed. Reactivity will negatively affect the reliability and validity of the research. In this situation, justified deception is ethical, which means the study subjects will not be informed of the research. The researcher can choose to fully participate in the activities, pretending being part of the group, or can observe behind the scenes, as a spectator, or through one-way see-through windows. A briefing will be provided to the study subjects of the research. Justified deception is ethical only if there is no risk or minimal risks associated with the study subjects. Observing children's behavior, whether in a classroom, a playground, a daycare center, or other places, needs consent from the children's guardians.

Physical accretion measures indicate regardless of which field observation methods are used the researchers collect data during different days or time periods, or during different situations.

The last field observation method, ***field interview***, is a method used rather often. As any other methods in collecting data, it requires careful preparation but follows somewhat different approaches and ethical principles. Interviews can be structured or semi-structured. Field interviewees are human beings, and the interview will be either face to face or through phones or internet. Therefore, interviews are direct interactions. Successful interviews thus require rapport and certain inter-personal skills. Interview questions are not set as strict as survey questionnaires and can change to respond to the specific probing during the interview. During an interview, the respondent should be doing most of the talking, not the interviewers. Phrasing interview questions is very important for the interviewer to get intended answers. Wording is especially important and offensive or sensitive questions should be avoided. For example, for some reason, the term "affordable housing" sounds offensive to certain local officials since they relate it to low-income housing. Therefore "workforce housing" was used in situations like that. "Welfare" is offensive to whomever receiving governmental subsidies. Therefore, "Subsidies" or whatever specific assistance programs should be used to avoid offending the interviewee. Therefore, the interviewer should be sensitive to the cultural, racial, and ethnic background of the interviewee before conducting the interview.

There are seven stages relating to interviewing (Babbie, 2012):

1. Thematizing. During this stage, concepts and theories are explored and theorized.
2. Design. Lay out the questions to ask, when to schedule the interview, what devices will be used and whether informed consent is needed.
3. Interviewing. This is the stage where interviewing takes place.
4. Transcribing. The interview content is transcribed and written as a transcription.
5. Analyzing.
6. Verifying and checking facts for reliability and validity.
7. Reporting.

Experimental Design

Experimental design is a commonly used method in natural and medical science. However, experiments or quasi-experiments can be used in social science and behavioral research. Experiments indicate the data collected are not from a natural setting; instead, they are from a constructed setting similar to an experiment. Strictly speaking, experimental design needs to satisfy an array of conditions, such as random selection of the experiment participants, requiring pre- and post- tests, independent variables, dependent variables, control group, experimental group, and stimulus. On the other hand, a quasi-experiment does not need to have all the required elements.

In an experiment, there are usually two groups of research subjects. One group is the experimental group, and another group is the control group. The participants in both groups need to be chosen or matched randomly. The ***experimental group*** receives a stimulus or the testing instruments, and ***control group*** receives a placebo or nothing. Then the outcome of the experimental group and control group, for example, effectiveness of a new medicine, measured based on the research questions is called ***dependent variable***. The presence or absence of the stimulus/testing instruments is called the ***independent variable***. The experimental and control groups are followed by the researcher longitudinally and the outcomes are recorded prior to (***pre-test***) and after the administration of the stimulus (***post-test***). Figure 4.9 is a simple illustration of a strict experimental design process.

Strict experimental design may be imperative in natural and medical sciences; however, in social and behavior sciences strict experimental design may not be necessary. Therefore, there are variations of experimental design and quasi-experimental design where one or a few elements of classical experimental design are missing.

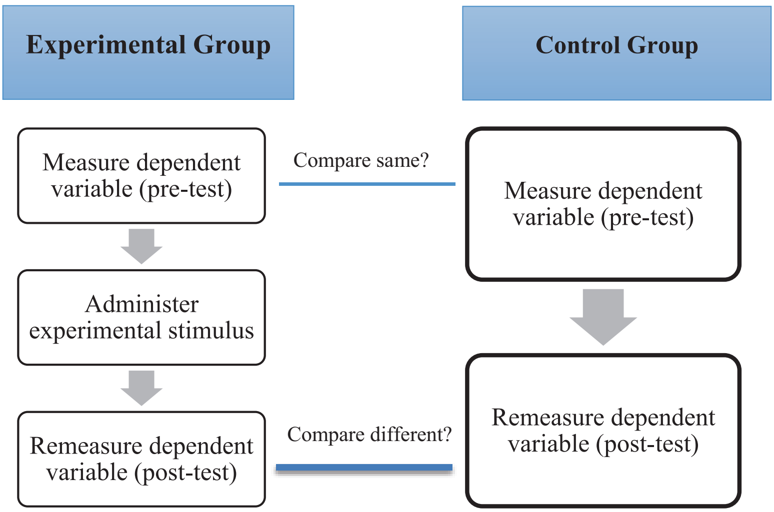


Fig. 4.9 Experimental design research process

For example, a transportation authority decided to establish express lanes on an interstate highway and charges tolls on vehicles using the express lanes. The Authority followed the revenue and expenses of the toll lanes for 2 years and then made the conclusion that the toll lane can effectively reduce rush hour traffic on regular lanes while maintaining a significant amount of profit margin. In this type of research, the Authority only follows the experimental group for a period of time after the establishment of the express lane. There are no pre-test and no control groups. Some other examples of variations of experimental design include monitoring the neighborhood impact of a Low Income Housing Tax Credit (LIHTC) project prior to and after the project is in place, where the control group is missing. Case studies of downtown river front development are an example of comparing multiple experimental groups to draw patterns of successes and failures in such development. Therefore, even though the strict classical experimental design will help to promote the reliability and validity of a study, variations of experimental design are generally acceptable in social science research.

Public Forums and Focus Groups

Public forums and focus groups are commonly used methods to solicit public input regarding a planning policy or project (Greenbaum, 1993). Therefore, public participation is an important method in collecting primary data in planning. There are different forms of public forums and focus groups. Some of them are more organized; for example, the participants take turns to express views and alternatives and then priorities are chosen among these alternatives. The *Delphi method* of soliciting opinions from the experts in the field is one of such methods. *Normative participation*, where each participant has the opportunity to express his/her concerns or opinions, will offer a more comprehensive approach to public participation since it is not just the most vocal ones speaking up. Brain storming, fishbowl participation, etc., are all variations of the normative participation method. Fishbowl participation means people take turns to go into the “fishbowl” in the center of the discussion room and actively participate in the forum or workshop.

Some other public forum and brainstorming sessions are less organized. Data collected from a public forum may be biased since the participation might be limited to certain residents in the jurisdiction. However, *focus groups* and other deliberately planned workshops, such as charettes, may reach more representative residents and experts. The number of participants in these workshops and focus groups should be limited to, for example, 20–30 people, to ease the facilitation process. A design *charette* is a commonly used technique to gather input from the participants. In a charette, the participants actively participate in the design and planning of a project by drawing or brainstorming different design concepts and scenarios and then prioritize the ones that are most accepted. To collect the most useful data effectively, public forums, workshops and focus groups need to be deliberately and carefully planned to choose the right combination of participants. During the COVID-19 pandemic, many of the public meetings, focus groups, and design charrettes have moved online, which requires the organizers to be mindful of inequality in internet access.

Data Collection Through Social Media, Remote Sensing, Real-Time Tracking, and Other Technologies

With the advancement of technologies and artificial intelligence, the collection of more complex big data and digital census is becoming a reality. Digital census has two types of meanings. The first indicates that the traditional census data collection has become digitalized, greatly reducing the time and manpower to digitize paper questionnaire responses. Digital census data can also mean real-time or digital data collected to reflect demographic characteristics, household characteristics, and other psychographic data such as consumption patterns, opinions, and perceptions (Zhu & Sobolevsky, 2018). Compared to the traditional census data, real-time digital census is more complex and dynamic. Some examples of the digital census data include the 311 calls of local municipalities, mobility tracking and contact tracing by mobile devices used during the COVID-19 pandemic, and other locally collected data such as the NYC (New York City) TLC (Taxi and Limousine Commission) Trip Record data. These data provide valuable big datasets for policy and research purposes (Desouza & Smith, 2016). However, privacy issues may arise for data collected through these means (Smith et al., 2012).

Social media, such as Facebook, Nextdoor, Reddit, LinkedIn, Instagram, and Twitter, also provides new platforms in data collection, integrating data collection applications used by third-party surveying vendors such as Qualtrics and Survey Monkey. Advantages of using social media as the platform include broad outreach to the targeted audience, convenience, and low costs; however, some prominent limitations include selection biases and sometimes it is difficult to reach the target audiences due to privacy restrictions and people's unwillingness to participate in the data collection process. The participants of data collection through social media may not reflect the real demographic composition of the study area since many residents may not be active on these social media platforms.

Remote sensing and internet of things (IoT) are other major sources of big data. Traditionally, remote sensing has been used in geography, geology, climatology, meteorology and other disciplines and industries. Currently, remote sensing and real-time monitoring are also the critical technologies in smart city implementation, such as in smart transportation, autonomous vehicles, smart homes, mobile apps, and wearable smart devices. The speed of evolution in data collection methods is astounding, going through from in-person interviews/surveys, telephone surveys, emailed internet surveys, to social media data collection, mobile device data collection, remote sensing collection, and other collection methods using new computer and internet technologies (Fig. 4.10). Researchers often utilize the combination of these methods to increase the representativeness and randomness of the data samples.

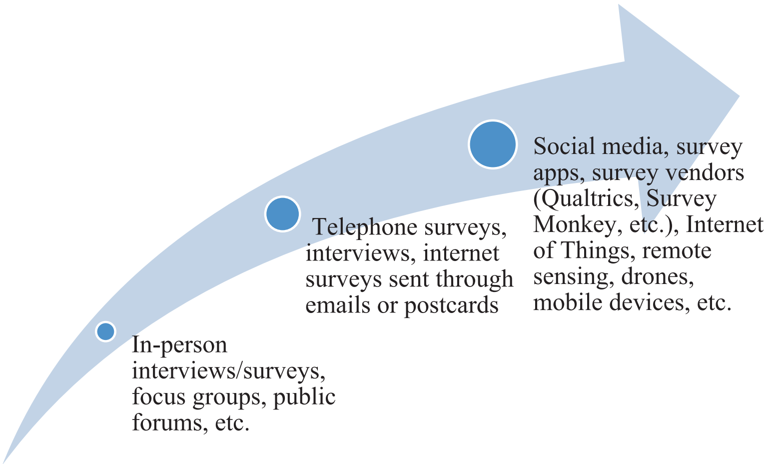


Fig. 4.10 Evolving data collection methods

The amount of data available sometimes can be overwhelming and selecting the right datasets for the research questions is important. Selecting the right methods to collect primary data is equally important if primary data will be used in the research. Representativeness, randomness, validity (accurate and correct answers), and reliability (responses are reliable and from targeted research subjects) are the fundamental principles one needs to follow to ensure the research findings are valid and reliable. In circumstances where data limitations cannot be overcome, data limitations and the limitations in research findings should be properly and sufficiently addressed in the research paper or report.

Analyzing the large amount of data collected with real-time technologies requires both sophisticated hardware and software resources. Under such circumstances the concept and application of machine learning and artificial intelligence has become more popular in urban and regional planning. These newer data analysis methods supplement the traditional descriptive, exploratory, and explanatory analysis in statistics, and provides powerful computing algorithm in urban analytics.

References

Andrade, C. (2020). The limitations of online surveys. *Indian Journal of Psychological Medicine, 42*(6), 575–576. https://doi.org/10.1177/0253717620957496

Babbie, E. (2012). *The practice of social research* (13th ed.). Wadsworth Publishing.

Baum, S. (2021). Secondary data in urban analysis. In S. Baum (Ed.), *Methods in urban analysis, Cities research series*. Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-16-1677-8_4

Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. Sage.

Gaber, J., & Gaber, S. (2007). *Qualitative analysis for planning & policy: Beyond the numbers*. American Planning Association.

Greenbaum, T. L. (1993). *The handbook of focus group research*. Lexington Books.

Smith, M., Szongott, C., Henne, B., & von Voigt, G. (2012). Big data privacy issues in public social media. 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST). https://doi.org/10.1109/DEST.2012.6227909

Williamson, C. (2008). *Planners and the Census*. American Planning Association. Planning Advisory Service Report.

Zhu, E., & Sobolevsky, S. (2018). *House price modeling with digital census*. arXiv, Cornell University. arXiv:1809.03834.

Web Resources

The U.S. Census Bureau: https://www.census.gov/

Desouza, K., & Smith, K. (2016). APA Planning Advisory Service (PAS) Report 585. *Big Data and Planning*.

Chapter 5
Data Presentation

The past half century has seen a tremendous increase in urban data, and tools and methods to analyze and present these data. Planners must present information to the public, clients, or key interest groups. For many professions, presentation is a one-directional process to inform the audiences. For instance, businesses present quarterly earnings to maintain public's confidence in them. College instructors make class presentations for the purpose of educating students. Clarity, logic, simplicity, and vividness are key elements for creating a more effective presentation. For planners, however, there are multiple purposes for a presentation. Planners need to inform the audiences of important issues about the community. Planners need to identify patterns, trends, and relations of different issues. Planners need to educate the public on the priority of focuses. Planners need to persuade the audiences and gain their agreements on further actions in a planning process. Presentation has become a part of the problem solving of a planning process. Effective data presentation is one of the determining factors for the success of a plan. This chapter focuses on issues related to data presentation in the planning area. First, the concepts of data, information, and geographical information are compared. Then, we assess the role of presentation in a planning process. The third and fourth sections of this chapter explore relevant factors related to presentation and propose ten rules for presenting information effectively for planners. At last, we address the importance of effective presentation for conforming to the AICP code of ethics.

Data, Information, and Geographic Information

Data are neutral descriptions of the properties of people, places, or events. They can be useful or not useful for a planner or a public administrator. For instance, all births or deaths need to be registered with local governments, in most cases, with the

---

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_5].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_5

Department of Health. Such data have simply existed since the early twentieth century for most states in the United States.

When data are transformed into a more meaningful form, it is often referred as ***information***. Data could be simply formatted or re-formatted to be more representable. Data can be grouped, re-organized, or represented in another form to provide more useful information to users. Data can be further analyzed to offer more insights about an issue, event, or activity. For instance, the graph on the left side of Fig. 5.1 presents raw birth data obtained from the Ohio Department of Health. This data was directly generated from individual birth reports, with one record representing one birth. Without further reorganization or manipulation, the data itself is not informative. This data can be regrouped based on another attribute, for instance, by mothers' age or by race. The graph on the right side of Fig. 5.1 presents number of births and percentage of births by mothers' age groups. This regrouped data, with proper formatting, provides meaningful information about birth activities by women's age group. Most births are from females in their twenties or early thirties. The Pivot Table function in Excel is a useful tool to group data. Box 5.1 presents brief instructions about how to conduct a Pivot Table analysis.

If information is geographically indexed, it is ***geographical information***. A geographical index can be a number, a name, or a sign that indicates a location. FIPS (Federal Information Processing Series) codes are a group of geographical indices for states and counties in the United States. These codes are available at the U.S. Census Bureau website (https://www.census.gov/geographies/reference-files/2018/demo/popest/2018-fips.html). The indices for smaller geographical areas, such as census tracts, block groups, and blocks (Fig. 4.5 in Chap. 4), are further developed based on state and county FIPS. Box 5.2 explains how census tracts and block groups are coded. Geographical information can be presented and analyzed in mapping software, for example, the ArcGIS.

YEAR	AGE	GENDER	CNTY	RACE
2000	28	M	06	1
2000	30	F	02	1
2000	39	F	33	1
2000	34	F	02	1
2000	28	M	02	1
2000	24	F	33	1
2000	23	M	02	1
2000	35	F	02	1
2000	29	M	02	1
2000	17	M	02	2
2000	32	F	02	1
2000	23	F	02	1
2000	18	F	02	1
2000	22	M	06	1
2000	28	M	02	1
2000	22	M	02	1
2000	28	M	02	1
2000	31	M	69	1
2000	25	M	69	1
2000	23	M	02	1
2000	27	M	02	1
2000	30	M	69	1

Administrative Birth Data

Mother's Age	Number of Birth	Percent of Total
Under 14	310	0.20%
15-19	18451	11.85%
20-24	41031	26.35%
25-29	43101	27.68%
30-34	34641	22.25%
35-39	15248	9.79%
40-44	2820	1.81%
45-49	113	0.07%
50+	6	0.00%
Total	155721	

Number of Birth by Mothers' Age Group

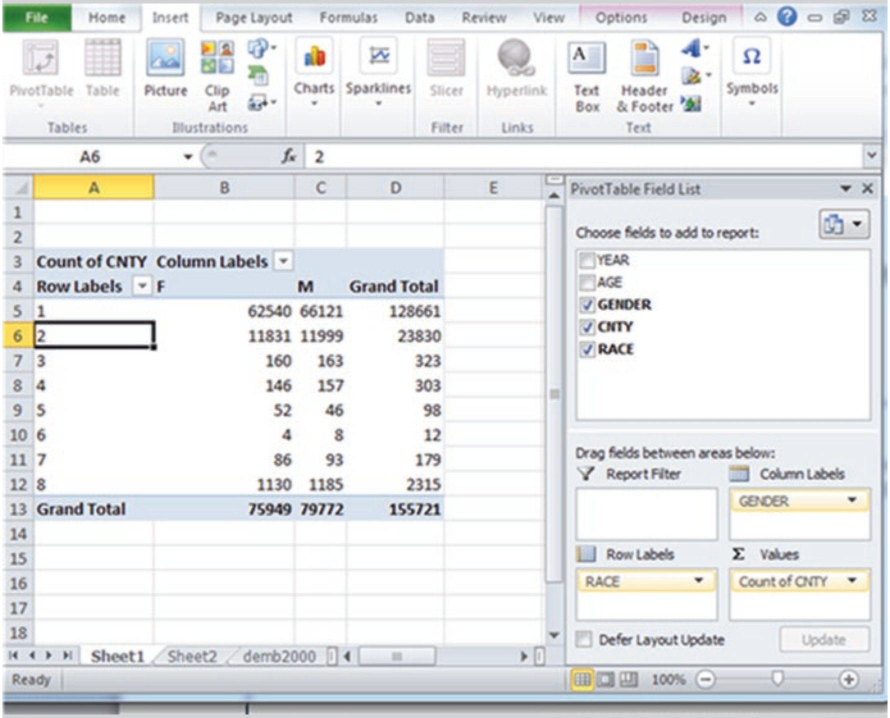
Fig. 5.1 Examples of data versus information

Box 5.1 Excel PivotTable Analysis

PivotTable analysis in Excel is a useful tool for planners to regroup and analyze data. We will use the Ohio birth raw data and illustrate how to conduct a PivotTable analysis in 2010 Excel.

The Pivot Table Analysis can be activated by clicking Insert > PivotTable.

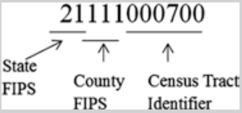
All the data will be automatically selected. We recommend reporting the pivot table as a new worksheet. The following graph illustrates how to create a table of number of births by gender and by race, by selecting GENDER as the column label and Race as the row label. In this dataset, there are eight races coded as 1 for White, 2 for American Africa, and 3–8 for other races. For ∑ values, common options are sum, count, and average in the field of planning practice.



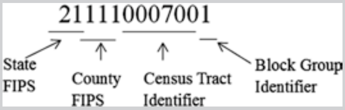
In the United States the Census Bureau is one of the primary providers of spatial information. Figure 4.5 illustrates the hierarchical relations among common census boundaries. Data at a higher hierarchical level can be obtained by aggregating the data at the lower hierarchy. This is **spatial aggregation**. For instance, the population of a county can be obtained by aggregating the amount of population of all census tracts within the county. Box 5.3 further explores possibilities of spatial aggregation within the Census data.

Box 5.2 Census Tracts and Block Groups Codes

A census tract code consists with state FIPS, county FIPS, and tract identifier. For instance, tract 7 in Jefferson county (with county FIPS 111), Kentucky (with state FIPS 21) is coded as:



A census block group code consists with state FIPS, county FIPS, tract identifier, and block group identifier. For instance, block group 1 in census tract 7 in Jefferson county, Kentucky is coded as:



Box 5.3 Data Aggregation based on Spatial Hierarchy

Figure 4.5 in Chap. 4 presents the spatial relations among common political and census designated boundaries. Summary level (SUMLEV) indicates the spatial hierarchy. For instance, if SUMLEV equals 150 for a record, this indicates that this record is an observation at the block group level.

The following matrix summarizes the applicability of data aggregation between different spatial boundaries. A grey cell indicates an applicable aggregation and the arrow points the direction. For instance, county level data can be aggregated into national level data. This matrix only lists boundaries that are commonly used by general planners. For boundaries not listed here, aggregation decisions should be made based on exploring spatial relations among the boundaries. A blank cell in the matrix suggests that further research is needed about the applicability. For instance, place boundaries are changing all the time, and are not guaranteed to be consistent with any other boundaries. But a BG-to-place aggregation is still possible if the BG boundaries happen to be consistent with the place boundary. Or if the planner can identify spatial overlay relations between BG and place boundaries, the aggregation can be made by using the areal interpolation method.

The Matrix of Data Aggregation Applicability among Boundaries

	Nation	State	MSA	County	Places	ZC5	CT	BG
State	↑							
MSA								
County	↑	↑	↑					
Places								
ZC5	↑							
CT	↑	↑	↑	↑				
BG	↑	↑	↑	↑			↑	
TAZ			↑	↑				
Block	↑	↑	↑	↑			↑	↑

MSA: Metropolitan Statistical Areas
ZC5: 5-digit Zip Code
CT: Census Tract
BG: Block Group
TAZ: Traffic Analysis Zone

Information and Planning

Friedmann (1987) explains planning as a process to link actions to knowledge. Forester (1988) argues that information supplies solution and constitutes a source of power for traditional planners to work within an administration. More recently, researchers and theorists have argued for communicative rationality, which emphasizes on inter-subjective decision-making, mutual understanding, and series of communicative actions (Healey, 1992; Innes, 1998). In a democratic society, planning in the public sector offers the public a platform to deal with urban issues collectively and interactively. Effective presentation of information ensures the influence of information in a planning process and is critical for the success of a plan. Figure 5.2 presents the role of data analysis and presentation in a planning process.

Planning practitioners transform data into information and geographical information and present it to the public. Information uncovers relations and patterns that are hidden in a set of data. Spatial information further presents spatial patterns. Researchers are more interested in synthesizing empirical evidence to build up knowledge to understand urban dynamics. For instance, in the past half century, planners and researchers have conducted enormous amount of empirical/theoretical works about urban sprawl to understand its causes, consequences, and solutions. With more people understanding and accepting this knowledge, it is easier for planners to promote more sustainable development patterns. Generating consensus is

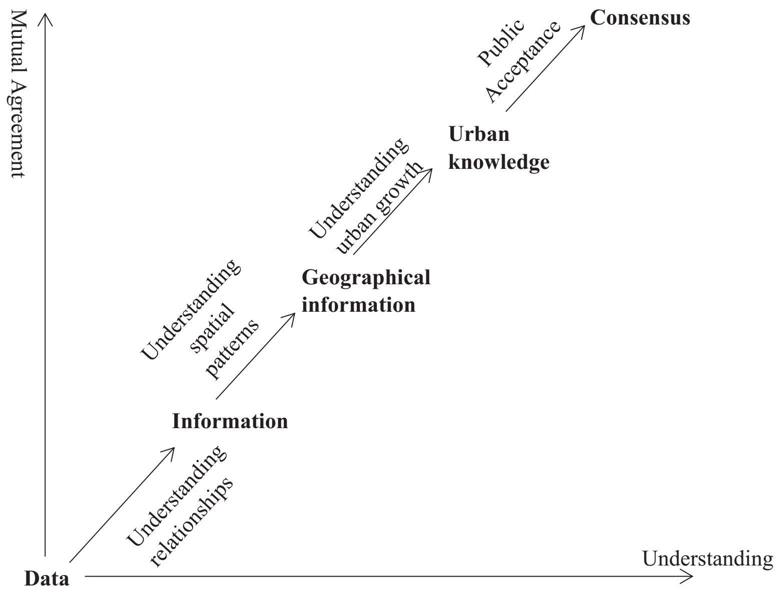


Fig. 5.2 Data, information, and consensus building

the goal of a planning process. An effective presentation of information is expected to better educate the public and to promote mutual agreements between different interest groups.

Data Presentation

Data are collected and analyzed to produce meaningful information. Data collection has been discussed in Chap. 4, and data analyses will be covered in Chaps. 7, 8, 9, 10, 11, 12, 13, 14, and 15. The following only presents issues related to improving presentation efficiency.

Audiences

People have varying perceptual ability to evaluate, process, and digest information, depending on their age, education level, and other personal attributes. Desired impacts of a presentation on audiences in a planning process include (1) understanding the information, (2) being interested in the issues, (3) being able to raise relevant questions that deepen the discussion, and (4) generating consensus. Undesired potential impacts of a presentation include unable to understand the information, misunderstanding information, being overloaded of information, disliking the information, and unable to generate consensus. Planners need to take consideration of audiences' interpretation constraints. Present the information in a way that facilitates planning goals, rather than the opposite.

Table 5.1 analyzes potential audiences for a presentation in the process of different types of planning, as introduced in the APA's Planning and Urban Design Standards (2006). It is critical to consider the nature and the scale of a plan, and to present information in a way that will encourage audiences' participation. For instance, Clarksville, IN is part of the Louisville Metropolitan Statistical Area. It is a small town with 21,714 population in 2010. In the process of preparing its comprehensive plan in 2015, three public meetings were conducted. A group of students from the planning program at the University of Louisville presented their findings related to land use, economy, and the environment to the public. It was found that the public was more interested in information that directly compared the town with its neighboring competitor, Jeffersonville, IN, rather than analytical results, for instance, spatial patterns of population density and distribution of environmental elements. Given this group of audience, planners should try to avoid any jargons that could cause confusions and should simplify their presentations to keep the audience focused. On the other hand, a transportation plan will involve community leaders and government officials from all levels. These people are equipped with proper background knowledge that will enable a deep discussion. Planners can adjust their presentation and select appropriate tools to fit the needs of different audiences.

Table, Graph, and Map

A **table** consists of columns and rows, and presents numerical, text, or symbolic information of a common theme. Row and/or column identifiers are used to identify information. For instance, Table 5.2 presents the employment time series of the United States over 2000–2019, and only has one set of row identifiers to identify time. This is a one-dimensional table. Table 5.3 uses a set of column identifiers and a set of row identifiers to present the employment time series of the U.S., Florida, and Kentucky. This is a two-dimensional table. There could have more than one set of row or column identifiers nested with each other. For instance, Table 5.4 uses two sets of column identifier (one for location and another for time), in addition to one set of row identifiers for race. This is a three-dimensional table.

A **graph** visualizes information and is helpful in presenting trends and comparisons. Figure 5.3 illustrates a column graph of the information presented in Table 5.2. Graphs are more effective in presenting relationships or trends than tables. With the employment trend as displayed in Fig. 5.3, a planner or a decision-maker can easily identify economic downturns. Box 5.4 presents common graphical presentation tools.

A **map** presents spatial information, and is a critical tool for planners to explore growth patterns, locational convenience, and environmental issues. For instance, the population density map (Fig. 5.4) enables regional planners to identify the spatial

Table 5.1 An analysis of audience by different types of plans

	Scale of analysis	Focuses of data presentation	Audiences
Zoning/rezoning	Property/ community	The relation between a given property and a community	Individual property owners
Comprehensive plans	Regional	All relevant issues	Local officials, community leaders, major institutions, investors, special interest groups, individuals
Urban design plans	Community	Design aspects	Local officials, community leaders, major institutions
Neighborhood plans	Community	Detail goals, policies, and future patterns of a neighborhood	Residents, businesses, nongovernment organizations, local officials
Downtown plans	Community	Downtown revitalization and visionary plan for the future	Governments, nongovernment organizations, community leaders, major institutions, special interest groups
Redevelopment area plans	Community	Redevelopment of a business district, a residential area, or an industrial site	Local officials, individuals, businesses
Economic development plans	Local/ regional	Stimulate economic growth and preserve jobs	Local officials, businesses, individuals, nongovernment organizations
Corridor plans	Community/ regional	Potential impacts of a public investment (e.g. highway, transit, utility, etc.)	Local officials, businesses, special interest groups
Transportation plans	Local/ regional	Maintenance of the existing system and plan for future facilities	Federal, state, and local officials, public transit provider, communities
Housing plans	Community	Address affordable housing needs and promote social integration	Local officials, communities, special interest group
Community facilities plans	Local/ regional	Provide necessary supports (educational, social service, cultural, recreational, healthcare, public safety, public utilities etc.) for future land use patterns	Officials, communities
Parks and open-space plans	Local/ regional	Protect natural resources, create recreational spaces, promote public safety, and reshape development patterns	Local officials, special interest groups, communities
Critical and sensitive area plans	Local/ regional	Protect critical and sensitive areas (e.g. wetland, water bodies)	Officials, special interest groups
Hazard mitigation plans	Local/ regional	Reduce long-term risk to human life and property	Officials, businesses

Table 5.2 U.S. 2000–2019 employments

Year	Employment
2000	114,064,976
2001	115,061,184
2002	112,400,654
2003	113,398,043
2004	115,074,924
2005	116,317,003
2006	119,917,165
2007	120,604,265
2008	120,903,551
2009	114,509,626
2010	111,970,095
2011	113,425,965
2012	115,938,468
2013	118,266,253
2014	121,069,944
2015	124,085,947
2016	126,752,238
2017	128,591,812
2018	130,881,471
2019	132,989,428

Table 5.3 2000–2019 employments of the U.S., Florida, and Kentucky

Year	U.S.	Florida	Kentucky
2000	114,064,976	6,217,386	1,513,722
2001	115,061,184	6,431,696	1,497,466
2002	112,400,654	6,366,964	1,462,517
2003	113,398,043	6,549,488	1,471,878
2004	115,074,924	6,864,987	1,489,497
2005	116,317,003	7,107,378	1,514,199
2006	119,917,165	7,535,515	1,552,012
2007	120,604,265	7,425,331	1,550,192
2008	120,903,551	7,366,571	1,570,800
2009	114,509,626	6,861,612	1,486,545
2010	111,970,095	6,626,558	1,456,790
2011	113,425,965	6,732,639	1,463,173
2012	115,938,468	6,932,382	1,481,323
2013	118,266,253	7,134,644	1,507,375
2014	121,069,944	7,441,584	1,535,417
2015	124,085,947	7,777,990	1,579,477
2016	126,752,238	8,169,642	1,603,173
2017	128,591,812	8,385,577	1,625,006
2018	130,881,471	8,669,611	1,642,234
2019	132,989,428	8,860,042	1,666,637

Table 5.4 Racial compositions

	United States		Florida		Kentucky	
	2000	2010	2000	2010	2000	2010
White	75.1%	76.3%	78.0%	77.3%	90.1%	87.5%
Black	12.3%	13.4%	14.6%	16.9%	7.3%	8.50%
Other	12.6%	10.3%	7.4%	5.8%	2.6%	4.0%

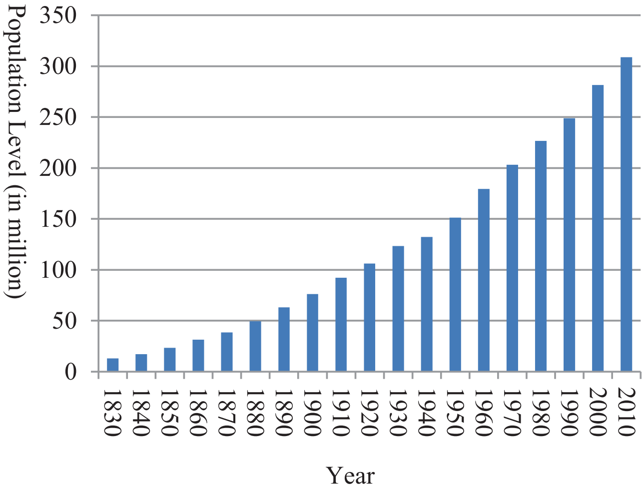


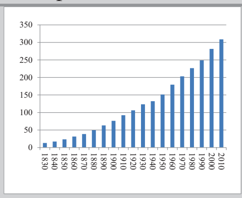
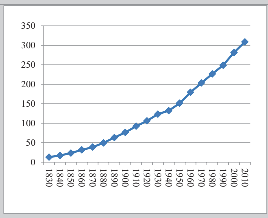
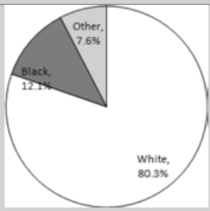
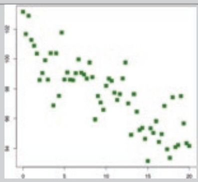
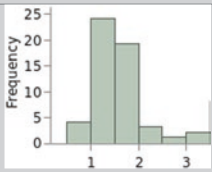
Fig. 5.3 2000–2019 U.S. employment

pattern of urbanization in the U.S. Basic elements of a map include title, frame, legend, scale bar, and north arrow.

Table, graph, or map, what should be used? Table 5.5 compares between these three sets of tools. If the goal of a plan is to uncover the uniqueness of a community, both tables and graphs are reliable tools to present symbolic and discrete information and illustrate differences. Tables are more effective in presenting multi-attribute information than graphs or maps. If the goal of a presentation is to provide a summary of the social-economic environment of a community, a table presentation should be chosen. For instance, the U.S. Census Bureau provides quick facts of the U.S., states, counties, cities, towns, and zip codes (The U.S. Census Bureau, 2020). These are summary information of a community and cover subjects of demography, economy, housing, poverty, education, and so on. Table 5.6 illustrates housing market quick facts of the U.S., Florida, and Kentucky. It is impossible to present all the information with one map. It is also unlikely to present the information with one graph, because of the varying scale of data.

Graphs are more effective than tables to identify trends over time and relationships within the dataset. However, tables preserve the most information. This information can be used in further analyses. If the audiences can further interpret tabular data and identify associations within the data themselves, tabular presentation

Box 5.4 Common Graphical Presentations

	Example	Comments
Column/bar graph		Use horizontal bars or vertical columns; Can be grouped or stacked; Present the trend for a single group; Compare levels between multiple groups.
Line/trend graph		Present one or multiple trends; Compare multiple trends.
Pie graph		A circle divided into portions; Compare parts as relation to a whole; Not an effective presentation when the number of parts increases; Doughnut chart is a variation of pie chart.
Scatter plot		Also named X-Y plot; Identify correlation between variables.
Histogram		Look like a column graph; Present frequency information.

allows for more in-depth conversations. Mapping is a tool for planners and researchers to identify and present spatial patterns to the public.

There are potentially greater confusions/errors/distortions with graphical and mapping presentation than with tabular presentation. The choices of presentation scale, format, and color could affect the effectiveness of a presentation. These will be further addressed in the next section.

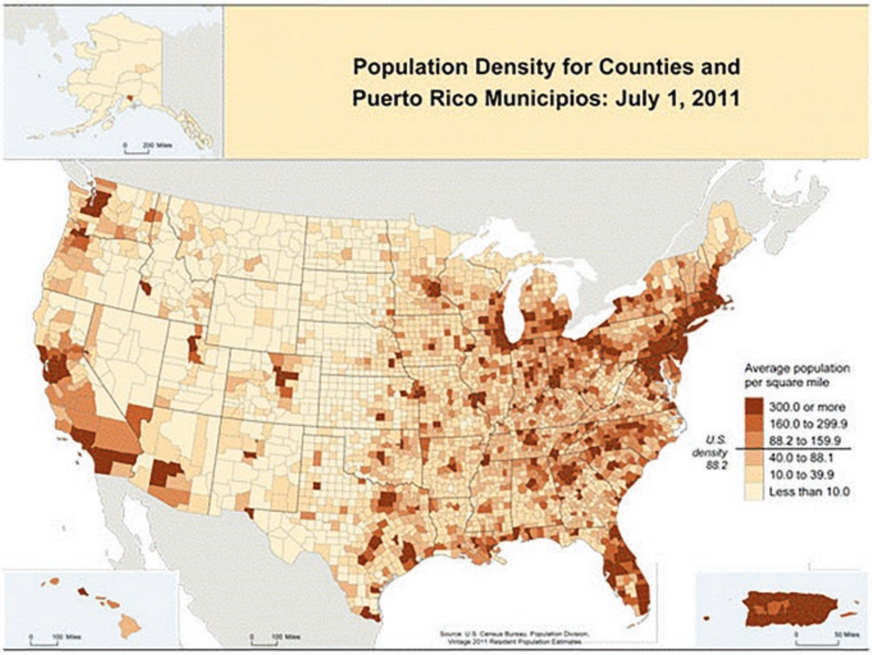


Fig. 5.4 Historical U.S. populations

Table 5.5 Strengths and weaknesses of tabular, graphical, and mapping presentations

	Table	Graph	Map
Illustrate difference	Yes	Yes	
Present multiple attributes	Yes	Limited	Limited
Summarize	Yes		
Identify trend		Yes	
Uncover relationship		Yes	
Identify spatial patterns			Yes
Potential to distort information?	Low	High	High

Ten Rules for an Effective Data Presentation

Wainer (1984) published an article about how to display data badly in the Journal of the American Statistician. The purpose of the paper is to encourage people to avoid bad practices and present information clearly and accurately. Some of his arguments are outdated, for example, the concern about image density. Most arguments are still meaningful. Starting with Wainer's list, we consolidate some of the rules and re-interpret them from a planner's view. We introduce new rules and propose the following ten rules for data presentation.

Table 5.6 Housing market information

	Florida	Kentucky	United States
Housing units, July 1, 2019, (V2019)	9,673,682	2,006,358	139,684,244
Owner-occupied housing unit rate, 2015–2019	0.654	0.672	0.64
Median value of owner-occupied housing units, 2015–2019	215,300	141,000	217,500
Median selected monthly owner costs -with a mortgage, 2015–2019	1503	1178	1595
Median selected monthly owner costs -without a mortgage, 2015–2019	505	375	500
Median gross rent, 2015–2019	1175	763	1062
Building permits, 2019	154,302	11,811	1,386,048

Source: The U.S. Census Bureau (2020)

Table 5.7 Popular writing styles

	Space between sentences	Font style	Font size	Page margin
Modern Language Association (MLA)	Double space	Recommend Times New Roman	12 pt	1″ on all sides
American Psychological Association (APA)	Double space	Times New Roman	12 pt	1″ on all sides
The Chicago Manual of Style	Double space	–	Not less than 10 pt., 12 pt. preferred	Between 1″ and 1.5”
The Harvard Style of Writing	Double space	Standard Font	12 pt	Uniform

Rule 1--Consider audiences' constraints and preferences in reading Writing styles have been codified to suit the needs of authors, readers, and publishers, and to facilitate the presentation of knowledge (Bazerman, 1988). Table 5.7 summarizes the requirements and recommendations related to formats from four popular writing styles in the areas of social and human sciences. There are basically two groups of standard fonts, Serif and San Serif. Serif fonts use small strokes on the ends of letters and symbols, and San Serif fonts do not have such strokes. Table 5.8 presents the words "Urban Planning" in popular Serif and San Serif fonts, all in the same size. Serif fonts are easier to read in print out format and have been commonly used in printing. However, San Serif fonts are easier to read than Serif fonts on internet. In the academic area, Times New Roman is preferred because of its readability in printed works.

Planners should also consider audiences' visual acuity and use design tools to help people see and digest information. Various visual tools (e.g., colored text, underline, local magnifying) can be used to highlight important information. Visual structure helps people to better understand information. Elements with visual structure include heading, subheading, vertical or horizontal space, or lines.

Table 5.8 Popular serif and san serif fonts

Serif fonts	Example	Sans serif fonts	Example
Century Schoolbook	Urban planning	Arial	Urban planning
Courier New	Urban planning	Comic Sans MS	Urban planning
Georgia	Urban planning	Tahoma	Urban planning
Times New Roman	Urban planning	Verdana	Urban planning

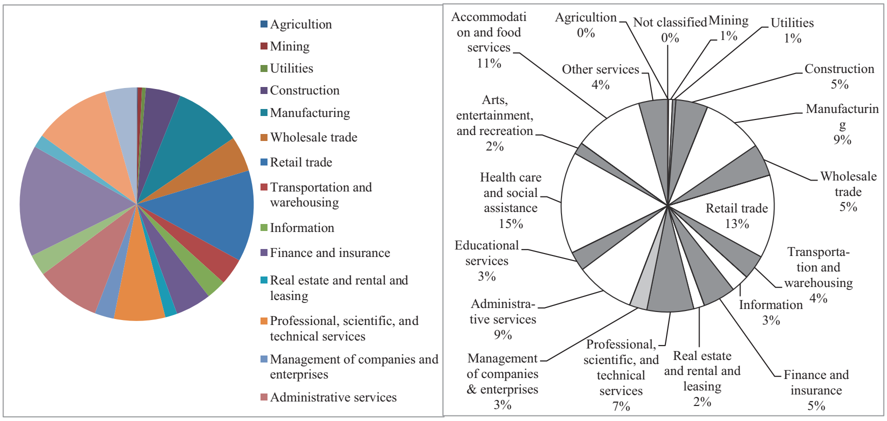


Fig. 5.5 The 2015 U.S. economy

Rule 2—Be careful with colors Color is a useful tool to display information. In graphs, colors can be used to indicate different type of activities. In maps, colors are used to display different types of roads, different levels of activity densities, or different types of soils. Do not overdo coloring. This is especially important when displaying a set of economic data with many industries. Figure 5.5 illustrates an example. It presents the U.S. employment distribution by industry in 2015. The figure on the left side uses the format automatically generated by Excel, i.e., using colors to identify industries. Theoretically, human eyes can distinguish millions of colors, hues, and shades. But practically, when the number of colors is more than 5, it becomes challenging for readers to match colors with their corresponding labels. The figure on the right side presents a solution to this problem, direct labelling. Only three colors (white, light grey, and dark grey) are used to differentiate between different industries. As for using colors in mapping analyses, the four-color theorem states that at most four colors are needed to display regions such that no two adjacent ones have the same color (Ringel, 2012).

It is important to think about the relation between the choice of color on a map and the story one is trying to tell. For instance, if the project is to map the distribution of population density and you use different colors to indicate different levels of density, do not randomly select colors. Instead, use a series of shaded colors, with a darker color indicating a higher level of density (Fig. 5.4). In addition, when mapping land-use activities, there is a traditional color list for different type of land uses.

The American Planning Association's recommendations are red for commercial, yellow for residential, forest green for agricultural, grey for transportation, purple for industrial, and blue for education and social activities (The American Planning Association, 2020).

Rule 3—Be careful about wording Planners need to be careful about ambiguous words, such as "sometimes" and "often", and pursue clarity in a presentation. As for negative words, such as not, un-, less, shorter, declining and so on, researchers in the medical area (Wright et al., 1998) recommend avoiding them, because people are more error-prone when involve negatives and positive information is more succinct in many cases. Some other researchers (Ito et al., 1998) find that there are stronger reactions from people on negative information, and this is call "negativity bias" in Psychology. The implications for planners are that negative information could be a barrier in consensus building, but it also can be a tool for planners to raise public's attention. For instance, a planner of a city needs to present the median household income of two communities within the city, $45,000 for community A and $50,000 for community B. The following are the three options:

Option 1: The median household income is $45,000 for community A and $50,000 for community B;
Option 2: The median household income of community B is $5000 more than community A; and
Option 3: The median household income of community A is $5000 less than community B.

Option 1 presents the facts with no further interpretation, option 2 interprets this as a positive finding for community A, and option 3 presents the same information as a negative finding for community B. Presenting the information with option 3 may cause anxieties in the public. If a planner's goal is to focus on regional scale comprehensive planning, it may not be necessary to go with this option. However, if it is a planning effort on promoting social equity and diversity, presenting the information with option 3 will help planners to gain the attention from the public.

Rule 4—Use comparisons Presenting information in a context helps audiences to better understand it. For instance, according to the employment data from the County Business Patterns database, the total employment of Kentucky has increased by 8.4% over 2010–2015. With this information, a planner can conclude that the Kentucky economy was growing. The 2010–2015 employment growth rate for the U.S. is 10.8%. With this additional information, the planner can further conclude that the growth of the Kentucky economy is behind the national average. The mother region, i.e., the bigger region that contains the subject area, is a good choice for making comparisons. For example, the United States is the mother region for Kentucky, and Florida is the mother region for Orlando, FL. A similar region offers another good choice. Louisville, KY and Cincinnati, OH are similar to each other. Both cities are in central counties on state boundaries and the growths are constrained by the Ohio River. Both economies have been affected by the economic transition from the manufacturing-based to the service-based, and then to the

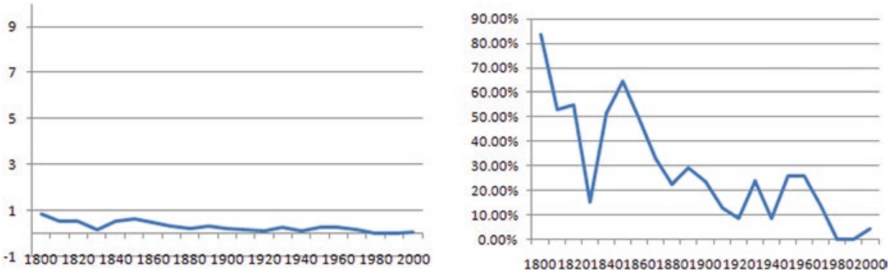


Fig. 5.6 Historical population growth grates of Jefferson County, KY

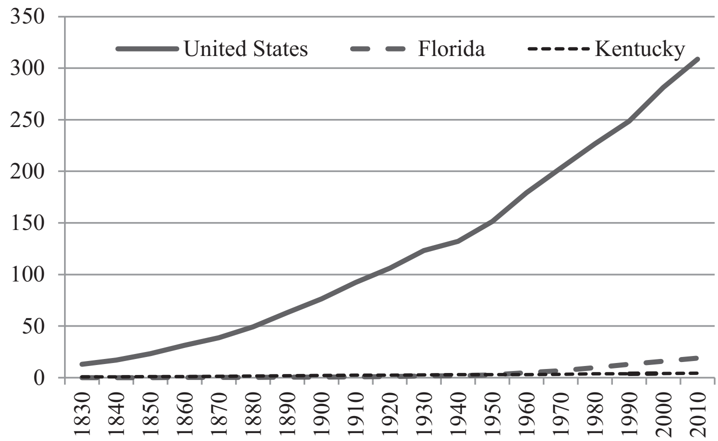


Fig. 5.7 Historical populations of the U.S., Florida, and Kentucky

information-based. Cincinnati can serve as a decent comparison for studying Louisville's economy.

Rule 5—Do not hide information For a graphical presentation, information could be hidden if the scale is not appropriate. The two graphs in Fig. 5.6 both present population growth rates over 1800–2000 for Jefferson county, KY. The one on the left side use an inappropriate scale for growth rate between −1 and 9, and the dynamics of growth trend cannot be shown. When the scale is appropriate (please see the graphs on the right side), people can identify the change of growth rate over time. When two groups of data with different scales are presented on the same scale, information could be hidden as well. Figure 5.7 presents U.S., Florida, and Kentucky population on the same graph, and Florida and Kentucky information is hidden.

Rule 6—Be Strategic about how much to present Judgements need to be made to balance between presenting clearly and presenting more. Figure 5.8 presents historical population growth rate information for the counties in the Louisville

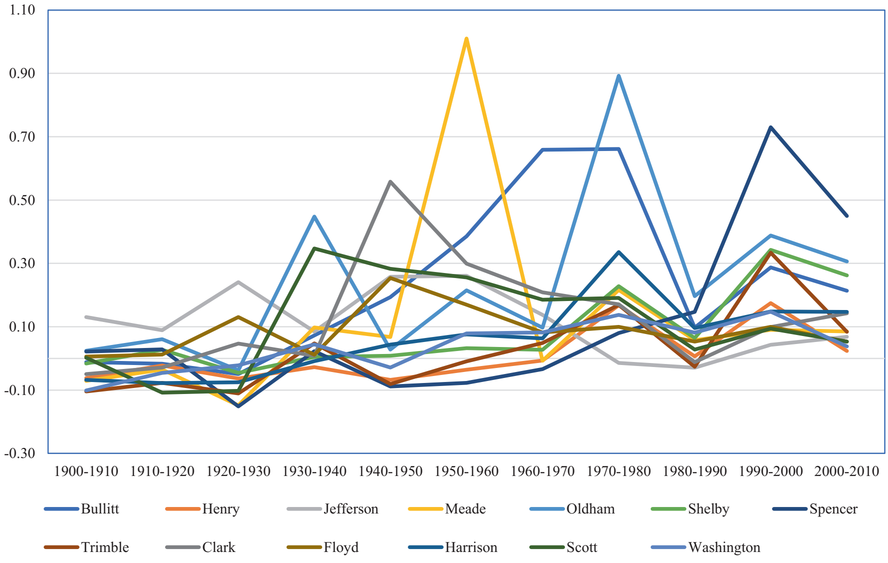


Fig. 5.8 Historical population growth rates for counties in the Louisville MSA

MSA. One can easily compile the historical information of the 13 counties in the MSA and generate this figure in Excel. This figure is not readable. Figure 5.9 provides another example of two groups of presentations of the same information. On the left side, three line graphs present historical population growth information. On the right side, the three graphs are consolidated as one. This graph directly informs the audiences that Florida population growth has been consistently higher than the U.S. average, but with a declining trend. Kentucky growth is more stable, but consistently below national averages. More information (with a similar scale) on the same graph enables a better and more comprehensive interpretation. But with too much information, the graph is not readable for any audience. The presenter needs make decisions about what is the optimal amount of information.

Rule 7—Respect the internal relation within data and be consistent All quantitative data has a natural order, the numerical order. Some qualitative data also has a natural order, for instance, the satisfaction level (that can be low, moderate, and high). Follow the natural order and do not change scale in between. This improves the logic and accuracy of a presentation. For instance, in Fig. 5.10, there is a temporal order for time information on the horizontal axis and a numerical order for the growth rates of the vertical axis. This figure provides misleading information about a declined growth rate in the 1990s. The figure presents 10-year growth rates except for 1990–1995 and 1995–2000. Figure 5.11 presents the population pyramid for Jefferson County, Kentucky. A population pyramid presents the distribution of population by age cohort and gender. This will be further explored in Chap. 7.

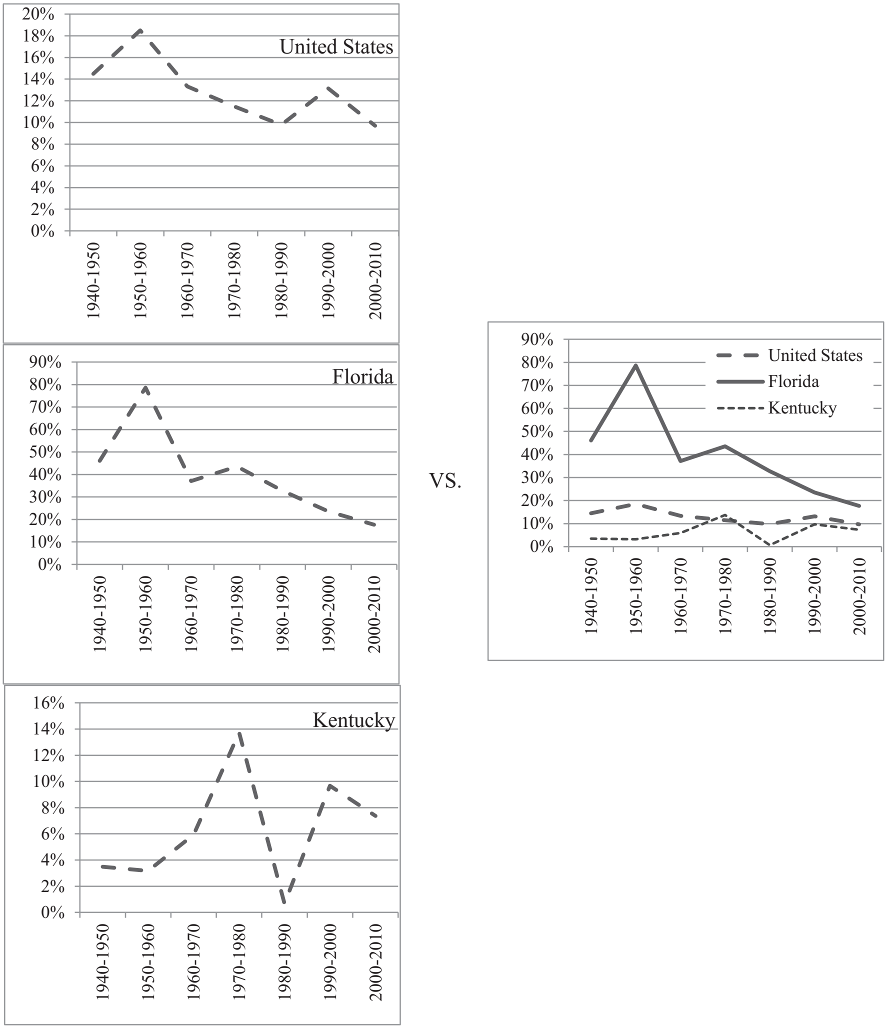


Fig. 5.9 Population growth rates of the U.S., Florida, Kentucky

Figure 5.11 can mislead audiences in thinking that Jefferson County had less population in the age groups 15–24 and 60–69 in 2005. This misleading information was created because the sizes of all other age cohorts are 5 years, but 1–2 years for the problematic cohorts.

Rule 8—Emphasize the important The presentation of data serves a set of purposes. The same set of multi-attribute data can be presented in different ways and emphasizes different issues. For instance, the two graphs in Fig. 5.12 are made based on the same information, percent of workers by industry and gender. If the focus of the analysis is to explore gender differences, the graph on the top serve the

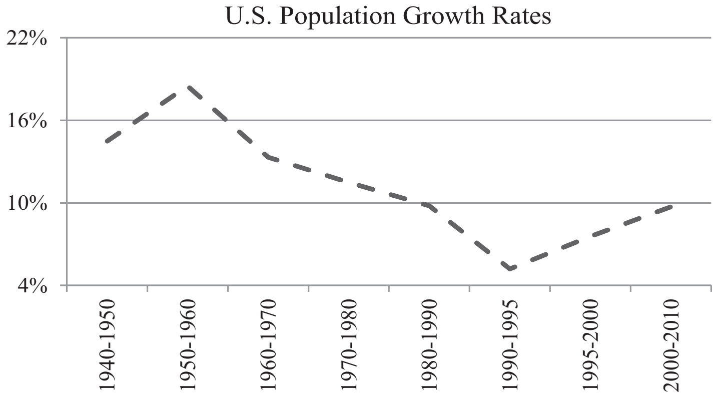


Fig. 5.10 U.S. population growth rates

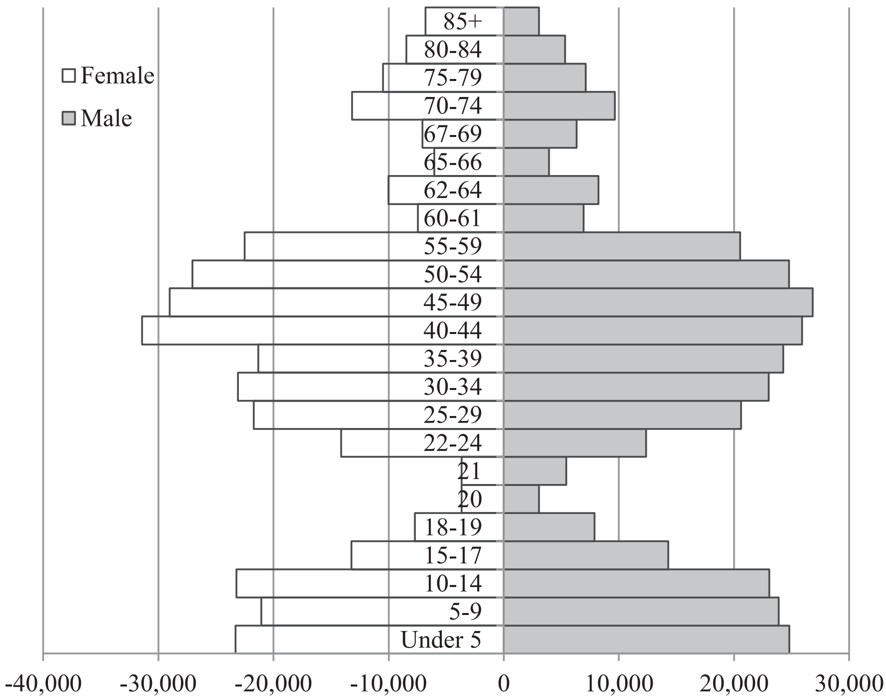
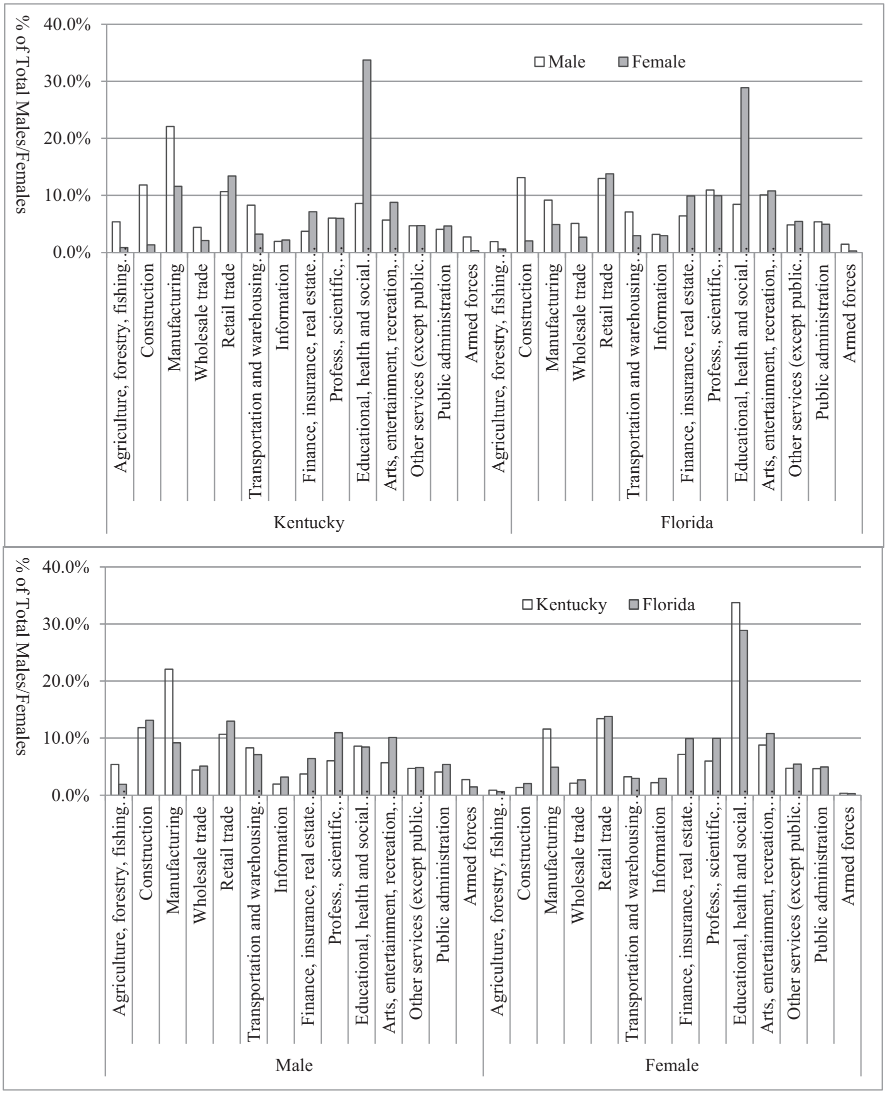


Fig. 5.11 2005 population pyramid of Jefferson County, Kentucky

topic better than the graph on the bottom. The graph on the top compares male and female differences. It suggests that for both states, there are more males working in Construction and Manufacturing industries, and more females working in Education, Health, and Social Service industry.



Data source: 2000 Census Transportation Planning Package

Fig. 5.12 Employment distribution by industry and gender. (Data source: 2000 Census Transportation Planning Package)

Rule 9—Be careful about templates Do not reply too much on software. Microsoft Excel is commonly used for making tables and graphs. Many of the misleading information mentioned above was created because people let the software decide on the layout of elements, the choice of colors, and the selection of scales. For instance, the graph on the left side of Figs. 5.5, 5.7 and 5.8 are all automatically generated by Excel.

PowerPoint is often used to create presentations. There are pre-determined presentation structures with title, subtitles, and text all arranged. These pre-arrangements encourage presenters to write more text and read through the presentation files. Such presentations are not attractive and cannot effectively instruct the audiences. Design templates may contain unnecessary ornaments and animations, acting as a strong visual block for a presentation. It could distract audiences from the focus of the presentation itself.

Rule 10—Consider limitation of printing We are in an information society, and highly rely on internet to circulate information. Planners post reports on websites for the public to access, and email documents directly to relevant interest groups. However, many people still prefer to print out documents to read. Colorful printing costs more. A colorful document is likely to be printed as black and white. In this case, all colors will become different hues of grey. The left graph in Figs. 5.5 and 5.8 will become even more difficult to read. To solve this issue, people may reply on filling and line patterns and different shades of grey, rather than colors, when creating a graph. Or planners can select colors that contrast with each other in both colorful and black-white printouts.

Planning Ethics

The AICP Code of Ethics and Professional Conduct (APA, 2020) lists planners' overall responsibility to the public prior to the responsibility to the clients, employers, the planning profession, and the colleagues. There are two principles related to data presentation. The first one is that "we shall provide timely, adequate, clear, and accurate information on planning issues to all affected persons and to governmental decision makers." It is planners' responsibility to provide clear and accurate information. Misinformation, intentionally or unintentionally, can mislead the audience and cause distrust from the public (Forester, 1988). Inappropriate presentation of information could lead to ambiguity, confusion, redundancy, and over complexity or over simplification, and impede a planning process. Inappropriate presentation could also lead to hiding or distorting information and misleading the public.

Second, the Code of Ethics states that "we shall give people the opportunity to have a meaningful impact on the development of plans and programs that may affect them. Participation should be broad enough to include those who lack formal organization or influence." This principle urges planners to consider the target audiences. Academic literature suggests that high income population are more likely to participate in local politics (Junn, 2000; Laurian, 2004). Practitioners also noticed lower levels of public participation for public meetings in low-income neighborhood during the same comprehensive planning process. To improve participation rate from residents in those neighborhoods, one thing a planner can do is to make the presentation understandable for this group of people.

References

APA. (2020). The AICP code of ethics and professional conduct. https://www.planning.org/ethics/ethicscode/. Visited on 12 Dec 2020.

APA (American Planning Association). (2006). *Planning and Urban Design Standards*. Wiley.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science* (Vol. 356). University of Wisconsin Press.

Forester, J. (1988). *Planning in the face of power*. University of California Press.

Friedmann, J. (1987). *Planning in the public domain: From knowledge to action*. Princeton University Press.

Healey, P. (1992). Planning through debate: The communicative turn in planning theory. *Town Planning Review, 63*(2), 143.

Innes, J. E. (1998). Information in communicative planning. *Journal of the American Planning Association, 64*(1), 52–63.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75*(4), 887.

Junn, J. (2000). The significance of race and class for political participation. Paper presented at the conference. In *Political Participation: Building a Research Agenda*.

Laurian, L. (2004). Public participation in environmental decision making: Findings from communities facing toxic waste cleanup. *Journal of the American Planning Association, 70*(1), 53–65.

Ringel, G. (2012). *Map color theorem* (Vol. 209) Springer Science & Business Media.

The American Planning Association. (2020). *LBCS standards*. https://www.planning.org/lbcs/standards/. Visited on 15 Dec 2020.

The U.S. Census Bureau. 2020. QuickFacts. https://www.census.gov/quickfacts/fact/table/US/PST045219. Visited on 15 Dec 2020.

Wainer, H. (1984). How to display data badly. *The American Statistician, 38*(2), 137–147.

Wright, P., Jansen, C., & Wyatt, J. C. (1998). How to limit clinical errors in interpretation of data. *The Lancet, 352*(9139), 1539–1543.

Chapter 6
Communicating the Research Findings

The purpose of conducting planning research often requires communicating the results and findings to a target audience or the general public. *Communication format* can be either oral or written, and oftentimes ***oral communication*** requires the assistance of a ***written communication***. Both the oral and written communication skills can be trained and this book will mostly cover written communication, although the tips and pitfalls of public speaking will be briefly discussed as well. Communication is especially important given the changing paradigms of public relations and technology. The shifting discourse from "the experts" to "influencers" or "community voice", the impact of politics on research and science, the growing distrust in government and institutions, the proliferation of "fake news", the exponential growth in big data and detailed information, and the rapid development in infographic and GIS communications and platforms have shifted the definition and execution of communication into different discourses than traditionally depicted (Wyeth & Johnson, 2021).

Effective Communication

Communication skills, along with writing and research skills, are one of the most important skills valued by practicing planners (Greenlee et al., 2015; Miller, 2019). Urban planners are constantly facing various stakeholders in the plan-making and implementation process. From communicating within the planning department to communicating between public employees and private consultants, to public hearings, planning commission meetings, public forums and workshops, and education and awareness of planning among citizens, communication is key in each of these daily activities. In *Planners' Communication Guide: Strategies, Examples, and Tools for Everyday Practice*, published by the American Planning Association, effective communications for planners require (APA, 2006):

- Building relationships with target audiences
- An effective and genuine listener

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_6].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_6

- Strong spoken, unspoken, and written communication skills
- Persuasive and consensus-building

Audiences for planners include colleagues within the planning department or the planning firm, public officials, such as the mayor, city manager, planning commissioners, and city or county commissioners, developers, residents, business owners, media, interest groups, and local employers, to name a few. Understanding the fundamental concerns of the audience and the political, economic, and development environment in which audience members live is the first step in connecting with them. We also need to probe for each audience's history with planning so that we are able to evaluate the trust levels from these audiences. Meanwhile, secondary data sources, such as the Census data, will help us understand the residents in a community. Interviews, focus groups, workshops, focus groups, and public forums are all primary conduits to help us understand what the audience cares, needs, and wants.

In order to reach out to major audiences when conducting a communications campaign, audiences need to be segmented demographically, geographically, psycho-graphically, or by professional roles. Demographic segmentation focuses on characteristics such as age, race, gender, ethnicity, income, immigration status, etc., depending on the planning issues at hand. For example, when planning for recreational facilities, age is an important factor since recreational needs usually differ among different age groups. Geographical segmentation stresses where people live, work, play, shop, and study. ***Psycho-graphical segmentation*** focuses on lifestyles, attitudes, values, and beliefs of the residents. Segmentation based on professional roles reaches the audience by their professional roles, which may help reveal intermediaries which influence planning decision-making (APA, 2006).

Plan the Communication

Before communicating ideas or research findings to the targeted audience, one needs to plan what and how to communicate and convey the information (Wyeth & Johnson, 2021). Ask the following questions:

- Who is the audience?
- Which print, online, and social media outlets and forums should be used?
- Which form of communication should be used? What is the structure of the information? Which writing styles should be used? Which visualization and infographic information should be included in the communication?
- Should the information be freely available for any audience?
- Should the communicated information allow comments? If so, how to control the validity and monitor the content of the comments?
- Can the conveyed information be easily accessible on mobile devices?
- Can people with impaired vision or hearing have easy access to the materials?

Communication with visual cues and information tends to attract readers' attention and the use of texts and graphs should be well balanced to achieve the best readability. Common visualization tools in planning are graphics (charts, tables, sketches, illustrations, diagrams, photographs, etc.), GIS maps, dimensional models and simulations, video, audios, webinars, and PowerPoint present presentations.

Regardless of the communication format, the following are the commonly followed principles for different types of communication mediums.

Technical Writing Skills

Writing is one of the fundamental skills enabling effective communication. Creative writing and technical writing are often distinctively different from each other. The fundamental difference between creative and ***technical writing*** is that technical writing is often objective and based on factual evidence and support, while creative writing can be subjective and may be based on imagination and creativity. The reality may be augmented or fabricated in creative writing. On the other hand, technical writing presents nonbiased, fact-based, or even scientific concepts, models, or information, with a purpose to educate or persuade. In some sense, technical writing is more challenging than creative writing, due to its vigor in presenting or disseminating the information to the audience. Overall, technical writing should abide by the following principles:

- Clear
- Concise and Precise
- Consistent
- Correct (use of grammar and spelling, correct information and facts)
- Relevant
- Conveying up-to-date information

There are a few basic rules in technical writing. The first rule is to use formal language instead of slangs and common idioms. Formal written language is the accumulation of formal vocabulary and sentences, when following proper grammatical conventions and punctuations.

The second rule is to use the tone of the third person to describe the fact or information, and pronouns such as "I" or "you" should be carefully and sparely used. For example, in a planning project report, students are required to describe the current streetscape conditions. A student writes, "When you drive along this street to downtown, you will notice that streetscape changes from being unattractive to appealing." Fundamentally this is not wrong; however, if this is a formal plan or a research report, it may sound informal. If this is storytelling, it sounds personable. If it is in the formal plan, the student may want to change it to "In the sections leading to downtown, streetscape appears more appealing than the sections farther away from downtown." Therefore, using such impartial tone describing the fact sounds more professional and formal in a planning document.

The age of "internet of things" and social media has created a large amount of informal words, acronyms and idioms. "u" was used to replace "you", "your" is confused with "you are", "its" is confused with it's, and all these mistakes or informal uses of words and sentences should not be used in technical writing. Short forms of words, such as "you're", "I'm", "it's", "it'll", etc., should be spelled out as "you are", "I am", "it is", or "it will", etc. Sentences should have at least the basic structure of subjects and basic predicates such as verbs, and full sentences should be used in technical writing.

When writing technical reports or presentations it is also crucial to keep the emotion neutral and out of the writing scheme, even though certain argument is not what you support. Neutrally presenting the factual information will be appealing to a larger audience, especially in a setting that the report or document needs to be presented or shared among residents, businesses, and officials alike. For example, local officials think that your claim that the ecological system of the community has been compromised due to a few manufacturing companies not following environmental regulations is fabricated by you and certain interest groups. You are angered by their judgement; however, when you present the facts in a memo or any other types of documents, you should not allow your emotion to cloud the writing tone of the document. Regardless of how difficult it is to retain your anger, factual information should be presented with facts without strong emotional involvement, unless you are using the document as the basis for campaigns.

Another rule of writing technical documents is to pay attention to the audience. Different styles of writing and the use of languages will be different depending on your potential audience. For example, jargons that are highly specialized should be avoided if the document is intended for the general public. A rule of thumb when writing to the general public is to use languages that can be easily understood by a middle-school student. However, if the audience is the expert in the field, by no means, jargons and sophisticated language will often add rigor to the document.

In technical writing, correctively using grammar and words is very important. Regardless of contents, typos and poorly executed grammatical conventions will greatly impact the readability of the document. Copyediting and proofreading is critical to convey the information and ideas to the targeted audience. The process of technical writing involves around a few cycles of planning, writing, revising, rethinking, rewriting, and revising, until the document achieves clarity, brevity, and simplicity for the audience to understand and digest. Audiences and purposes of the document should be identified in the planning stage before drafting the outlines. For information not clear or foreign to the audience, further definition and explanation is often needed. Brevity should never leave critical information unexplained. Words used in spoken language but sounds informal should be avoided.

Types of Communication Conduit

Research and practice indicate that well-poised visual materials, layout, and formatting enhance text narratives tremendously. Formatting, narrowly speaking, does not include the maps, photos, and other content-related visual materials in a document. Generating professional and visually appealing maps can be reviewed from Chap. 5 of this book. Processing photos and other visual materials require certain techniques and the use of software, such as AutoCAD for architectural or engineering drawings, Adobe Photoshop for image or photo professing, and Microsoft Publisher or Adobe InDesign for document layout and publishing. Formatting usually indicates how the document is designed regarding the layout, font size, simple graphics, line spacing, format of the page numbers, alignment, margins, orientation, columns, etc.

For novel writers or presenters it is imperative to look for examples demonstrating best practice communication format, and then use these to guide the design. Office communication, such as emails, memoranda, and the staff reports, oftentimes does not need deliberate formatting, as long as one follows the basic guidelines (e.g. a memorandum would have headings such as memorandum, with addressing to, from, date, and subject). Formatting is mostly necessary for professional reports, papers, newsletters, and PowerPoint Presentation. Although there are sets of guidelines and principles regarding formatting for each type of communication conduit, common sense regarding professionalism often prevails. The following section will briefly explain the formatting rules and pitfalls under each type of written communication for research results.

Emails

Emails are popular communication tools in contemporary society and are sometimes used to communicate research results. Emails need to have a subject line, precise and to the point. There is no need to capitalize the initials for each word in the subject line, but doing so is socially acceptable. The body of the email should almost always start with a salutation, such as Dear Madeline, Dear Dr. Smith, Hello Sandy, etc. Default text size, font type, and color is commonly used; however, important information may be highlighted with bold, italic, larger, or colored text. To show respect, use sincerely, best regards, courteously, etc. before the signature line. Professional emails should have a full signature if possible, indicating the sender's name, title, affiliations, and contact information. Professional emails should have full sentences, and avoid excessively using anonyms and abbreviations. Overall, a professional email should abide by the following guidelines:

- Include a concise subject line.
- Always include a greeting (e.g. Dear Joe/Mr./Mrs./Ms./Dr., or Hello/Hi if you know the person really well) and a closing (Best regards/sincerely).
- Use business language and business format, check spellings carefully, and avoid confusing acronyms.
- Never use all capitalization letters.
- Be cautious about emotions, jokes and sarcasms, and avoid gossiping.
- The body of the email should be concise.
- Reread before hitting the "send" button.

Memorandum

Memoranda, often shortened as memos, are a form of formal letters and commonly used in workplaces. Research results may be formally communicated through memos. Memos should be concise and often should be limited to one to two pages. However, memos have gained popularity in planning communication, and sometimes planning memoranda have elements usually included in a staff report, such as key research findings. In such circumstances, a memo may be much longer. If additional information is needed, an appendix may be attached to a memo. Memos should almost always have a title "Memorandum" on the top center of the letter, followed by "Date, To, From, Re: Subject" left-aligned. Signature of the sender may be placed following the printed name of the sender. Usually a straight line across the paper will divide the content and headings of the memo. Underneath the straight line, the content of the memo should be precisely addressed per audience. If needed, subheadings may be added in the memo to enhance readability, if it is more than one page. Subheadings should be bold. After the content of the memo, names of the people who will be copied should be listed as well, for example, cc: Anthony Jones. If there are additional materials accompanying the memo, "enclosures" should be added as well.

Font size is usually 12 and font type is usually Arial or Times New Roman, to make the memo appear more professional. The following is an example of a short memo; however, certain memos may be much longer, and sometimes can be as long as more than 10 pages.

[Insert Agency Name and Logo]

Memorandum

Date: 10/20/2021

To: Jane Doe

From: Joe Doe, the Planning and Zoning Department

For: The Mayor and the City Council

Subject: Recommending the Creation of a Historic Preservation Advisory Board

---

On August 15, 2021 the City Council adopted the recommendation by the Planning and Zoning Department about establishing a Historic District within the city. To facilitate the citizen participation and the planning process of drafting the historic preservation ordinance, making the implementation of historic preservation accountable, the city's Planning and Zoning Department deems it necessary to establish a historic preservation advisory board. The board will

- consist of six members from the city and all are required to be the current residents within the city.
- The chair and vice chair of the board are required to have a background in history or historic preservation. All the other members could be from other various backgrounds.
- Demographically the board members should strive to best represent the city's demographic composition.
- All the board members should participate in the city's ethics training workshop.
- The process of appointing the board members should follow the same procedure as appointing members for other existing boards within the city.
- The board is required to have a regular meeting of once a month and all meeting materials and notices will be prepared by the appointed staff person Joe Doe from the Planning and Zoning Department.

Please contact Joe Doe if you have any questions.

Enclosure: The City's Board Appointment Policy and the map of the Historic District

cc: Jane Doe, Director of Planning and Zoning

John Doe, Assistant Director of Planning and Zoning

Staff Report

Staff report is a common type of communication tools among urban planners and public policy makers. Staff reports are concisely-written reports that professional planners use to "digest a particularly development proposal or permit request; assess site conditions; relate the proposal to the comprehensive plan, subplans, and functional plans; and make a recommendation", among all other functional purposes (Meck & Morris, 2004, p. 4). Similar to memoranda, the objectives of staff reports are to inform the intended audience, so that the audience will take action based on the report, within a set time frame. Meck and Morris (2004) indicate that there are 11 key elements of a staff report (p. 4):

- Cover sheet
- Project description, when applicable, including legal description which is necessary for a rezoning
- Factual information about the site and surrounding area, such as zoning, land use, land survey, public facilities and services serving the area, traffic counts, transportation network, floodplain or wetlands, and other pertinent information.
- Staff analysis, including presentation of decision-making criteria for the recommendation, applicable impact analysis, and consistency assessment of the proposed action with all applicable plans.
- Description of information yet to be submitted
- Comments from other agencies
- Staff recommendations, including conditions if appropriate
- Maps displaying subject property
- Photographs of the property, if appropriate
- Information submitted by applicant as attachments
- Written comments from citizens as attachments

Staff reports should be well-organized with all the elements. The cover page should include key information such as project name, report preparer's name, action requested, applicant name, date of hearing, etc. Sometimes, other information about the site, such as location, existing zoning, etc., may be included on the cover sheet.

When writing, one should get to the key points quickly in a staff report. The staff report should be technical and defensible, yet not too technical and thus readable to the general audience. The format should be consistent and interesting. Well-organized and designed matrixes, tables, charts, maps, and photos add tremendous value to the report. The report should explain acronyms and avoid jargons. When presenting the factual information, subjective information should be avoided. Since the staff report will be presented to stakeholders and community officials in a limited time frame, being concise, clear, thorough, and accurate with appropriate details is key to convey what the planner intends to communicate.

Staff reports should follow the three Cs (Swift, 2011):

- ***Compliance***. The report should identify how the project complies with municipal codes, land use, comprehensive plan, and regulations in environment, utility, traffic, and other elements.
- ***Consistency***. The report should describe whether the proposed project is consistent with the goals and objectives of approved plans.
- ***Compatibility***. The report should assess whether the project is compatible with surrounding land uses.

As with emails and memoranda, staff reports are usually single-space typed. Subheadings should be bold and/or with a larger font size. Staff reports should not include more than two levels' subheadings. If visual materials are included in the staff report, they should be tightly placed around the text or in-line with the text.

Research Papers

In addition to following the basic framework and structure of writing a *research paper*, such as title, author name, abstract, key words, introduction, literature review, data and methods, analysis, results, conclusions and discussions, and references, research papers may be presented as single-space typed (published or in-print) or double-space typed (draft format, or submitted in an academic setting). Subheadings should be bold and/or with different sizes. If the paper is intended to be published, the authors need to follow the author guideline of the publisher or the target journal to format the paper. Times New Roman is almost always used for research papers. The main text is usually size 12 font size, but the subheadings may have a larger size, especially for level 1 headings (e.g. abstract, key words, introduction, etc.). Maps, charts, and other visual materials should be properly placed in the research paper. Captions for figures should be placed below the figures and captions for tables should be placed above the tables. In-text citations are often required for academic papers, but sometimes footnotes or endnotes are acceptable. References need to following common citation styles, such as the APA (American Psychological Association), MLA (Modern Language Association), or Chicago/Turabian styles. The style guidelines for all these can be easily found through web searches. Commonly APA is used in many research papers.

Research or Planning Reports

For many planning practitioners, *research reports* are often used instead of research papers. However, research reports can be written in house, or by hired planning consultants. In practice, planners in many public planning agencies do not directly produce research reports or planning documents. They often review development plans and permits, conduct public meetings, enforce land use regulations and zoning codes, and ensure compliance with comprehensive plans and other mandates. Oftentimes, skills of preparing research reports are essential for planners in private agencies, even though some private agencies function as a planning department and have their line (e.g. customer service) and/or staff (e.g. advisory or support) members.

Research reports require more efforts in formatting and document design than research papers. Research reports often have a cover page (with title, author name, agency, and some simple graphic design, such as a photo of the local community, and the logo or seal of the community). This is followed by acknowledgements, and/or team members, Table of Contents, List of Tables (if applicable), List of Figures (if applicable), Executive Summary, and then the body of the report. The body of the report should also include subheadings with different levels. If a report is comprehensive and long, multiple levels of subheadings will greatly add readability to the report. For a typical planning document, such as a plan, or a research report, two to three-levels of subheadings oftentimes suffice. Box 6.1 explains how to categorize the subheadings into different levels so that a table of contents can be automatically generated by Microsoft Word, when prompted.

Executive summary is the backbone of the research report, especially when the report is presented to the audience. Executive summary should briefly introduce what the report is about, the methodology of the research, and then focus on summarizing the key findings from the research. If policy recommendations are derived based on the research findings, these policy recommendations need to be concise, to the point, and relevant.

Research reports can be portrait or landscape oriented; although many professional planning documents prefer landscape oriented, with two columns each page. Columns can be inserted through Microsoft Word, or any publishing software, such as Microsoft Publisher and Adobe InDesign. Single-line spacing may appear cluttered, and therefore, 1.15 might be better. Font sizes can be flexible as well, and mostly range from size 10–13. Font types are usually not Times New Roman, but Georgia, Helvetica, PT Sans & PT Serif, Open Sans, Quicksand, Verdana, Rooney, Karla, Roboto, Ubuntu, Lato, and Futura are often used since these are deemed the most readable by Vistaprint.com. Justified alignment (the tab with four equal-length parallel lines under the "Paragraph" tab of Microsoft Word) is often used so that texts will be aligned for both the left and the right sides of the margins.

In-text citations are usually not necessary for research reports and these citations are oftentimes put in the endnotes. However, the reference or works cited list in the endnote section is suggested to follow the common citation styles.

When placing visual materials into the document, authors need to be particularly mindful since these materials may shift texts and formatting around. Tight wrap, or in-line insertion of these visual materials will be sufficient; however, size of the materials needs to be balanced with the main text. Captions of the visual materials need to be added (by right clicking the photos, maps, tables, or charts and choose "insert captions…") so that a list of tables/figures may be automatically created when prompted. Credits should be addressed and permission to use should be granted if the author uses copyrighted materials from third parties.

When necessary, page breaks should be added to create a section (such as a chapter), or a standalone table/chart/maps/photo on one page. Using page breaks will prevent texts from being shifted to the next section when adding or deleting contents.

Infographics

Infographic techniques are combinations of techniques using graphs to demonstrate data or information. Infographics are a collection of charts, graphs, and images with minimal narratives to convey a central message (Fig. 6.1). The purpose of infographics is to deliver the message in a way that is eye-catching and easy to understand. Infographics can be incorporated into a staff report or a regular planning report. They are not common in empirical research papers. Infographics is similar to a mini-poster with key information presented in one place. Various software applications are specifically designed to generate infographics, and many of these

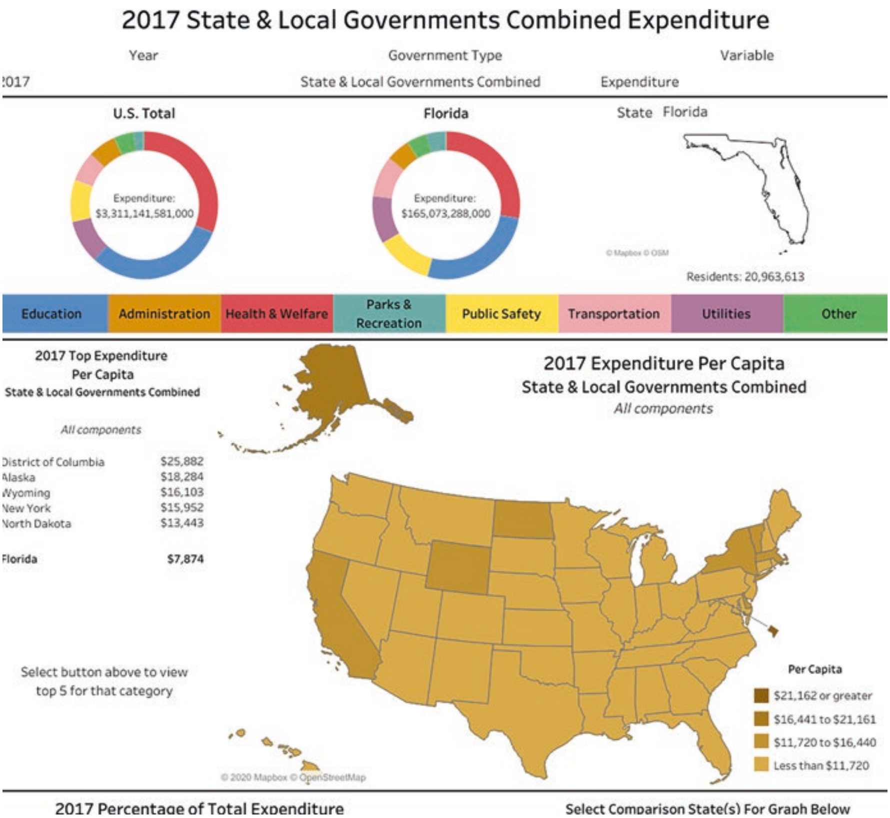


Fig. 6.1 An example of infographics. (Source: U.S. Census Bureau. https://www.census.gov/library/visualizations/interactive/state-local-snapshot.html)

newer applications provide more powerful visual representation tools than the traditional Microsoft Office series. Some of the notable online applications are Canva, Venngage, Piktochart, Visme, Infogram, Vizualize.me, etc. As with many other applications, they offer free versions with limited functionality, and premium paid versions with more versatility and functionality.

A variation of infographics is computerized applications delineating key statistics of data analysis results. Some of these applications are dynamic, taking advantage of animation technology in describing trends or changes longitudinally. Hans Rosling, a renowned Swedish professor in International Health, the chairman of Gapminder Foundation which developed the Trendalyzer software system, has created stunning visualization tools to delineate various statistics.

Proofreading

After completing any writing project, it is critical to proofread the end product. This applies to any form of writing, regardless whether it is an email, a memorandum, staff report, or a research paper. Proofreading will detect any organization, structural, grammatical, and spelling mistakes, given that the writer has solid foundations on grammar and writing.

Proofreading should focus on three aspects of the writing: ***Content***, ***Format***, and ***Mechanics*** (Fry & Kress, 2006, 254). Adapted from Fry and Kress (2006), contents include checking for:

Sticking to the topic and consistent point of view
Good and sufficient sources for information
Sound organization
Check for facts
Use of appropriate visual materials
Tailor to target audience
Sufficient details and description
Format includes:
Appropriate title
Use quotations correctly and sparely
Proper use of headings and subheadings
Legible labels and titles for visual materials
Include a list of references
Page numbers are properly placed
Cover page includes title, author names, and date
Mechanics include checking for:
Proper uses of complete sentences
Consistent verb tenses
Spelling
Punctuations
Capital letters, possessive format, and other writing conventions
Consistent paragraph indentations, line spacing, font size and font types

Box 6.1. Insert Table of Contents, List of Tables, or List of Figures in Microsoft Word

Microsoft Word has been the most popular word processing software since the widely use of technologies in writing and preparing presentation materials. When writing a research report, formatting is one of the key elements that the author needs to pay close attention to. For example, insert page breaks to separate different subsections of the report, cover page, acknowledgment, table of contents, list of tables, list of figures, executive summary, introduction, different subsections under the main text, references, appendices, etc. Page numbers should not start from the cover page, and ideally preamble sections, such as acknowledgments, table of contents, list of tables/figures, should have a separate page sequence, e.g. using Latin numbers, compared to starting the executive summary, or introduction (when the executive summary is placed right after the cover page). Figures can be wrapped or in line with text by right-clicking the figures and choose "Wrap Text" options. Paragraphs can be set as left, right, or both sides aligned, or first line suspended by a few characters or hanging. Line spacing is also set under the "Paragraph" tab of the "HOME" menu. Margins and page orientations can be set through the "PAGE LAYOUT" menu. Overall, one can play with all the tabs and menus in the software to get an idea about what each entails.

In a report, Table of Contents (TOC) and List of Tables/Figures should be automatically created so that any update made on the entire document will be reflected easily in an updated TOC and List of Tables/Figures.

When writing, the author needs to include subheadings for each subsection of the report. After completing the draft report, the author inserts a page break to hold the place for a Table of Contents (TOC). Then all the subheadings in the draft need to be assigned to a style under the "styles" table of the "HOME" menu. "Heading 1" should be used as the first-level heading, and "Heading 2" and "Heading 3" are the second and third-level headings. Any headings more than three levels are not necessary. Most of the table of contents usually only include two levels of headings. The author may change the font color and size if he or she does not like the default setting. After assigning heading levels to all the important headings, go back to the page where the TOC will be inserted. Then click "REFERENCES" and find "Table of Contents" with the dropdown arrow. Choose a style and confirm the selection. A TOC will then be created automatically. Each time a document has been updated which might affect the change in page numbers, the TOC can be updated by clicking it and choose "Update Table…".

To create a List of Tables/Figures, the author needs to add captions to the tables or graphs by right-clicking a table or graph, choose "Insert Caption…". Use Figure or Table to delineate the caption for a table or a figure. Tables are in tabular format with numeric values in each cell. Figures include charts, graphs, photos, maps, and other visual materials. Once all the tables and figures have captions, go back to where the List of Tables/Figures will be inserted. Then go to "REFERENCES", choose "Insert Table of Figures". Make sure change the "Caption label" to "Table" if it is a list of tables, or "Figure" if it is a list of figures. Click "OK" and the list will be created. The list can be easily updated if page numbers have shifted due to revisions.

PowerPoint Presentation

With the advancement of technology, presentation often relies on various software to deliver the intended content. Compared to the traditional whiteboard/blackboard technique, presentation software provides much more flexibility and a broader functionality when handling visual materials, animation, and hyperlinks to internet information. Although there are various presentation software, open or paid, the most commonly used is Microsoft PowerPoint. Principles of crafting presentation files are similar, regardless of the platform used to carry out the presentation.

Presentation files should mostly be used to convey summarized information with important details. Therefore, the biggest pitfall to avoid is to directly copy and paste the content from a Word document into the presentation file. Bullet points are the most commonly used format when adding texts to the presentation. The bullet points should be concise with short sentences. Each slide usually should not have more than five bullet points. Long paragraphs are not easily processed when being presented, due to the time limit of presentations. Too much texts in the presentation often makes the presentation boring and not to the point. Depending on the nature of the content to be presented, oftentimes a balanced presentation of texts and visual materials will greatly enhance the effectiveness of the presentations. This is particularly true for presenting the information in urban and regional planning, since maps, charts, and tables are often used in the field to convey planning information.

As in a report or a paper, citation should be used in the presentation as well. Data sources, citations, and graphic credits should be used in the presentation, unless all these are the original work by the author. As in any written documents, copy right needs to be respected. Background design and the use of visual materials should be professional with neutral color palettes. Visual materials should be highly relevant to the content of the presentation and be professional. Animation and any additional formats should be minimally used so that they do not become distractive. Presenting the content will be covered in the public speaking skills section.

Public Speaking Skills

In the discipline of urban and regional planning *public speaking skills* are nearly equally important as technical writing skills. Workplace presentation is often required during meetings, and presenting a planning recommendation to the planning commission and the general public is often the norm, regardless whether one works in the private or public sector. Additionally, planners are often actively engaged in public participation through workshops, focus groups, interviews, and public meetings. In all these circumstances public speaking and facilitation skills are critical to the success of these venues. Few of us are naturally born public speakers. Therefore, learned skills and large amounts of practice incorporating these skills makes one a better speaker over time.

The process of engaging in public speaking or facilitating involves ***preparation***, ***execution***, and ***evaluation***. Listing the presentation key points is a critical step in preparation. Depending on the nature of the meeting or workshop, the speaker may need to notify the stakeholders through websites, social media, and news outlet. Sometimes speakers will be the facilitators and planning and preparation are therefore critical to ensure the success of the meetings or workshops. During the meetings, the key to success is how to effectively engage the audience throughout the entire meeting. Disinterests from the audience often signals the inadequacies of the speakers or the facilitators in capturing their attention. Disinterests have more negative impact on the outcome of the meetings in public meetings than in professional or academic conferences. Therefore, engaging the audience is critical during the execution of meetings or workshops. Towards the end of the meeting or afterwards, evaluation of the meeting effectiveness will help the speakers or facilitators objectively assess what have been effective and what may be lessons to learn for any future meetings or workshops. Public speaking and facilitation skills cannot be obtained by simply reading tips and pitfalls to avoid; however, the following section hopes to provide some fundamental guiding principles for the reader to understand the specific presentation skills, facilitation skills, and storytelling skills during the entire process of a meeting or workshop.

Engaging the Audience

There are a few fundamental presentation skills: *clear articulation*, *actively engaging the audience*, and *adjust based on the reaction of the audience*. Clear articulation indicates that the speech or presentation needs to be coherent, logical, and with minimal interruptions of thoughts. Voices need to be clear with appropriate volume, pace needs to be not too slow, nor too fast, and verbal clutches such as “um”, “you know”, and “you know what I mean”, etc. should be kept to a minimum. Keeping proper eye contact and shifting eye contact among the audience is also very important to show the speaker is confident and the speaker is genuine in connecting with the audience. Presentations should also avoid too much dry content to keep the audience's interests. Stage freight should be positively channeled so that it does not affect the presentation negatively. Take deep breaths prior to the presentation. Talk slowly and allow yourself time to gather thoughts.

If possible, the presenter should actively engage with the audience by tailoring messages to motivate audiences. It is critical to develop talking points from messages by considering audience perceptions and what information they need and want. Relevant, important, and beneficial information for the audience will catch their attention better. The first 5 minutes of public speaking often indicates whether the speaker will lose or retain the audience' interests. Actively checking with the audience during the presentation will help prolong the attention span. When presenting, proper eye contact also helps the speaker scan the reactions of the audience through facial expression and other gestures. The speaker then can use these clues to adjust the presentation.

To achieve the goals of clear, coherent, and interesting presentations, the speaker needs to plan ahead, sometimes rehearse to get familiar with the content and the timing so that he or she does not extend beyond allotted time. Speakers should avoid excessive hand gestures and may sometimes use humor to lighten up the presentation. However, if you are not used to using humors in public speaking, do not use it since sometimes improperly used humor may backfire.

Visual materials, *computerized simulations*, and *stories* are effective tools in engaging the audience. Charts and graphs should be large enough to be seen and information in each chart or graph should not be crowded. Visual aids should be used properly without distracting the audience (Dandekar, 2003). PowerPoint slides should not have too much texts on one slide and the text size should be easy to read. Depending on the type of audience, planning acronyms, such as TIF (Tax Incremental Financing), ADUs (Accessory Dwelling Units), TDRs (Transfer of Development Rights), should be carefully used.

Oftentimes elected officials are the audience of a meeting or workshop. When presenting written reports to them always remember to be concise, to the point, and use interesting, nonthreatening, easy-to-understand supporting facts tailored to these officials. When communicating with elected officials, American Planning Association listed seven key points for success (APA, 2006):

- Understand the political context
- Know the views of the elected officials and whom they represent
- Form common ground with elected officials
- Develop relationships and build trust
- Tailor messages to specific audiences and situations
- Select the right messenger and deliver the message effectively
- Use repetition and persistence

Overall, presentation skills need practice, and presentation need to be prepared and delivered. When making preparations and presentations, the presenter needs to follow sets of guidelines to ensure success. Figure 6.2 shows some of the key points of presentation (Storey, 2003).

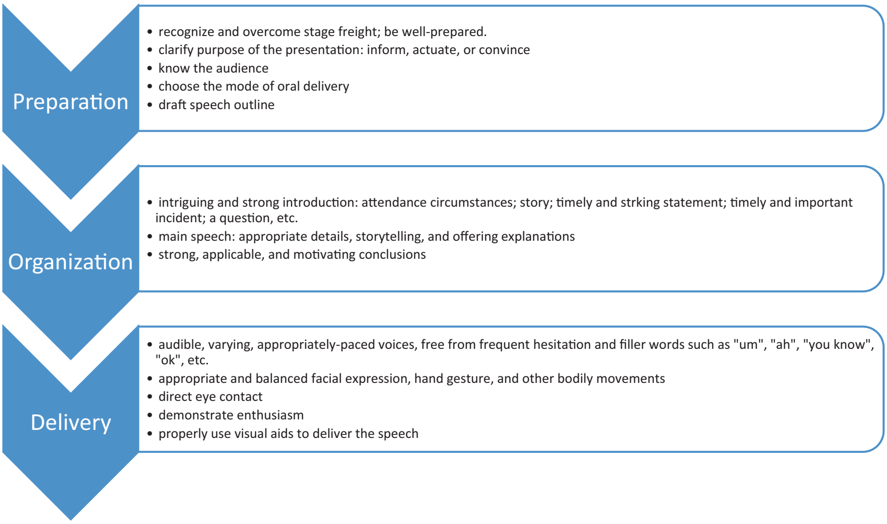


Fig. 6.2 Presentation process. (Adapted from Storey (2003))

Facilitation Skills

Public meetings, professional or academic conferences, and public workshops are various ways to disseminate research findings. Oftentimes a planner or a planning researcher needs to facilitate these meetings. Although the skills of public speaking can be applied in these meetings, being a facilitator requires organizational skills to ensure the success of these meetings.

Organizing the meetings is the first step of facilitating. Preparations need to be done prior to the meetings. The facilitator needs to plan the agendas included in the meeting and prepare all the related materials or handout. The facilitator then needs to ***practice*** or rehearse how to start the meeting, help the audience stay engaged throughout the meeting, and conclude the meeting. The key talk points to capture the audience's attention needs to be carefully crafted. The following list includes key messages that a facilitator needs to address before the meeting (APA, 2006):

- Core planning messages
- Specific issues to be discussed
- Define segments and tapestry of the audience
- Intended and anticipated outcomes
- Talk points based on core planning messages and different audience groups
- List supporting facts to illustrate core planning messages and specific issues
- List visual aids, personal stories, and testimonials to assist in facilitating the meeting

Use technology, such as PowerPoint Presentation, simulation tools, and virtual reality tools to facilitate the meetings. Appropriate use of visual materials will also add appeal to the audience, and therefore attracting more attention from the audience. An important role of a facilitator is to manage time efficiently so that the meeting can be conducted in a reasonable time limit without sacrificing content and quality.

Conflicts arise in meetings, particularly public meetings and workshops. A facilitator should maintain a neutral and calm position to make sure the conflicts do not go out of control.

Storytelling Skills

In recently years, storytelling has gained its popularity in public engagement, due to the relevance of stories to the audience. Stories are also more interesting compared to many other types of communication methods. Storytelling is a type of social and cultural activity sharing stories and information. When presenting research findings, storytelling may be incorporated into the messages to increase engagement of the audience. Storytelling involves three key players, storyteller, audience, and the story, which is defined as the storytelling triangle (Fig. 6.3).

When telling stories, the storyteller needs to pay attention to the following elements of a good story (Passi, 2019):

- ***Plot*** is the most important part of a story. Plots need to be engaging, relevant, and with key details supporting the central message.
- Well-defined ***character*** that the audience can relate to or trust.
- ***Setting***: where and when the story takes place.
- ***Dialog***: how the stories connect with the storyteller and/or the audience.
- ***Melody***: adding music corresponding to the tone and message of the story will add appeal to the story. However, this may not apply when simply using a story to accentuate the presentation.
- ***Deco***: visual materials related to the story.
- Unexpected yet meaningful ***spectacle***. However, spectacle should not be allowed to control the story; instead, the plot should be always strong enough to carry the story.

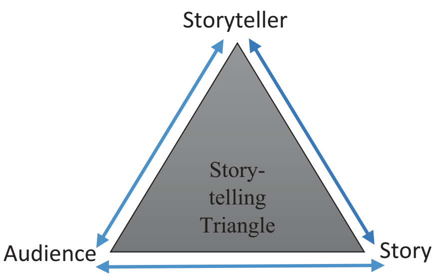


Fig. 6.3 Storytelling triangle

Fig. 6.4 Rules of storytelling. (Adapted from Peters (2018))

Peters (2018) summarized the following six rules of great storytelling (Fig. 6.4), stressing clarity, relevancy, simplicity, focus, and interesting characters.

References

APA (American Planning Association). (2006). *Planners' communication guide: Strategies, examples, and tools for everyday practice*. American Planning Association Press.

Dandekar, H. C. (Ed.). (2003). *The planner's use of information*. Planners Press, American Planning Association (APA).

Fry, E. B., & Kress, J. E. (2006). *The reading teacher's book of lists* (5th ed.). San Francisco.

Greenlee, A. J., Edwards, M., & Anthony, J. (2015). Planning skills: An examination of supply and local government demand. *Journal of Planning Education and Research, 35*(2), 161–173.

Meck, S., & Morris, M. (2004). *Formatting and writing the staff report. Zoning Practice*. American Planning Association.

Miller, E. V. (2019). Assessing the preparation of undergraduate planners for the demands of entry-level planning positions. *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X19873129

Passi, A. (2019). *Elements of good story telling*. UX planet. Retrieved on Sept. 20, 2020 from: https://uxplanet.org/storytelling-elements-31b2a6a7e373.

Peters, B. G. (2018). *6 rules of great storytelling* (as told by Pixar). Retrieved on 20 Sept 2020 from: https://medium.com/@Brian_G_Peters/6-rules-of-great-storytelling-as-told-by-pixar-fcc6ae225f50

Storey, A. W. (2003). Speaking skills for presentation. In H. C. Dandekar (Ed.), *The planner's use of information* (2nd ed.). Planners Press, American Planning Association (APA).

Swift, S. (2011). Getting the most out of staff reports. American Planning Association, *PAS QuickNotes*.

Wyeth, S., & Johnson, L. (2021). Communicating urban research. In S. Baum (Ed.), *Methods in urban analysis, Cities Research Series*. Springer Nature Singapore Pte Ltd.. https://doi.org/10.1007/978-981-16-1677-8_11

Chapter 7
Statistical Analysis

To plan for a community, a planner first needs to understand the community. This session introduces basic tools of statistics in relation to urban issues. Merriam-Webster dictionary defines statistics as "a branch of mathematics dealing with collection, analysis, interpretation, and presentation of masses of numerical data." This may sound challenging. However, everyone is conducting statistical analysis every day. In the morning, you drive from your home to your school or office, and you may avoid the highway that goes through the downtown. Why? You have been stuck in traffic at those locations many times in the past. You have collected your own traffic condition data. You brain does the calculation and informs you that it is very likely to have traffic again at those locations, even though you did not get any statistical training. To conduct statistical analysis, there have to be repeated activities that can be translated into a specific variable. For example, you have daily home-to-work commute, and you accumulate traffic condition data for your local streets. A planning agency may survey local residents about their neighborhood satisfaction level. This is a repeated activity over a group of survey respondents. The Census Bureau provides the same demographic and economic data for all block groups. This is a set of repeated activities across geographies. With data ready, statistical skills (e.g., descriptive analysis, correlation analysis, regression analysis) will help an analyst better uncover patterns and relations. This chapter covers the basics of statistics, including summary statistics, the probability theory, hypothesis testing, simple linear regression, and spatial statistics.

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_7].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_7

Summary Statistics

The increasing amount of urban data available at small geographies, such as blocks and block groups, presents challenges and opportunities to planners. Many times, a planner needs to provide the public a concise update about a city. What is the population composition and is the distribution even across the city? What is the current income level and income disparity? What is the housing market condition? Summary statistics provide tools to address such questions.

Summary statistics enable planners to present a set of data with a common theme with several representative statistics. For instance, a planner needs to present household income information to the public. She gathers income information for the 750 block groups within the city. Directly presenting these records does not provide any meaningful information. Instead, the planner can present the summary statistics, which typically includes number of observations, mean, median, modes, range, and standard deviation. The planner may also present a histogram that shows the distribution of household income.

Measures of Location

A measure of location summarizes a set of numbers into a typical value, or a number that represents the overall level of this group of numbers.

Mean is the average of a set of numbers. It is calculated by adding up all the numbers and dividing it by the number of observations in the dataset. The traditional notation for a sample mean is $\overline{x}$, and the notation for a population mean is μ.

It can be problematic to use mean to study urban issues, because distribution is not even in an urban environment. For example, income distribution is not even. If a planner uses mean to measure the income level of a community where Jeff Bezos, the richest person in the world, lives, the extremely high income of Jeff Bezos dominates the mean value. As a result, this mean value does not represent the income level of most households living in this community. In this case, median is more reliable.

Median is a value sitting in the center of a set of data. For instance, 2017 median household income in the U.S. was $61,372. This indicates that in 2017, half of U.S. households had income more than $61,372, and the other half less than $61,372. To find the median of a set of numbers, sort the number from the lowest to the highest. Find the value that is in the middle. If there is an even number of observations in the dataset, the median is the average of the middle two values.

For example, a neighborhood has 45 households. The assessed housing values are:

269,900	170,570	221,220	226,110	182,140	246,000	185,000	258,890	257,530
255,000	249,100	226,000	243,000	187,590	165,260	250,000	264,730	225,000
230,840	220,600	266,560	199,500	268,000	181,500	207,000	221,070	224,420
212,000	259,770	261,000	214,000	201,850	194,800	209,000	209,310	234,230
214,950	229,830	255,000	177,930	242,000	259,240	190,830	250,000	231,220

What are the average and median housing values? The total number of housing units is 45. The total housing value is to add up all the numbers: $10,149,490. The average housing value is $10,149,490/45 = $225,544. To get the median value, the first step is to sort the values from the smallest to the largest. There are 45 values. The middle one is the 23nd value. The 23nd value is $226,000. The median housing value is $226,000.

Modes describe peaks of the frequency distribution of a data set. The first mode is the value that appears the most frequently. The second mode is the value that appears the second most frequently. Similarly, one can define the third, fourth, fifth, …, modes, if there are any. For a set of discrete data, mode can be easily identified by constructing a frequency table.

Frequency Table

A ***frequency table*** is a useful tool for analyzing qualitative and discrete data. It presents numbers of observations by classes. Classes must be mutually exclusive and collectively exhaustive of each other that means that every observation must belong, and only belongs, to one class. Table 7.1 presents an example. The Louisville Metro government conducted a survey to support their comprehensive plan making. One question asks how often a respondent walks half mile or more. The second column of Table 7.1 presents frequency information, i.e. how many people for each type of walking activity category. The third column presents relative frequency information, i.e. percent distributions. The first mode of this dataset, which accounts for 34% of the total frequency, is never walking half mile or more. The second is nearly every day, which accounts for 18.9% of the total observations.

A ***relative frequency table*** is also called a percent distribution table or a percent density table. Cumulative percentage information can be calculated based on this percent information, as shown in the last column in Table 7.1. For instance, 56.9% of the survey respondents walk once a month or less frequently.

Table 7.1 Frequency and Relative Frequency about Whether Walk Half Mile or More in Jefferson County, Kentucky

	Frequency	Relative frequency	Cumulative percentage
Never	254	34.0%	34.0%
Less than once a month	117	15.7%	49.7%
Once a month	54	7.2%	56.9%
Once a week	134	18.0%	74.9%
Nearly every day	141	18.9%	93.8%
No response	46	6.2%	100.0%
Total	746		

Data source: The Louisville Metro Government

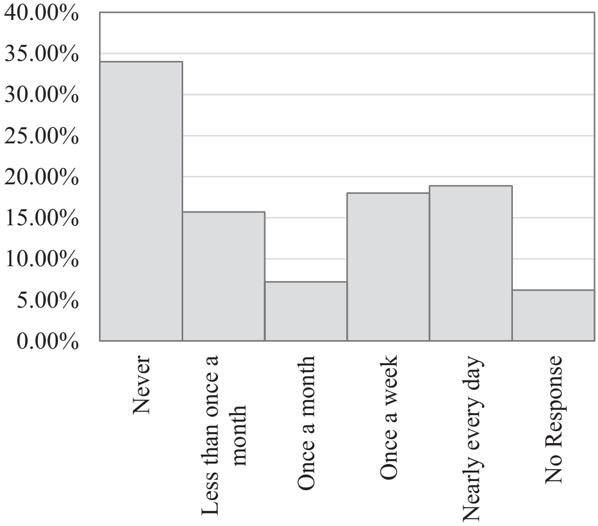


Fig. 7.1 Histogram about whether people walk half hour or more in Jefferson County, KY

A ***histogram*** is a virtual presentation of a frequency table. With frequency or relative frequency table ready, a planner can easily create a histogram in Excel, by inserting a column chart. Figure 7.1 is a histogram showing relative frequency information about people's decision about whether to walk half mile or more.

For a continuous dataset, one needs to first group data into intervals or bins, count the frequency of observations in each bin, and then, create a histogram graph. Figure 7.2 presents the U.S. household income histogram based on 2014–2018 ACS 5-year estimate data.

Measures of Dispersion

Mean and median represent the overall location of a set of data. For instance, according to 2014–18 American Community Survey five-year average, the median household income is $82,459 for New York County, NY, where Manhattan is located at. This is much higher than the national figure, $60,293. A public officer may draw a conclusion that households in New York County are wealthier. However, to fully understand the community conditions related to wealth, it is important to explore differences (dispersion) among New York County households.

A ***range*** measures the dispersion size of a set of data. It is the difference between the largest value and the smallest value, as

$$\text{Range} = \text{Largest value} - \text{Smallest value}.$$

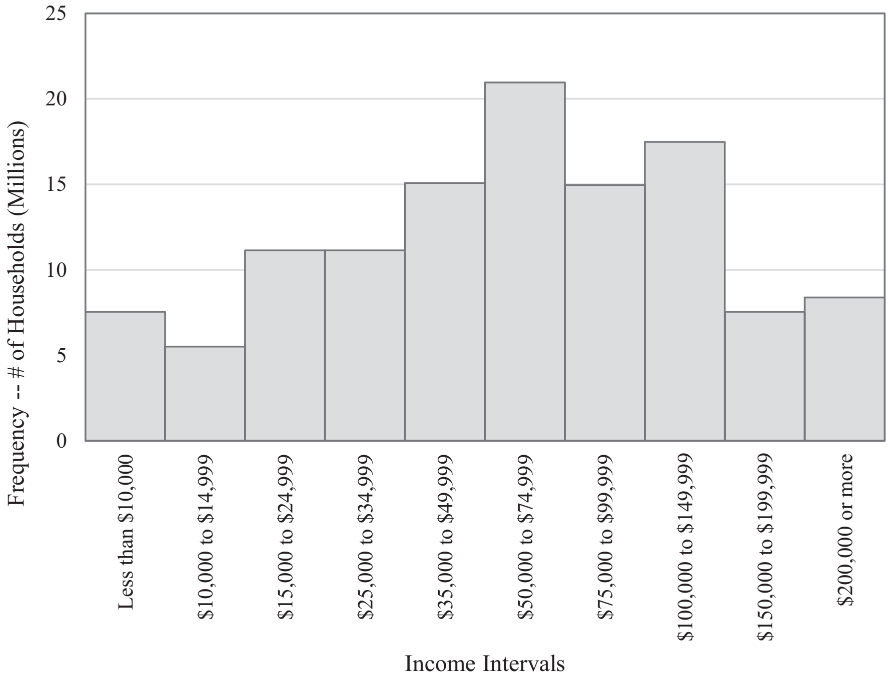
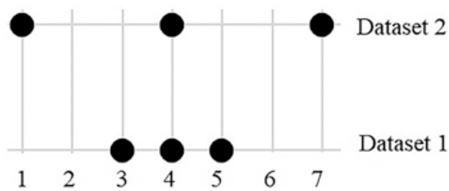


Fig. 7.2 Histogram of U.S. household income (2018 ACS 5-year estimates)

Fig. 7.3 Two datasets



A larger range indicates a higher level of dispersion within the data. Planners and administrators should interpret the range within their contexts. For instance, the range of housing value in a mixed-use neighborhood may be larger than the range of a traditional single-family neighborhood. This higher range suggests a higher level of diversity in household income. Another example is that for a city, block group level percent of black population varies between 0% to 99%. This large range indicates a concentration of black population in several blocks, i.e. a lack of diversity in population by race across the city.

The ***variance*** and the ***standard deviation*** measure the average squared distance and distance between a set of numbers. Figure 7.3 presents two datasets. Dataset 1 consists of 3, 4, and 5, with the mean and median both 4. Dataset 2 takes the values of 1, 4, and 7, with the same mean and median 4. Sample 2 is obviously more dispersed than sample 1.

We can use these two examples in Fig. 7.3 to illustrate the concepts of sample variance and sample standard deviation. The tradition notation of a sample variance is s^2 . It measures the level of spread between numbers within a dataset and can be calculated as:

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}$$

where x_i is the ith number of the dateset, n is the number of observations of the dataset, and $\bar{X}$ is the mean of this dataset and is calculated as $\sum_{n}^{i=1} x_i / n$. The notation for sample standard deviation is s, which is defined as:

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}}$$

Table 7.2 presents the calculations for sample variance and sample standard deviation of the two datasets in Fig. 7.2,

For a population, the population variance and standard deviation are donated as σ^2 and σ, respectively. They are calculated as:

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \mu\right)^2}{n}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \mu\right)^2}{n}}$$

Table 7.2 Sample variance and standard deviation calculations

Dataset 1			Dataset 2		
$\boldsymbol{x_i}$	$\bar{X}$	$\left(x_i - \bar{X}\right)^2$	x_i	$\bar{X}$	$\left(x_i - \bar{X}\right)^2$
3	4	1	1	4	9
4		0	4		0
5		1	7		9
$\sum_{n}^{i=1}\left(x_i - \bar{X}\right)^2 =$		2	18		
$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1} =$		1	9		
$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}} =$		1	3		

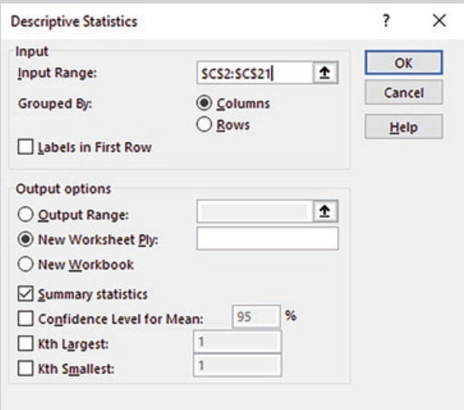
where μ is the population mean and n is the number of observations of the population. Excel provides multiple descriptive statistical analysis tools. One can use Excel functions and directly calculate a specific descriptive statistic, as illustrated in Box 7.1. One also can use the Data Analysis Toolpak and generate descriptive statistics directly (Box 7.2).

Box 7.1 Excel Functions for Descriptive Statistics

- AVERAGE – Calculate mean
- MEDIAN – Calculate median
- MODE – Find the first mode
- STDEV.S – Calculate sample standard deviation
- STDEV.P – Calculation population standard deviation

Box 7.2 Descriptive Statistics in Data Analysis ToolPak

Descriptive Statistics is a tool to quickly calculate summary statistics (e.g. mean, median, mode, range, standard deviation,…). It is available in Microsoft Excel 2016 or newer versions. To use the tool, you must install the Data Analysis ToolPak first. Descriptive Statistics is one of the tools in the Data Analysis ToolPak. The following figure is the Descriptive Statistics dialog box.



Probability

The concept of probability is related to randomness. ***Randomness*** means that (1) the result of an event cannot be determined in advance, and (2) when there is a large number of repetitions of the same event, some order will emerge. Tossing a coin is a classic example. When one tosses a coin, the outcome is either head or tail. No one will be able to tell the outcome of a toss in advance. However, when a coin is continuously tossed many times, one would expect that half of the tosses have heads and the other half tails. It is implausible for planners to conduct repetitive urban experiments to uncover inherent orders. However, in urban studies, repetitions of an event can naturally occur over time or across spaces. For instance, a community planner could study past annual population changes, assess whether there is some sort of pattern, and determine the probability of future growth. A city planner could study whether segregation exists based on analyzing demographic diversities in thousands of neighborhoods in a city.

Basic Concepts

To properly understand probability, we will first introduce the concepts of trial, outcome, and event. A ***trial*** is an action or a series of actions. An ***outcome*** is the result of a trial. An ***event*** is an outcome or a collection of outcomes. The ***sample space*** includes all possible outcomes. ***Probability*** is a numerical measure of the likelihood that an event will occur. Table 5.3 offers two examples to illustrate the above concepts. For example 1, the trial is to toss a coin one time. Having a head and a tail are the two possible outcomes. The interest here focuses on the event of having a head (H), i.e., what is the probability of having a head, which is presented as P(H). For example 2, the trial is to toss a coin twice. There are four possible outcomes. The interest is to study the event of having at least one head for the two tosses. The probability of this event occurs can be presented as P (have at least one head for two tosses) or P(HH, HT, TH).

Relations Between Events and Facts of Probabilities

The probability of event A to occur is presented as P(A) or Prob(A). The probability of any event ranges between 0 and 1, i.e. $0 \le P(A) \le 1$. There are three popular relations between events. If event A consists of all the outcomes from the sample space

Table 7.3 Examples of trials, outcomes, and events

	Trial	Outcome	Event
Example 1	Toss coin one time	Head (H)	Have a head (H)
		Tail (T)	
Example 2	Toss coin two times	First time head, second time head (HH)	Have at least one head for the two tosses (HH, HT, TH)
		First time head, second time tail (HT)	
		First time tail, second time head (TH)	
		First time tail, second time tail (TT)	

that do not belong to event B, and vice versa, A and B are complementary of each other. In this situation,

$$P(A) + P(B) = 1.$$

Two events are mutually exclusive if they do not contain any outcomes in common. In this case, the probability of A or B occurring is the addition of the probability of A occurring and the probability of B occurring, as:

$$P(A\,\text{or}\,B) = P(A) + P(B).$$

If two events are not mutually exclusive, they have a common set of outcomes. For instance, for example 2 in Table 7.3, event A contains outcomes with at least one head for two tosses (HH, HT, and TH). Event B contains outcomes of at least one tail (HT, TH, and TT). There is a common set of outcomes of HT and TH. (HT and TH) is the intersection of the two events, and the probability for this intersection to occur can be presented as P(A and B), i.e., the probability of events A and B happening at the same time. The probability of A or B occurring can be calculated as:

$$P(A\,\text{or}\,B) = P(A) + P(B) - P(A\,\text{and}\,B).$$

If two events are independent, i.e. the incident of one event having no effect on the probability of the other, the probability of A and B occurring is a product of the probabilities of the two events, as:

$$P(A\,\text{and}\,B) = P(A)^{*}\,P(B).$$

Assigning Probabilities

The probability of an event can be estimated by observing the relative frequency with which the event occurs with a large number of trial repetitions. Steps include:

1. Repeat experiment trials;
2. Identify outcomes;
3. Identify the frequency of an event; and
4. Calculate the probability of an event.

Let us have an example of tossing a coin. The experiment is to toss coin twice. A person conducted the experiment 2000 times. Results are

Outcome	(HH)	(HT)	(TH)	(TT)
Frequency	480	512	523	485

What is the probability of at least one head for two tosses? The total number of experiment trials is 2000. The event is having at least one head, which includes HH, HT, and TH. The number of times that the event occurred is 480 + 512 + 523 = 1515. Therefore, the probability of having at least one head for two tosses can be calculated as: P (having at least one head) = 1515/2000 = 0.7575.

Let us have another example about a retailer. A retail store wants to know where their customers come from, such that they can plan for their advertisement activities accordingly. The store surveyed their customers' locations. The resulting data is:

Neighborhood	A	B	C	D	E	F	G	Total
Frequency	1000	1250	780	300	210	170	500	4210

What is the probability of customers coming from neighborhoods A, B, and C? Here the trial is to survey a customer for her residential neighborhood. There are 4210 trials. The event is customers coming from neighborhoods A, B, or C. The number of times that the event occurred is 1000 + 1250 + 780 = 3030. Therefore, the probability can be calculated as: P(customers coming from neighborhoods A, B, and C) = 3030/4210 = 0.7197.

Using Probability for Planning Issues

Probability is a useful tool for planners to identify problems, predict future activities, and make plans accordingly. Many urban activities cannot be predicted in advance because of their inherent randomness, but their patterns can be analyzed with the probability theory. For instance, a transportation planner would like to know residents' commuting routes to properly plan for a future expansion of the existing street network. Nevertheless, a commuter's route choice is not fixed, but the probability of selecting a route in the future can be determined based on past decisions (Bovy, 2009). For an environmental planner, 100-year floodplain is one of the standards that government agencies used for regulating new development. It is a concept about probability. Hydrologists counted flooding frequency at a location for at least 10 years and identified areas with 1% of probability to get flooded every year. These areas are expected to get flooded once every 100 years (Di Baldassarre et al., 2010).

Probability Distributions

Random Variable and Probability Distribution

A variable is a symbol describing some quantity or information. For instance, in Fig. 7.3, we can define dataset 1 as variable Y, which can take a value of 3, 4, or 5. In the planning area, a variable often describes an attribute of entities or processes.

The attribute can vary from one to another. Examples of variables to describe people are age, gender, ancestry, disability, education, employment, and so on. Example of variables for housing units are tenure, number of rooms, year structure built, and so on.

A ***random variable*** is a variable for which the value depends on a random phenomenon. This means that the value is unknown, but there is an inherent probability relation among all possible values. For instance, a person's gender is unknown before conception. However, most newly fertilized eggs have a 50% chance of being a biological male and another 50% chance of being a biological female.

For a random variable X, the ***probability distribution*** associates all outcomes with their corresponding probabilities, which is denoted as $f(x)$ or $P(X = x)$. For a possible outcome, x, the probability is $P(X = x)$. Any probability distribution must satisfy two rules: (1) the probability of a particular outcome takes a value between 0 and 1, and (2) the sum of all probabilities for all possible outcomes must equal 1. A ***cumulative probability distribution*** specifies the probability of an event having an observed value less than or equal to a specific value of x. It is often denoted as $F(x)$ or $P(X \leq x)$.

Expected Value

A planning agency is interested in knowing the public's satisfaction level about its zoning service. A survey indicates the probability distribution in Table 7.4. The agency needs to know the satisfaction level rather than the distribution. Expected value addresses this issue.

The ***expected value*** of a random variable is derived based on all possible values of this random variable and the probabilities of these possible values. It is denoted as E(X), and can be calculated as:

$$E(X) = \sum_i x_i P(X = x_i)$$

where x_i is the ith possible value of random variable X and $P(X = x_i)$ is the probability of variable X to take the value of x_i.For the planning agency (Table 7.4), the expected value of the public's satisfaction level is 5*0.65 + 4*

Table 7.4 Probability Distribution of Zoning Service Satisfaction Level

Satisfaction level		Probability
Very satisfied	5	65%
Satisfied	4	15%
Neutral	3	5%
Not satisfied	2	10%
Very disappointed	1	5%

0.15 + 3*0.05 + 2*0.1 + 1* 0.05 = 4.25. This expected satisfaction level is between satisfied (4) and very satisfied (5).

Probability Mass Function

A ***probability mass function*** is a probability distribution of a discrete random variables, i.e. of a trial with a limited number of outcomes. It associates each possible value of this random variable with its corresponding probability.

The ***discrete uniform distribution*** has a finite number of outcomes and each outcome has an equal probability to occur. Suppose there are n possible outcomes. The probability of a specific outcome is:

$$f(x) = \frac{1}{n}$$

where n is the number of possible outcomes. The discrete uniform distribution is one of the most used probability distributions in urban areas. For instance, there are 100 neighborhoods in a city and a planner needs to analyze their demand for a public service. If there is no additional information available, the planner may assume that every neighborhood has an equal probability, 1/100, to demand for the service.

The ***binomial distribution*** is a discrete distribution, consisting of n independent Bernoulli trials. A ***Bernoulli trial*** has two outcomes, fail or success. The probability of success for a trial, denoted as p, is the same for all trials. The binomial probability mass function describes the probability of k times of success out of n trials, given the probability p of success for each trial, as:

$$f(k,n,p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

This probability can be found in Excel by using function BINOM.DIST.

The ***geometric distribution*** is constructed based on an independent infinite number of Bernoulli trials. Each trail has a probability of success p and a probability of failure 1-p. The interest is to explore the probability of the first success occurring at the kth trial, which is:

$$f(k,p) = (1-p)^{k-1}\, p$$

The ***negative binomial distribution*** is a generalization of the geometric distribution. It examines the probability of having x failures prior to the kth success in independent Bernoulli trials with a probability of success p. The probability of a geometric distribution or a negative binomial distribution can be found in Excel by using function NEGBINOM.DIST.

The ***Poisson distribution*** is to model the probability of the number of events (k) occurring within a given time or space interval, as:

$$f(k,\lambda) = \frac{\lambda^k}{k!} e^{-\lambda},$$

where λ is the average number of events in the given time or space interval. This is a useful probability distribution for studying urban traffic (Shankar et al., 2003). The probability of the Poisson distribution can be obtained in Excel by using function Poisson.DIST.-

Probability Density Function

For a continuous variable, there are an infinite number of possible outcomes. Therefore, it is implausible to find the probability of a specific value occurring. People study probability density instead of probability for a continuous variable and use a probability density function to present probability density for all possible values. For a specific value x, the probability density presents the probability of this variable taking a value within an extremely small interval around x.

The ***normal distribution*** is the most important probability distribution in statistics because the Central Limit Theorem, which will be introduced next, justifies the use of this distribution in many applications. The normal distribution has a symmetric and bell-shaped curve. A specific normal distribution depends on the mean (μ) and the standard deviation (σ), and is denoted as N(μ, σ). Figure 7.4 presents two

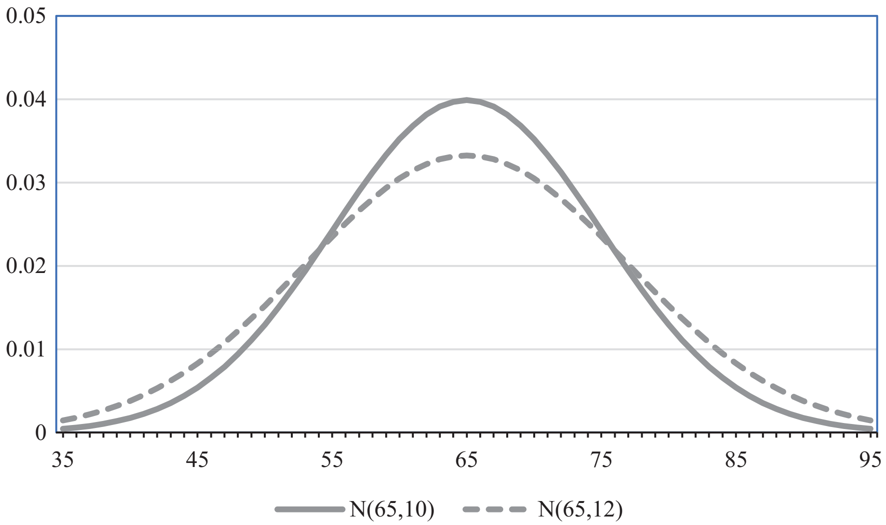


Fig. 7.4 Normal density function

examples. One follows the normal distribution of mean 65 and standard deviation 10, i.e., N(65,10), and the other follows the normal distribution of mean 65 and standard deviation 12, i.e., N(65,12). Both curves center around their means and are asymptotic, which means that the curves come closer to the horizontal axis, but never touch the axis.

Area under a normal distribution curve presents probability density information. The total area under the curve is 1, representing the probability of 100%. The probability of the variable falls in a range corresponds to the area under the curve for this range. For instance, in Fig. 7.5, the area under the curve between 55 and 75 is 0.683, and therefore, the probability of this variable taking a value between 55 and 75 is 0.683, i.e., $P(55 \leq X \leq 65) = 0.683$. Similarly, the probability of this variable taking a value greater than 75 is 0.1585. The probability of this variable taking a value less than 55 is 0.1585.

In addition to the symmetric and the asymptotic feature, a normal distribution has another two characteristics. The mean, median, and mode are equal. The area under the curve within one standard deviation of the mean is 0.683, the area within two standard deviations is 0.955, and the area within three standard deviations is 0.997.

The ***standard normal distribution*** is a normal distribution with mean 0 and standard deviation 1. It is traditionally denoted as $Z \sim N(0,1)$. It is a common practice to convert a normal distribution to the standard normal distribution, such that probability information can be easily found in a standard normal distribution table (Appendix 1). To convert a point (x) from a normal distribution $N(\mu, \sigma)$ to the corresponding point (z) from the standard normal distribution, one needs to calculate the z-score or z-statistic, as:

$$z = \frac{x - \mu}{\sigma}$$

The ***t-distribution*** is another important distribution which enables people to estimate population parameters based on sample data and make statistical inferences. The t-distribution has a symmetric and bell-shaped curve that is centered at 0, just like a normal distribution. But the t-distribution is more spread out than a normal distribution, as illustrated in Fig. 7.6.

Degrees of freedom (df) is a parameter for a t-distribution. It refers to the maximum number of logically independent values. Degrees of freedom can be calculated based on the sample size, as the size of the data sample minus one. The higher the degree of freedom, the closer the t-distribution is to the standard normal distribution. Appendix 2 presents a t-distribution table.

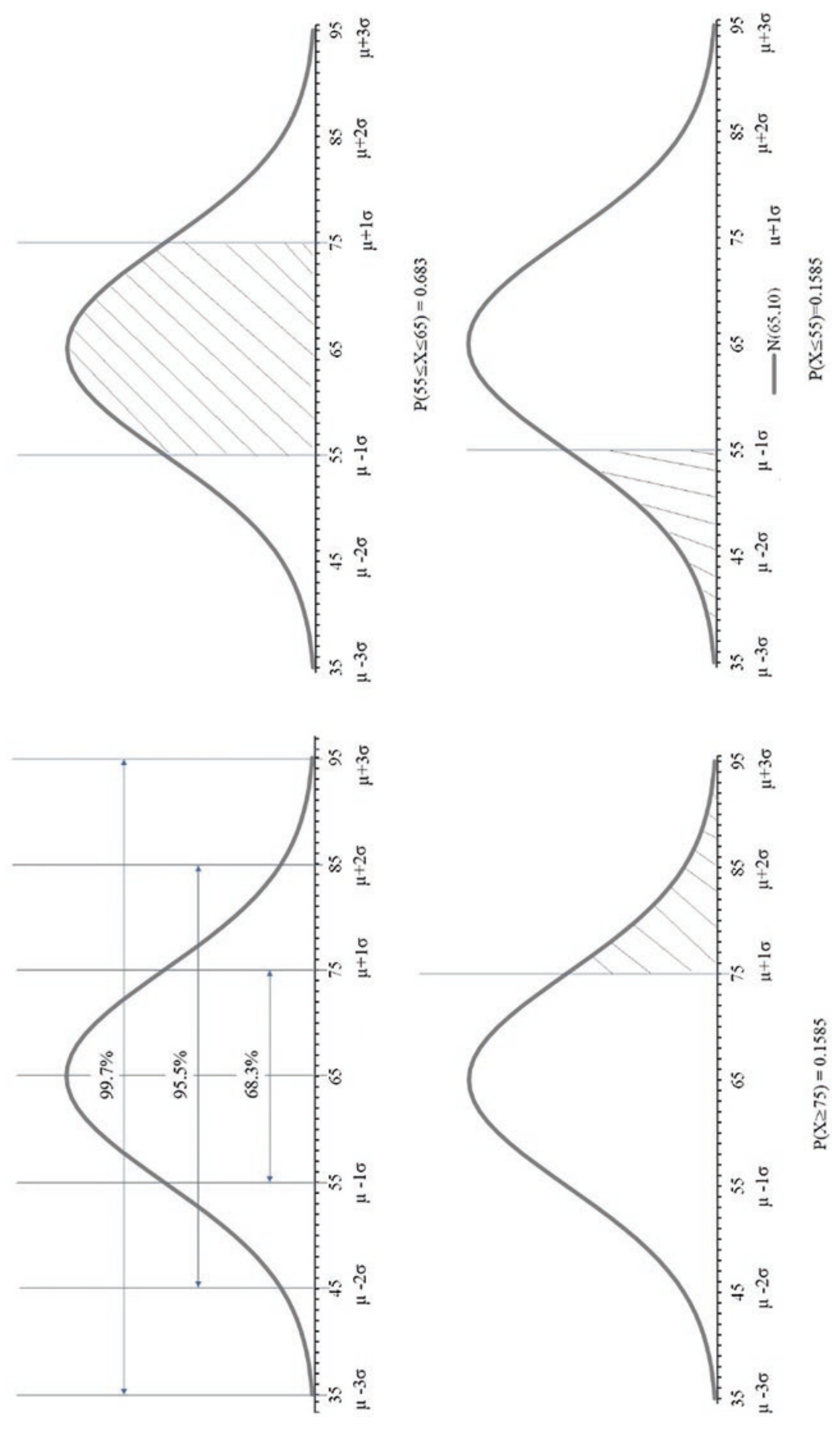


Fig. 7.5 Normal distribution and probability

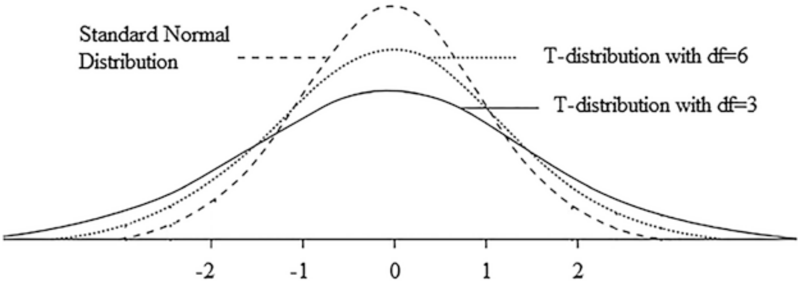


Fig. 7.6 The standard normal distribution and the T-distribution

For finding the probability information for different distributions in Excel, please see Box 7.3.

Box 7.3 Excel Functions Related to Probability Distributions

- BINOM.DIST (S, T, *p*, FALSE) – Return the probability of having exactly S number of success out of T trials, given the probability of success for each trial is *p*.
- BINOM.DIST (S, T, *p*, TRUE) – Return the cumulative probability of having at most S number of success out of T trials, given the probability of success for each trial is *p*.
- NEGBINOM.DIST (F, S, *p*, FALSE) – Return the probability of having F number of failures before the Sth success, given the probability of success for each trial is *p*.
- NEGBINOM.DIST (F, S, *p*, TRUE) – Return the cumulative probability of having at most F number of failures before the Sth success, given the probability of success for each trial is *p*.
- POISSON.DIST(x, M, FALSE) – Return the probability of having x events given that the average number of events is M.
- POISSON.DIST(x, M, TRUE) – Return the cumulative probability of having at most x events given that the average number of events is M.
- NORM.DIST(x, μ, σ, FALSE) – Return the probability density associated with x in normal distribution with mean μ and standard deviation σ.
- NORM.DIST(x, μ, σ, TRUE) – Return the cumulative probability associated with x in normal distribution with mean μ and standard deviation σ.
- NORM.S.DIST (z, FALSE) – Return the probability associated with z in a standard normal distribution.
- NORM.S.DIST (z, TRUE) – Return the cumulative probability associated with z in a standard normal distribution.
- NORM.S.INV(p) – Return the z value for a probability p in a standard normal distribution.

- T.DIST (x, df, FALSE) – Return the probability density associated with x in a t-distribution with degree of freedom df.
- T.DIST (x, df, TRUE) – Return the cumulative probability associated with x in a t-distribution with degree of freedom df.
- T.INV (p, df) – Return the t value for a probability p in a t-distribution with degree of freedom df.

The Central Limit Theorem

The ***Central Limit Theorem*** enables the study of a population based on samples. Given a population with the mean μ and the standard deviation σ, if there are a sufficiently large random samples drawn independently from the population with replacement, the distribution of the sample means will be normally distributed with mean μ and standard deviation $\sigma / \sqrt{n}$. For instance, in a city with a half million households, if a planner selects 1000 random samples from this population to study their household income and calculate the average household income for each sample, these averages will follow a normal distribution centered around the true average household income of the whole population. When the population standard deviation is unknown, one can estimate it based on the sample standard deviation *s*. In this case, the distribution of the sample means will follow a t-distribution with degree of freedom *n*-1. The theory provides insights into why one can conduct z-tests and t-tests based on samples.

Confidence Interval

When using a sample data of size *n* to study a population with mean μ and standard deviation σ, the Central Limit Theorem addresses that the sample mean follows a normal distribution with mean μ and standard deviation $\sigma / \sqrt{n}$. A ***confidence interval*** can be derived based on this probability distribution. It provides a range of values, with an upper and lower limit around the sample mean, for which one is confident that this interval has captured the population mean. The confidence interval of a population mean can be written as [$\bar{x} - z\frac{\sigma}{\sqrt{n}}$, $\bar{x} + z\frac{\sigma}{\sqrt{n}}$], or $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$, where z determines the probability or the confidence level that this confidence interval will contain the actual population mean. Common confidence levels are 0.9, 0.95, and 0.99, indicating that the probability of a confidence interval to capture the true value is 90%, 95%, and 99%. Associated with this probability values, z values are 1.6449, 1.9600, 2.5758, which can be found in Appendix 1 or in Excel using the NORM.S.INV function.

When population standard deviation is unknown, one can use sample standard deviation (s) to construct the confidence interval, as [$\bar{x} - t\frac{s}{\sqrt{n}}$, $\bar{x} + z\frac{s}{\sqrt{n}}$], or

$\bar{x} \pm t\frac{s}{\sqrt{n}}$, where t determines the confidence level. The actual t value can be found in a t-distribution with degrees of freedom n-1 in Appendix 2. In Excel, T.INV is the function that returns the t-value based on inputs of the desired probability and the degrees of freedom.

American Community Survey datasets provide margin of error information for every variable. Table 7.5 provides an example. These Margin of Errors are 90% confidence intervals. For instance, in terms of total number of housing units, the sample data suggests that at the 90% confidence interval, the true value is in the range of [138,539,906 – 4033, 138,539,906 + 4033].

Hypothesis Test

Hypothesis testing is a method to evaluate the truth of two mutually exclusive statements based on a sample of data. This is a useful method in the public sector because many decisions are made based on assessments of situations. For instance, a university may decide to continue its collaboration with the UPS on the UPS Tuition Assistance Program if the university finds that the program has helped the university in increasing the graduation rate. A planning department may modify its rezoning process if it is found that the current process is not effective. The goal of a hypothesis testing is to verify whether a statement is true.

General Steps of a Hypothesis Test

Most hypothesis tests involve four steps of hypothesis setup, data collection and test-statistic calculation, making decision rules, and drawing a conclusion.

- Hypothesis Setup

A hypothesis testing starts with a research question, which can be translated into a statement and then into hypotheses. For instance, suppose that the current national

Table 7.5 Housing occupancy status in the United States

	Estimate	Margin of error
Total housing units	138,539,906	±4033
Occupied housing units	121,520,180	±153,217
Vacant housing units	17,019,726	±155,617
Homeowner vacancy rate	1.5	±0.1
Rental vacancy rate	6.1	±0.1

Data source: 2018 American Community Survey 1-year Estimate average household income is known as $65,000. A planner would like to know whether the city is wealthier, i.e., whether households in the city are wealthier than the national average. On the optimistic side, she may be interested in proving a statement of that "the average household income in the city is greater than $65,000." A hypothesis consists with two statements that are mutually exclusive and collectively exhaustive. Mutually exclusive means these two statements do not coincide. Collectively exhaustive means that these two statements encompass the entire range of possible outcomes. This implies that if one hypothesis is true, the other must be false. These two hypothesis statements are called null and alternative hypothesis. The tradition is to use the more conservative statement as the null hypothesis, and the other one as the alternative. Therefore, we can develop the following set of hypotheses statements related to the household income example:

H_0: The average household income of the city is not higher than $65,000
H_1: The average household income of the city is higher than $65,000

The null hypothesis (the null) is typically denoted as H_0 and the alternative hypothesis as H_1 or H_a .

- Data Collection and Test Statistic Calculation

The next step is to collect data to test whether the null hypothesis is true. The two hypotheses are complementary. If H_0 is true, then H_1 is false and verse visa. In the city household income example, the planner may collect a sample of household income data (following an appropriate sampling strategy). The sample mean can be calculated accordingly.

The Central Limit Theorem states that given a sufficiently large sample size, this sample mean follows a normal distribution. As for what is a sufficient large sample size, the rule of thumb is to have the sample size larger than or equal to 30. Next one needs to select an appropriate distribution for conducting the test and calculate the specific test statistic. Table 7.6 provides information about how to choose the right method for testing population mean. The details of specific methods will be further explained in the following section.

- Decision Rules

In the previous step, one can calculate a test-statistic based on the available sample data. This test-statistic follows a normal or a student's t distribution. For instance,

Table 7.6 Hypothesis Testing for Population Mean

Sample size	Population follows normal distribution?	Know population standard deviation?	Distribution to use	Test statistics
≥30		Yes	Standard normal distribution	Z-statistic
		No	T Distribution	T-statistic
<30	Yes	Yes	Standard normal distribution	Z-statistic
		No	T Distribution	T-statistic

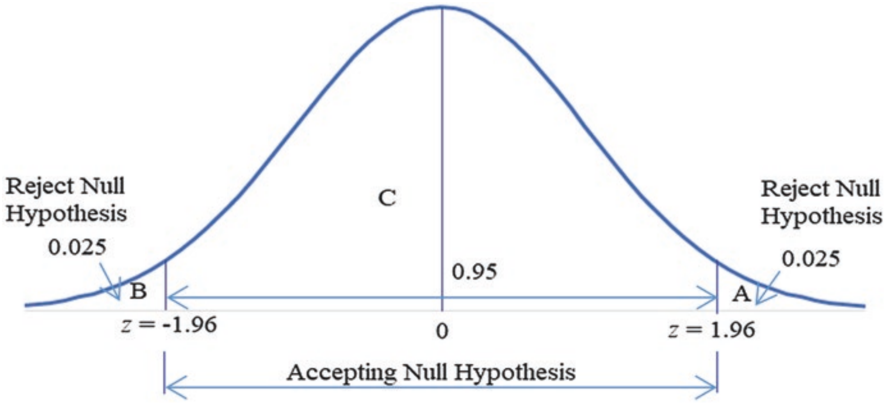


Fig. 7.7 Decision rules

if the test-statistic follows a normal distribution, the calculated value of the z-statistic takes a value on the horizontal axis in Fig. 7.7. A set of decision rules must be made to divide all the possible z-values into two ranges, one for accepting the null hypothesis and another for rejecting the null hypothesis. To simplify the understanding of the decision rule, we use "accepting the null hypothesis" instead of the traditional "fail to reject the null hypothesis" since the null hypothesis is not necessarily true when the data does not support the alternative hypothesis. However, the null hypothesis has not been disproven and therefore most likely to be true.

The standard normal distribution is a normal distribution with mean 0 and standard deviation 1. It is traditionally denoted as Z ~ N(0,1). It is a common practice to convert a normal distribution to the standard normal distribution, such that probability information can be easily found in a standard normal distribution table (Appendix 1). To convert a point (x) from a normal distribution N(μ, σ) to the corresponding point (z) from the standard normal distribution, one needs to calculate the z-score or z-statistics as $z = \frac{x - \mu}{\sigma}$.

A hypothesis test uses a significance level to weight the relation between ranges of accepting and rejecting the null hypothesis. A significance level takes a value between 0 and 1. It corresponds to the probability of the z-statistic or t-statistic to occur in the rejection region. Common choices of significance level are 0.01, 0.05, or 0.1. With a smaller significance level, it is more difficult to reject the null hypothesis and the test is more conservative. In Fig. 7.7, the rejection region includes areas A and B, with the significance level 5% (0.025 + 0.025 = 0.05). The rules are: if the calculated z-statistic falls in the acceptance range, accept the null hypothesis and reject the alternative hypothesis, and if the calculated z-statistic falls in the rejection range, reject the null hypothesis and accept the alternative hypothesis.

- Make a Decision

The final step of a hypothesis test is to assess the calculated z-statistics in relation to the acceptance and rejection regions and make decision to accept or reject the null hypothesis.

One Sample z-Test of Population Mean

When the population standard deviation is known and samples are randomly selected, a *z-test* is appropriate. To conduct the z-test, one needs to set up the hypothesis, convert the sample figure to its corresponding z-statistic, define the acceptance and rejection region for decision rules, and draw a conclusion. We will illustrate the details of a z-test with the following two examples.

In the first example, the planner wants to explore whether children's weight in a distressed community equal their expected weight. Children of age 10 are known to have a mean weight of $\mu = 85$ pounds. A complaint is made that the children living in a distressed community are underfed. To assess this complaint, a random sample of 100 children from this community are collected and found to have a mean weight of 80.94 pounds. It is known that the population standard deviation is 11.6 pounds. Based on the sample, does the average children's weight in this distressed community meet the expectation of 85 pounds? For this scenario of testing for equality, this planner should use a 2-sided z-test. The first step is to set up the hypothesis, as:

H_0: The average children's weight in this community is 85 ($\mu = 85$)
H_1: The average children's weight in this community is not 85 ($\mu \neq 85$)

The next step to is calculate the test-statistic. Here the sample size ($n = 100$) is larger than 30 and the population standard deviation ($\sigma = 11.6$) is known, one should calculate a z-statistic. According to the Central Limit Theorem, the sample mean should follow a normal distribution around population mean ($\mu = 85$) and standard deviation $\sigma / \sqrt{n}$. Therefore, the z-statistic can be calculated as $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{80.94 - 85}{11.6 / \sqrt{100}} = -3.5$.

The third step is to define decision rules. It is common to use 0.05 as the critical value. Now look at the hypothesis, the alternative is about inequality. This indicates a two-sided test, i.e. there are two equal size rejection regions on both sides. The rules are: accept the null hypothesis if the calculated z-statistic is in the range of [−1.96, 1.96]; or reject the null hypothesis if the calculated z-statistic is less than −1.96 or greater than 1.96 (Fig. 7.7).

The last step is to make the decision based on the z-statistic. The calculated z-statistic is −3.5, which is less than −1.96 and falls in the rejection region. Therefore, the decision is to reject the null hypothesis, and accept the alterative. The average weight of 10-year old children in this community is not 85.

The test of equality addresses the question about whether the population mean equals an expected figure. It does not address that the population mean exceeds or less than the expected figure. In the example above, the planner may want to further explore whether children's weight in the distressed community is higher or lower than the expected weight. In this case, a one-sided z-test should be used. The new hypotheses can be setup as:

H_0: The average children's weight in the distressed community is great than or equal to 85 ($\mu \geq 85$)
H_1: The average children's weight is less than 85 ($\mu < 85$)

Next, it is to calculate the test-statistics, as $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{80.94 - 85}{11.6 / \sqrt{100}} = -3.5$.

The third step is to define decision rules, using 0.05 as the critical value. Now look at the hypothesis, the alternative is less than 85. This indicates a one-sided test, as shown in Fig. 7.8. The rejection region is located on the side that is consistent with the direction of the alternative hypothesis. The rules are: accept the null hypothesis if the calculated z-statistic is greater than or equal to-1.645; and reject the null hypothesis if the calculated z-statistic is less than −1.645.

Lastly, one can make the decision. The calculated z-statistic is −3.5, which is less than −1.645 and falls in the rejection region. Therefore, the decision is to reject the null hypothesis, and accept the alterative. The average weight of 10-year old children in this community is less than 85.

One Sample t-Test for Population Mean

Many times, the population standard deviation is unknown. In this situation, one needs to use a t-test. There are one-sided and two-sided t-tests, depending on how the hypotheses are formulated. In this section, let us focus on a two-sided t-test. One sided t-test should follow similar steps as for one sided z-test, as illustrated in the previous section, but with a t-statistic calculated.

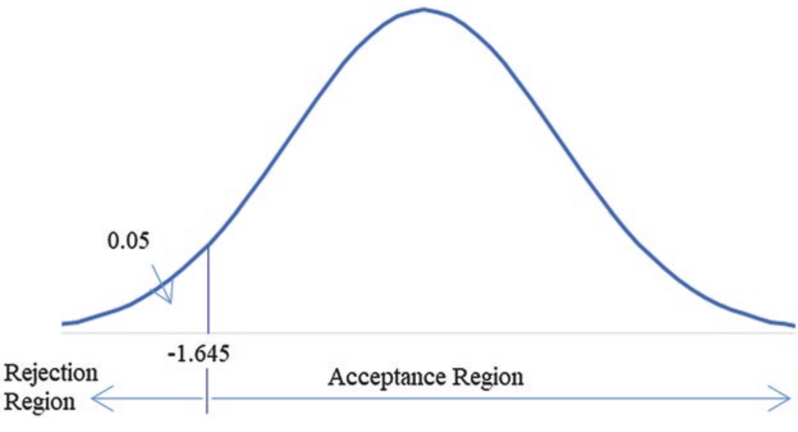


Fig. 7.8 One-sided t-test

With the population standard deviation not available, one can use the sample standard deviation (s) to calculate the t-statistics, as:

$$t = \frac{\overline{x} - \mu}{s / \sqrt{n}}$$

When deciding on the acceptance and the rejection regions, one should use the t distribution with degrees of freedom n-1. If the calculated t-statistics falls is the acceptance region, the decision is to accept the null hypothesis. Otherwise, it is to reject the null hypothesis and accept the alternative.

Let us continue with the example in the previous section. To assess this complaint that children in a distressed community are underfed, a random sample of 100 children from the community are collected and found to have a mean weight of 80.94 pounds, with a sample standard deviation 4 pounds. Does the average children's weight in this community meet the expectation of 85 pounds? The first step is to set up hypotheses, as:

H_0: The average children's weight of this community is 85 ($\mu = 85$)
H_1: The average children's weight of this community is not 85 ($\mu \neq 85$)

Then, calculate the test-statistic. Here the sample size is fairly large (>30) and the population standard deviation is unknown, one should calculate a t-statistic. The sample standard deviation is 4. Given the null hypothesis is true, the population mean (μ) is 85. The t-statistic can be calculated as $t = \frac{\overline{x} - \mu}{s / \sqrt{n}} = \frac{80.94 - 85}{4 / \sqrt{100}} = -10.15$. This t-statistic follows a t distribution with mean 0 and degrees of freedom 99 (number of observations minus 1).

The third step is to define the decision rules. Let us use 0.05 as the significant level. Now look at the hypothesis, the alternative is not equal 85. This indicates a two-sided test, i.e. there are two equal size rejection regions on both sides, as shown in Fig. 7.9. The rules are to accept the null hypothesis if the calculated t-statistic is

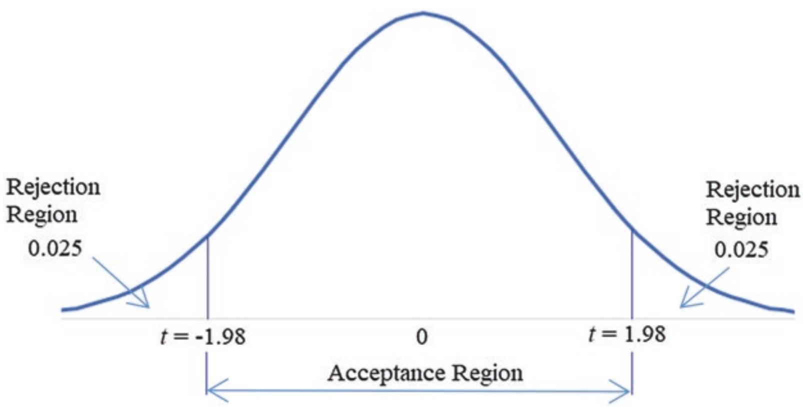


Fig. 7.9 Two-sided t-test

in the range of [−1.98, 1.98], and reject the null hypothesis if the calculated t-statistics is less than −1.98 or greater than 1.98.

The last step is to make the decision. The calculated t-statistic is −10.15, which is less than −1.98 and falls in the rejection region. Therefore, the decision is to reject the null hypothesis, and accept the alterative. The average weight of 10-year-old children in this community is not 85.

Two Sample t-Test for Difference

Often a planner or a public administrator will have two samples (two different groups of data) and would like to assess whether there is a difference. For instance, within the university, is there a difference in salary between males and females? In a Public School System, do magnet schools have higher educational quality than non-magnet schools?

- Paired Two Sample t-Test

A paired t-test compares two population means where the observations in one sample can be paired with the observations in the other sample, for instance, before and after a treatment of the same subject. For every pair of observations, one can calculate the difference. The comparison between these two populations is equivalent to an evaluation of the differences. One can use one-sample t-test to assess the differences.

For example, an instructor would like to evaluate whether the planning method course has helped students improving their analytical skills. To assess this, she gave students a test before they took the class and the same test after they took the class. Table 7.7 presents the grades.

The instructor would like to know whether students' grades are improved after they take the course. The first step is to set up the hypotheses, as

H_0: The grades are not improved, i.e. (Grades after) - (Grades before) ≤0;
H_1: The grades are improved, i.e. (Grades after) - (Grades before) >0;

Next, calculate the test-statistic. Here the sample size is 30 and population standard deviation of the difference is unknown, and we should calculate a t-statistic. The sample mean and the sample standard deviation are 8.4 and 13.5, respectively. Given the null hypothesis, the population mean (μ) is 0. The t-statistic can be calculated as: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{8.4 - 0}{13.5/\sqrt{30}} = 3.41$. This t-statistic follows a t-distribution of mean 0 and degrees of freedom 29, which is the number of observation (30) minus 1.

The third step is to define the decision rules. Let us use 0.05 as the significant level. Now look at the hypothesis, the alternative is greater than 0. This indicates a one-sided test, with the acceptance region on the left side, as shown in Fig. 7.10. The rules are to accept the null hypothesis if the calculated t-statistic is less than 1.6991; otherwise reject the null hypothesis.

Table 7.7 Test grades before and after a planning method course

Student ID	Grades before	Grades after	Grades after - grades before
1	60	79	19
2	70	65	−5
3	55	75	20
4	80	95	15
5	75	74	−1
6	50	85	35
7	40	75	35
8	20	15	−5
9	70	60	−10
10	90	91	1
11	75	90	15
12	85	89	4
13	77	75	−2
14	64	70	6
15	89	91	2
16	90	91	1
17	95	92	−3
18	30	70	40
19	55	60	5
20	56	57	1
21	68	66	−2
22	89	95	6
23	30	50	20
24	70	80	10
25	66	80	14
26	70	60	−10
27	45	75	30
28	90	91	1
29	77	90	13
30	65	62	−3

At last, make the decision. The calculated t-statistic is 3.41. It does not fall into the acceptance region. Therefore, the decision is to reject the null hypothesis and accept the alternative. The grades are improved.

Two-Sample t-Test

In the case of having two groups of unpaired data, if the variances of the two populations are equal, one needs to use Two-Sample T-test with Equal Variance. If the variances do not equal, one needs to use Two-Sample T-test with Unequal Variance.

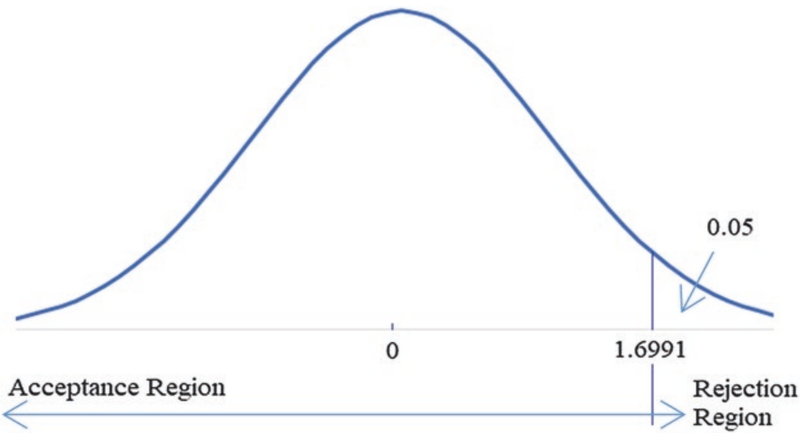


Fig. 7.10 The decision rule –the students' grades case study

- Two-Sample T-test with Equal Variance

Suppose that there are two samples, X and Y, from two populations following normal distributions and with population mean μ_x and μ_y, respectively. The sample size is n for sample X and m for sample Y. Both populations have the same variance. To test the equality between the two population means (μ_x and μ_y), the following t-statistic should be calculated:

$$t = \frac{(\overline{x} - \overline{y})}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

where s_p is the pooled variance. It is calculated based on sample variances s_x and s_y, as:

$$s_p = \frac{(n-1)s_x^2 + (m-1)s_y^2}{m+n-2}$$

This t-statistic follows a t distribution with mean 0 and degrees of freedom $n + m - 2$.

- Two-Sample T-test with Unequal Variance

In the case of unequal variances between the populations, for testing the equality of population means, the t-statistic should be calculated as:

$$t = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

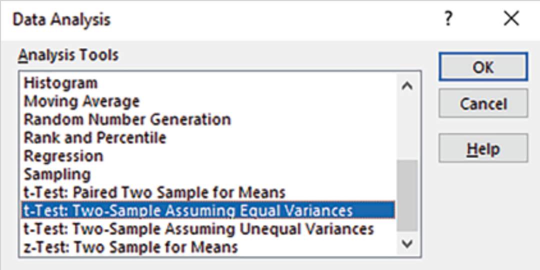
The degree of freedom for the distribution of this statistic is

$$\frac{\left(\frac{s_x^2}{n}+\frac{s_y^2}{m}\right)^2}{\left(\frac{s_x^2}{n}\right)^2/(n-1)+\left(\frac{s_y^2}{m}\right)^2/(m-1)}.$$

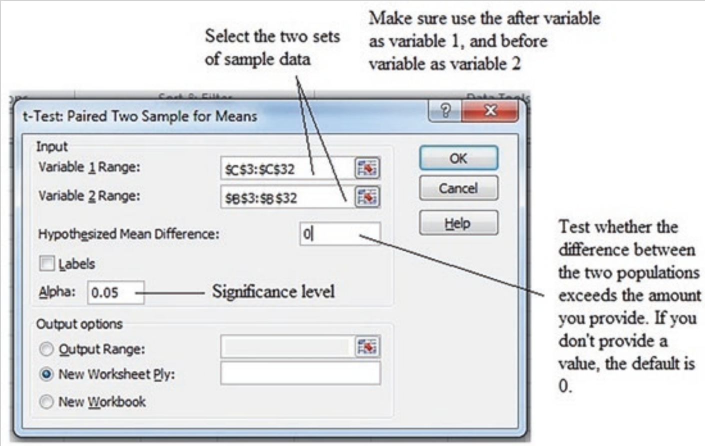
The Data Analysis Toolpak in Excel offers various tools for different types of two-sample z tests or t tests, as shown in Box 7.4.

Box 7.4 Hypothesis Test Tools in Data Analysis ToolPak

After the Data Analysis ToolPak is installed, you can access Data Analysis tools in Data Tab. It offers three t-test tools and one z-test tool.



The following figure illustrates the use of paired two sample t-test. The uses of other tests are similar.



Let us work on an example. Neighborhoods A and B are adjacent to each other. In 2016, there were 41 housing transactions in neighborhood A, and 30 transactions in neighborhood B, as shown in the following table.

Transaction prices in Community A (X)		Transaction prices in Community B (Y)	
$425,000	$385,000	$417,000	$513,140
$450,000	$295,000	$306,500	$342,000
$565,000	$324,500	$357,520	$292,000
$472,500	$530,000	$343,000	$374,000
$568,000	$512,000	$297,500	$335,000
$532,000	$575,000	$496,000	$342,130
$755,000	$360,000	$521,690	$425,000
$499,000	$375,000	$513,000	$359,900
$399,900	$480,000	$372,610	$395,000
$302,000	$339,900	$485,000	$319,000
$355,000	$510,000	$635,000	$465,000
$476,000	$500,000	$360,000	$386,280
$550,000	$435,000	$370,000	$294,000
$311,000	$315,000	$420,000	$294,000
$330,700	$530,000	$385,150	$317,500
$330,000	$545,000		
$754,000	$335,000		
$565,000	$430,000		
$352,330	$318,000		
$402,000	$616,700		
$425,000			

Based on these housing transaction data, a planner would like to assess whether the average housing prices of these two communities equal.

The first step is to set up the hypotheses, as:

H_0: The average housing prices of these two communities equal, i.e., X = Y
H_1: The average housing prices of these two communities do no equal, i.e. $X \neq Y$

The next step is to find out the test-statistic in Excel. To do so, use Data > Data Analysis, and choose t-Test: Two Sample Assuming Unequal Variances. Select X as variable 1 and Y as variable 2, and 0.05 as the critical value (Alpha). Figure 7.11 presents the excel outputs. The t-statistic is 2.568.

The third step is to define decision rules, using 0.05 as the significance level. Two-sided t-test fits our hypothesis. Excel returns information for both one- and two-sided tests. Based on the information above, we develop our decision rules based on the information for a two-sided test (Fig. 7.12). The rules are accept the null hypothesis if the t-statistic falls in [−1.9949, 1.9949]. Otherwise, reject the null hypothesis.

The last step is to make a decision. The calculated t-statistic is 2.568. It falls into the rejection region. Therefore, the decision is to reject the null hypothesis

	X	*Y*
Mean	451964.1463	391130.6667
Variance	13396978905	7032220131
Observations	41	30
Hypothesized Mean Difference	0	
df	69	
t Stat	2.568018661	
P(T<=t) one-tail	0.006198496	
t Critical one-tail	1.667238549	
P(T<=t) two-tail	0.012396992	
t Critical two-tail	1.994945415	

Use this information to set up the acceptance and rejection regions for an one-sided test

Use this information to set up the acceptance and rejection regions for an two-sided test

Fig. 7.11 2-Sample unequal variance T-test results from excel

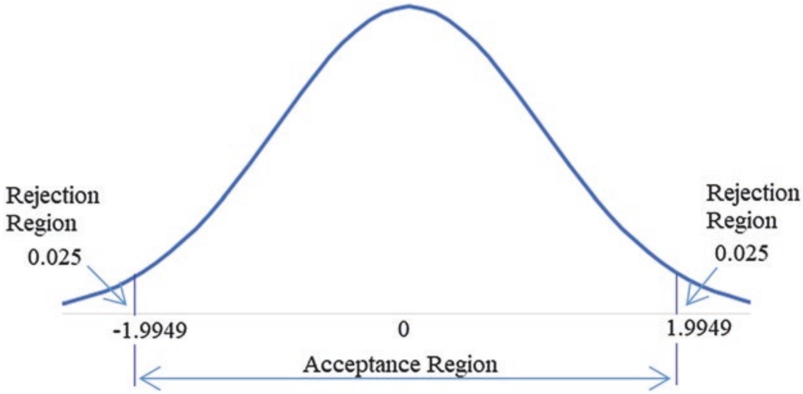


Fig. 7.12 The decision rule –the housing value case study

and accept the alternative. The average transaction prices of these two neighborhoods do not equal.

Type I Vs. Type II Error

A hypothesis test is a useful decision-making tool. However, one could make mistakes. Let us use the example of an Enterprise Zone (EZ) program evaluation to explore possible errors in hypothesis testing. An EZ program uses tax incentives to encourage businesses to move into distress areas. Policy makers are eager to know whether this program is effective in increasing job opportunities in target

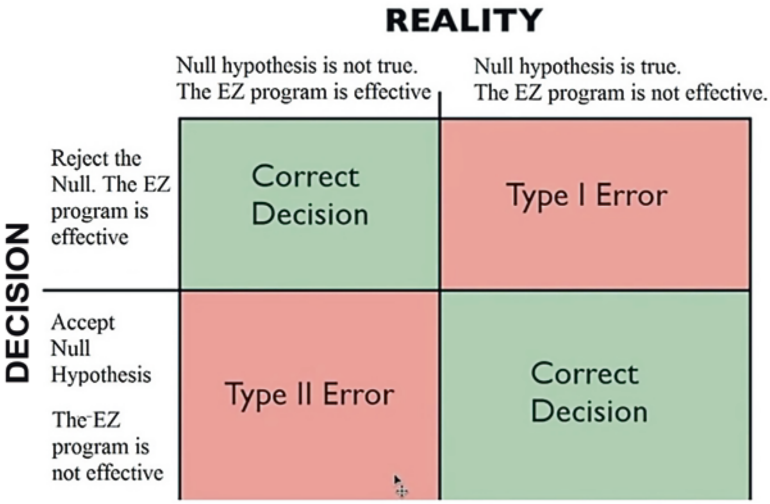


Fig. 7.13 Type I vs II errors

neighborhoods and revitalize these neighborhoods accordingly. The hypothesis can be formulated as:

H_0: The EZ Program is not effective.
H_1: The EZ Program is effective.

There are two possible errors in a hypothesis testing (Fig. 7.13). When the program is not effective, but the decision is that the program is effective, ***type I error*** occurs. It is a situation in which the null hypothesis is true, while it gets rejected. When the program is effective, but the decision is that the program is not effective, ***type II error*** occurs. This is a situation in which the null hypothesis is not true, while it gets accepted.

What are the possible consequences of these two types of errors? For type I error, the program is not effective, while you recommend that it is effective. Based on the result, the government will keep on allocating budget to be used as tax incentives, while this program has no effects on local growth. This is a waste of public money. For type II error, the program is effective, while you recommend that it is not effective. Because of the result, the government may decide not to adopt the EZ program. The status quo will be maintained. There are no benefits and no loss of local budget for supporting the program.

In choosing a significance level for a hypothesis test, one is deciding how much she wants to risk committing a type I error, i.e. rejecting the null hypothesis when it is true. The area of the rejection region represents the likelihood of committing a type I error. Type II error is related to the power of the test, which is the probability of rejecting the null hypothesis when it is false. One minus type II error is the power of a test. Therefore, the lower the type II error, the higher the power of a hypothesis test.

The Simple Linear Regression

Dependent Vs. Independent Variable

To address dependent and independent variables, one must first explore the concept of causal relationship or causation. ***Causation*** indicates that one event causes or contributes toward the occurrence of another event. A ***dependent variable*** is the variable that is of interest for planners or researchers and is the outcome of many other factors. An ***independent variable*** is a variable that we expect to have an effect on the dependent variable. For instance, if the interest is to study individual income, we may collect information related to gender, education, and age. We will use gender, education, and age to explain the differences in individuals' income. These variables are independent variables, or explanatory variables. Individual income is the independent variable, or response variable.

The Concept of Correlation

To justify a regression analysis, it is important to first explore the correlation between the independent and the dependent variables. If there are not significant correlations, causal relations do not exist. It is not necessary to spend time for further conducting a regression analysis.

A ***correlation analysis*** tests whether two variables are linked, i.e., whether the change in one variable is associated with the change in the other variable. A correlation analysis between two variables yields one number, which is between −1 and 1. This is called the correlation coefficient. It is denoted as r. If $r = 0$, it indicates that is no correlation between these two variables. If $r > 0$, it indicates a positive relation, i.e., with the increase in one variable, one would expect an increase in the other variable. If $r < 0$, it indicates a negative relation, i.e., with the increase in one variable, one would expect a decrease in the other variable. In Excel, the function for finding out the correlation between two variables is CORR.

Simple Linear Regression

A ***simple linear regression*** is a statistical model to explore the relations between two variables: the dependent variable and the independent variable. In the case of population trend analysis, the dependent variable is P (population) and the independent variable is year. The regression function can be written as:

$$P = \alpha + \beta * Year + \varepsilon$$

Year	Population
1910	2289905
1920	2416630
1930	2614589
1940	2845627
1950	2944806
1960	3038156
1970	3218706
1980	3660777
1990	3685296
2000	4041769
2010	4339367

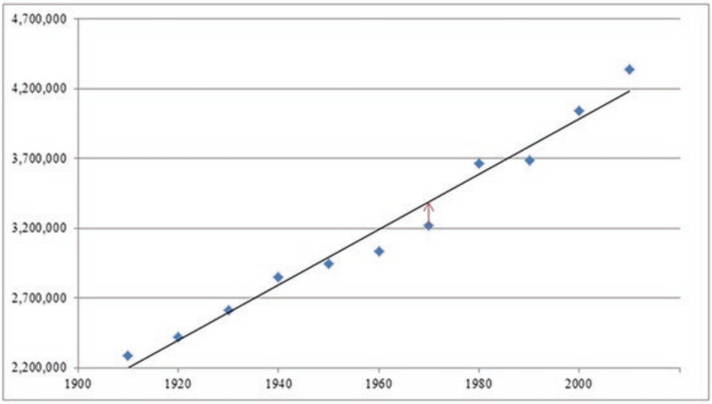


Fig. 7.14 Simple linear regression for Kentucky population

where P is the population level for a specific year and ε is an error term. The goal is to use the independent variable *Year* to explain population level. More specifically, the goal is to find the values for α and β, which are parameters or coefficients for this regression model. α is the intercept. It is the expected population level when *Year* equals 0. β indicates the impact of *Year* on P. β is also the slope for this regression line as presented in Fig. 7.14. When the absolute value of β is bigger, the slope of the line is deeper and the impact of *Year* on P is higher, with a more dramatic annual population increase or decline. Using Kentucky as a case study, the regression analysis will yield a trend as shown in Fig. 7.14. In the figure, every dot presents the actual population at a time, and the corresponding location on the trend line is the estimated population by the regression curve. For instance, the actual 1970 Kentucky population is 3,218,706, while the regression curve suggests another value. The distance between the actual value and the suggested value (as marked as red arrow line in the graph) is the error. It can be understood as the amount that one would regret (i.e. not willing to take). The curve that a statistical software provides you is the one with the minimum level of total regrets (or total error). A statistical software minimizes the total squared errors for all data points. Therefore, sometime people refer a linear regression analysis as a least squared regression.

Regression Analysis in Excel

There are two ways to conduct simple linear regression analysis in Excel. One is to make a scatter plot, add a trend line, and then add the regression equation and R-square. The other way is to use the Regression tool in Data Analysis Toolpak. Results produced by using this method are similar to the results of other statistical

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.987621396
R Square	0.975396021
Adjusted R Square	0.972662245
Standard Error	110364.1241
Observations	11

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4.34584E+12	4.35E+12	356.7945	1.50218E-08
Residual	9	1.09622E+11	1.22E+10		
Total	10	4.45546E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-35767494.29	2062738.466	-17.3398	3.18E-08	-40433732.9	-31101255.7
Year	19876.53364	1052.280635	18.889	1.5E-08	17496.10946	22256.95781

Fig. 7.15 Simple linear regression outputs in excel

software, such as SPSS and SAS. Figure 7.15 presents the regression outputs produced by Excel for using Year information to explain Kentucky population levels.

- The Fitted Regression Model

The coefficients values correspond to α and β values in the regression question, with $\alpha = -35767494.29$ and $\beta = 19876.53364$. Therefore, the fitted regression model is:

$$P = -35767494.29 + 19876.53364 * Year$$

When t increase by 1 (i.e., 1 year), Kentucky population increases by approximately 19,877.

- Standard Error and t-Stat

The standard errors for the two variables are used to calculate t-statistics for testing whether the parameters are significantly different from 0. For instance, for variable *Year*, the null hypothesis is that the coefficient equals 0, i.e., the *year* variable does not have an impact on the dependent variable; and the alternative is the coefficient does not equal 0, i.e. the *year* variable has an impact. For testing this, the t-statistic is calculated as the fitted coefficient value divided by its standard error, 19876.53364/1052.280635 = 18.889, which is the t stat listed on the next column.

- P-value

P-value is the probability of observing the t-statistic (t Stat) given the null hypothesis is true. A smaller p-value provides a stronger support for accepting the alternative hypothesis. The P-value for *Year* is 1.5E-08 (0.000000015), an extremely small number and almost close to 0. If we use .05 as the significance level for the hypothesis test about whether *Year* has an impact, with the P-value smaller than the significance level, we will accept the alternative hypothesis. Given the null hypothesis is true, the probability to observe a t-statistic of 18.889 is very minimum, 1.5E-08. However, now we do observe it. Therefore, the decision is that the null hypothesis is false, and the alternative is true.

- Confidence Interval

Lower 95% and Upper 95% are the lower and upper bounds of the 95% confidence interval for the coefficients. For variable *Year*, at the 95% confidence interval, the coefficient value is in the range of [17496.10946, 22256.95781].

- The ANOVA Table

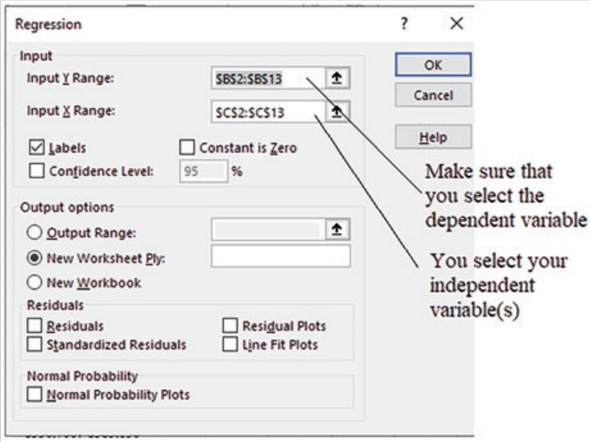
The ANOVA table presents variations among data and attributes these variations to different sources (i.e., regression, residual, and total). SS, or SSE, is the abbreviation for sum squared error. SSTotal, 4.4556E+12, is the total squared errors (i.e. variations) in the dependent variable. SSRegression, 4.34584E+12, is the amount of squared variations that are explained by the regression model. SSResidual, which can also be named as SSError, is the amount of squared variations that are not explained by the regression model and remain as errors or residuals. Ideally, we would hope that the regression model will explain all the variations of the dependent variable. In this case. SSRegression equals SSTotal and SSResidual is 0. MS is mean squared error and is calculated as the sum squared error divided by the corresponding degrees of freedom listed in the table. For instance, MSRegression is the mean squared variations that can be explained by the model and MSResidual is the mean squared error.

The F value is calculated as the MSRegression (4.35E+12) divided by the MSResidual (1.22E+10), yielding F = 356.7945. Significant F is the p-value associated with this F statistic for testing the hypothesis about whether the independent variable(s) in the model reliably estimate the dependent variable. Here the P-value is very small, suggesting that the model is a reliable model for explaining the population level in Kentucky.

- R-square

R-square is another thing worth mentioning. This figure tells what proportion of variations in the data can be explained by the model. It is calculated as SSRegression divided by SSTotal. In this specific example, this simple model explains 97.54% of variations within the data (Box 7.5).

Box 7.5 Regression Tool in Data Analysis ToolPak
After the Data Analysis ToolPak is installed, you can access Data Analysis tools in Data Tab. It offers Regression tool.



Spatial Statistics

Conventional linear regression methods assumes that observations are independent of each other and uncovers cross-sectional relations. However, strong interactions exist among urban activities between locations within a city or between cities. Therefore, the observation at one location can have an impact on observations at other locations. Spatial statistics offer tools for analyzing such spatial data.

- Spatial Dependency

The Tobler's First Law of Geography states that "everything is related to everything else, but near things are more related than distance things." (Tobler, 1970, p. 234). Activities taking place at a location depend on the activities in neighboring areas. For instance, areas near a retail center are expected to be attractive to retail activities or to residential activities. There is a positive dependency among these activities. Nevertheless, the presence of a manufacturing factory or a sewage plant may discourage residential developments in neighboring areas. This offers an example of negative dependency among activities across the space.

- The Weights Matrix

A weights matrix quantifies the spatial relationships that exist within a dataset. It is a square matrix, with equal numbers of rows and columns. For a dataset with n spatial entities, the generalized form of the weights matrix (***W***) is:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$

where w_{ij} represents the spatial relation between the ith and the jth spatial unit. There are three common types of weights matrix: neighboring matrix, distance matrix, and interaction matrix. A neighboring matrix is a binary matrix presenting contiguity relations among spatial units. If two spatial units are neighbors and share a common border, $w_{ij} = 1$; otherwise, $w_{ij} = 0$. In a distance matrix, w_{ij} is developed based on the distance between the spatial entity i and j. It can be the distance itself, the reversed distance, the reversed exponential distance, or other form of the distance. In an interaction matrix, w_{ij} measures the level of interaction between entities i and j, for example, the number of people commuting between these two locations.

- Spatial Autocorrelation

Spatial autocorrelation measures the level of dependency between observations at different locations. A commonly used statistic is the Moran's I, which is formulated as:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - \overline{x}\right)\left(x_j - \overline{x}\right)}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\right) \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}$$

where n is the number of observations, i.e. spatial units, x_i and x_j are the observed values at locations i and j, $\overline{x}$ is the mean of the observations, and w_{ij} is the weight index that presents the spatial relation between locations i and j. The Moran's I statistic takes a value between −1 and 1, with a positive value indicating a positive spatial autocorrelation and a negative value indicating a negative autocorrelation. A value close to zero indicates no spatial autocorrelation.

- The Spatial Autoregressive Model

The spatial autoregressive (SUR) takes the form of:

$$y = \rho W y + \beta X + \varepsilon$$

where y is the dependent variable, X is a set of independent variables, ***W*** is the spatial weights matrix, ε is the error term. ***Wy*** controls for the impacts from neighboring regions' activities. ρ measures the strength of spatial dependency. If ρ equals 0, there is no spatial dependency, and this SUR model is simplified into a conventional linear regression model.

Appendix 1: Standard Normal Distribution

Standard Normal Cumulative Probability Table

Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

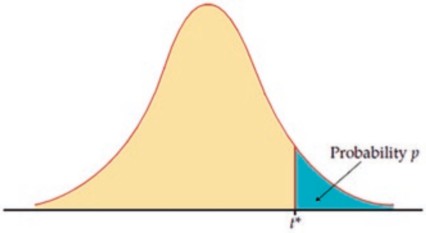
Standard Normal Cumulative Probability Table

Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Appendix 2: T-Distribution

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

References

Bovy, P. H. (2009). On modelling route choice sets in transportation networks: A synthesis. *Transport Reviews, 29*(1), 43–68.

Di Baldassarre, G., Schumann, G., Bates, P. D., Freer, J. E., & Beven, K. J. (2010). Flood-plain mapping: A critical discussion of deterministic and probabilistic approaches. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55*(3), 364–376.

Shankar, V. N., Ulfarsson, G. F., Pendyala, R. M., & Nebergall, M. B. (2003). Modeling crashes involving pedestrians and motorized traffic. *Safety Science, 41*(7), 627–640.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(sup1), 234–240.

Chapter 8
Qualitative Data Analysis

Chapter 4 introduces the typology of data, where ***quantitative data*** indicate numeric data and ***qualitative data*** are those non-numeric. In general, data can be measured at four levels: nominal, ordinal, interval, and ratio, where ratio data are those that can be calculated and manipulated in any mathematical formulas and equations, and nominal data have the most restriction in terms of mathematical manipulations. Qualitative data are mostly nominal data, which are subjected to very limited mathematical manipulations. In certain situations, quantitative and qualitative data can be converted into each other. For example, gender, as a variable, traditionally has two attributes, male or female, although more recently the variables in many studies are no longer binary. Gender is usually an important variable explaining many different socioeconomic phenomena, but it cannot be directly used in statistical models when being kept in the qualitative format. In this case, qualitative data can be converted to quantitative format, such as 0 as female and 1 as male, and then the variable "gender" can be used in statistical models as a dummy variable. For example, dichotomous data, such as the yes and no questions, are qualitative data, and can be converted to the values of 0 and 1, which are often called ***dummy variables***, and then used in mathematical models. Other than that, frequency analysis is pretty much the only manipulation for the dummy variables. There are other types of qualitative data, such as documents, music notes, images, videos, focus groups, public meetings, interviews, or other data that are mostly descriptive, void of quantitative details. Analysis of qualitative data is therefore, very different from quantitative data.

Qualitative data search for explanatory patterns and relationships, and often can link data collection, analysis, and theory better than quantitative analysis. *Quantitative data analysis* usually follows a set of methods, such as regression analysis or various statistical procedures to reveal relationships and test hypotheses; however, *qualitative data analysis* is less standardized and researchers use various methods to find patterns and relationships. Data in qualitative analysis are sometimes less precise, context-oriented, and sometimes have multiple meanings. Oftentimes, qualitative data analysis complement quantitative data analysis to provide real-life explanations to the findings from quantitative data analysis. Qualitative data analysis predominantly involves around coding and categorizing data to reveal patterns or themes (Wong, 2008). When analyzing qualitative data, it is critical to

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_8].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_8

"connect particular data to concepts, advanced generalizations, and identify broad trends or themes" (Neuman, 2011). For example, in statistical analysis, a researcher finds that homeownership, particularly, being a homeowner, is significantly positively related to satisfaction with one's neighborhood. However, the researcher may not know why being a homeowner may relate to higher neighborhood satisfaction. Note here that we did not use a causal relationship, but simply a relationship, since proving causal relationships needs specific types of statistical methods. A regression model often cannot prove any type of causal relationships. There are a few ways that the researcher will find out why being a homeowner often contributes to increased neighborhood satisfaction. The first approach is to explore what other similar studies have found out why. The second approach is to select a sample of homeowners, ideally stratified based on characteristics, such as choosing homeowners from different age, income, education levels, racial and ethnicity background, marital status, whether having school-aged children or not, and other important factors, to conduct an interview. The factors used to determine how to structure the interview sample are based on reviewing extant literature about which household-level characteristics mostly contribute to neighborhood satisfaction. Therefore, qualitative data will better explain the statistical findings from sophisticated mathematical models.

Therefore, it is imperative for planners to understand the value of qualitative research, since many scholars often are carried away by statistical models, especially by using the most advanced and sophisticated models without carefully thinking about the underlying theoretical and explanatory reasoning of using these mathematical or statistical models. This chapter will start with an introduction of the types of qualitative data, followed by an analysis of qualitative data. Then the chapter concludes with a brief introduction of common software used in analyzing qualitative data.

Discovering Patterns in Qualitative Data

Qualitative data are not quantitative in nature and often are nominal data, which is the lowest level of statistical measurement, and thus qualitative data has the least mathematical calculation capacity. However, if you are conducting research of all the Long Range Transportation Plans, which are often written by local Metropolitan Planning Organizations (MPO), from a certain state of the U.S., you can still find patterns of similarity among all the plans, and consequently identify differences. You can use key words, such as mobility, accessibility, smart transportation, etc. as a criterion, which are often determined by the research questions, and count how many times these key words are mentioned in each planning document. Then you can go to the places that the key words appeared and read in more details to understand whether there are any unique policies or actions that these key words relate to. Based on the comparison between all the documents you are able to tabulate the similarities and differences to help answer your research questions.

The key questions to ask when looking for patterns are, (1) How can these patterns be used in a broader context to help address the research questions; (2) Are there any deviation or outliners from the patterns and how to explain these deviations? (3) Do the patterns suggest additional data need to be collected? (4) How do these patterns relate to the theories or previous studies on the topic (Frechtling & Sharp, 1997)?

Generally speaking, there are five ways to discover patterns:

- Frequencies
- Magnitudes
- Structures
- Processes
- Causes and Consequences

Frequencies

Measuring how often key words or concepts appear in the qualitative data is often the first step to reveal any patterns in these data. For example, you interviewed local business leaders and other key informants about the effectiveness of initiatives that local government implemented to promote immigrant-operated businesses. After the interview is transcribed, you found the initiatives mentioned by the interviewees about the positive impact, then you count the frequencies that these initiatives are mentioned regarding the positive impact. A simple table to illustrate the frequencies of each initiative in promoting business activity will then help explain which initiatives have the most impact. Typical narratives about these positive impact should then be presented to help strengthen the frequency analysis.

Magnitude

Using the immigrant-operated businesses as an example, if the frequency analysis reveals certain initiatives are mentioned more than others it is often beneficial to explore how these initiatives are mentioned by the interviewees. Therefore, the level of impact, magnitude, should be revealed when specifically looking at detailed responses regarding the relationship between these initiatives and how much impact the interviewees have felt.

Structure

One of the important purposes of qualitative data analysis is to reveal structure and relationships among different concepts. For example, when exploring the meaning of home of a certain group of people, interviews and surveys may reveal the relationship between material items, social interactions, and the meaning of home for the research subjects.

Processes

Qualitative data may reveal the process associated with behaviors, opinions, or interactions. For example, analyzing the process of public meetings may help planners understand the needs and wants of their residents. A SWOT, Strengths, Weaknesses, Opportunities, and Threats, brainstorming process is often conducted in planning and public policy to help identity community needs. Processes may also be evaluated based on certain criteria to derive the most effective processes.

Causes and Consequences

One of the advantages of qualitative data is that they are able to help a researcher to pinpoint the potential causes of certain phenomena. Quantitative data analysis often reveals the statistical relationships among dependent and independent variables, but often fails to explicitly explain causal relationships. For example, when examining why a population cohort is not interested in participating in the planning process, key informants or a representative sample of population may be selected to take surveys and interviews regarding the likelihood of them participating in planning and public decision-making process, and the potential reasons why they are disinterested. Qualitative data may also reveal consequences associated with certain behavior or act, or the potential impact the behavior or act has.

Types of Qualitative Data

Observations and Ethnographic Data

Observing actions or behaviors of other people is one of the most commonly used methods in gathering qualitative data. Ethnographers, who often stay within the study population, observe the day-to-day life of certain groups of people, and then analyze and write about the life of these people. Some ethnographers are disguised, while others are not. Regardless of the approach the observers take to obtain data, they need to strictly follow research ethics (e.g. informed consent, confidentiality and privacy protection of the research subject, no harm or minimum risk to the research subject, etc.), local traditions, regulations, and laws. In urban and regional planning, observations can be used in observing the usage of certain public facilities and how people interact in these facilities, which may help the planners identify any issues associated with the facilities and possible improvements to be implemented. Planners can also observe residents' behavior and opinions in public meetings or forums.

Observation data should be carefully recorded; however, audio or video recording of the individuals being observed need to be approved by these individuals or their guardians. In an academic setting, whenever in doubt about the proper procedures of following research ethics, the Internal Review Board (IRB) should always be consulted. Practice planners should also follow appropriate procedures regarding video or audio recording, or use of photos in published materials. Local laws sometimes stipulate that whenever individuals attend any public meetings or forums they agree being recorded. The same procedure often applies to surveillance recordings in any public properties.

In order to generate meaningful analysis of patterns and relationships, observations often need to be done in extended time period and for multiple times. The frequency or time duration of observations depends on the nature of the study and the specific research questions.

Observations have advantages and disadvantages when being used as a data collection method (Mahoney, 1997). Some of the advantages include providing direct information about individual or group behavior and opinions and the environment is natural, unstructured, and flexible. However, in situations when the research subjects know they are observed, they may behave differently, therefore, resulting in possibly atypical or misleading findings and conclusions. Well-structured observations may be time consuming, and sometimes expensive. Observers need to be highly trained and preferably should be content experts in order to identify key patterns and relationships. Personal perceptions of the observer may also distort the analysis process of observatory data. Regardless of these limitations, well-intentioned observations may help gather meaningful qualitative data to help in research or practice.

Interview Data

As mentioned in Chap. 4, interview is one of the most commonly used data collection method, particularly relating to qualitative data collection. Opinion and behavioral data are most often gathered through interviews, particularly when the subject matter is complex, and/or highly sensitive, or when detailed information is needed (Mahoney, 1997). In urban and regional planning or public policy, answers to the following questions may benefit from an interview process (Patton, 1990):

- What does the project, program, or policy look and feel like to the participants? To other stakeholders?
- What are the experiences of the participants?
- What do stakeholders know about the project, program, or policy?
- What thoughts do stakeholders have concerning project/program/policy operations, processes, and outcomes?
- What are participants' and stakeholders' expectations?
- What features of the project/program/policy are most salient to the participants?
- What changes do participants perceive in themselves as a result of their involvement in the project, program, or policy?

Traditionally, interviews should be conducted in a comfortable, nonthreatening, and easily accessible location that provides privacy and is equipped with recording devices, promoting involvement and interaction without distractions. However, with the advancement of technology, some of the contemporary interviews can be conducted with the comfort of at home, through phones, emails, or video conferencing. Some of the rules still apply under remote settings, such as protection of the privacy of the participants, minimal distractions, recording consent, etc. Persons who conduct the interviews should have some trainings in interview and understand the nature of the data to be gathered.

Focus Groups or Design Charrette

Data gathered from focus groups or charrettes are often qualitative data as well, although researchers or practitioners may distribute survey questionnaires at those focus group or charrette meetings. Focus groups or charrettes usually discuss a few focal points and are facilitated by an experienced facilitator or professional, although charrettes tend to be more design-oriented soliciting opinions and data. Focus groups are commonly used for the following purposes (Mahoney, 1997):

- identifying and defining problems;
- identifying strengths, weaknesses, opportunities, threats, and recommendations;
- assisting with interpretation of quantitative findings;
- obtaining perceptions of project, program, or policy outcomes and impacts;
- generating new ideas.

Comparing to interviews, focus groups allow interaction of the participants, which may speed up the generation of ideas and help to achieve consensus. Focus groups should not be used in situations where the subject matter is sensitive and personal. When there are multiple sensitive issues to be discussed, interviews may be a better option. Therefore, using interviews or focus groups depends on the research project and questions, and whether interaction and discussion is important in gathering data.

Documents and Artifacts

Various documents may be used as the sources of qualitative data; for example, planners are constantly using written case studies, public records, existing plans and project descriptions, and existing ordinances and regulations to help make decisions about a certain planning project, program, or policy. They synthesize what they read and learned from these reports, and use the synthesized information to help guide their own work. This process of reading, summarizing, and synthesizing is formally called content analysis. The literature review section in Chap. 3 is also one type of content analysis.

An *artifact*, made or created by a human being and often has cultural or historic significance, may be the "data" as well in situations where analyzing these artifacts is necessary. The analysis of these artifacts is called content analysis. For example, one analyzed 20 comprehensive plans and found that in average each plan mentions the word "resilience" 13 times, with a standard deviation of 4. These indicate that 98% of the plans mention the word "resilience" between 1 time and 25 times (confidence interval of three standard deviations).

When using documents or artifacts as the data source, it is crucial that the documents and artifacts are reliable and accurate.

Analyzing Qualitative Data

Cross-Case Analysis

Cross-case analysis indicates analyses involving more than one case, for example, more than one neighborhood about how residents view affordable housing development within the neighborhood. Cross-case analysis can be either variable-oriented analysis or case-oriented analysis.

When we are interested in how residents in general view the impact of affordable housing development, we may choose a few neighborhoods with these development projects, ideally should be different types of neighborhoods, such as low income, middle income, high income, and can be even further stratified based on homeownership rate, racial composition, and other variables to make the selection of neighborhoods more representative. Then interviews or surveys, either to the households or key informants based on various sampling techniques, among these selected neighborhoods should be carried out. This nomothetic approach is often variable oriented.

In another situation, what interests the researchers are the different reactions to affordable housing development in wealthy neighborhoods. Then the researchers should only focus on wealthy neighborhoods and conduct more specific studies about residents' reactions to these projects. This idiographic approach is then case oriented approach.

Grounded Theory Method

Grounded theory method is an inductive approach to research in which theories are generated solely from an examination of data, rather than being derived from theories or findings of other studies. This approach starts with observations rather than hypotheses. The benefit of using Grounded Theory Approach without laying out initial hypotheses is that the analysis does not need to be restricted by the established theories where the hypotheses are based upon. For example, American Community Survey (ACS) provided abundant data about demographic, occupational, and commuting characteristics of the households in the sample. A planner is interested in how household income relates to commuting choices, particularly the decision of cycling to work. The planner then run a simple correlation analysis and possibly even a multivariate regression analysis to see the likelihood the household would cycle, controlling for other demographic characteristics. If the planner understands more advanced methods, such as data simulation and machine learning techniques, he or she may even generate more sophisticated models to capture the complex relationship between income and cycling. After analyzing multiple years' data to increase reliability of the study, the planner finds that income is not a straightforward explanatory factor to the decision of cycling to work. To further investigate the anomaly in the pattern, the planner may conduct qualitative data collection to confirm the theory he or she found by analyzing the data. Grounded Theory Method can be used in quantitative data analysis as well, when there are no given hypotheses. In this way, exploring the data without the limitation of hypotheses may help the researcher or planner discover unusual patterns or theories not found from other studies.

Constant Comparative Method

This method is a component of the Grounded Theory Method in which observations are compared with one another and with the evolving inductive theory. Constant comparative analysis is the most commonly used method analyzing qualitative data. Constant comparative research can be conducted inductively, deductively, and abductively. ***Inductive research*** is the method of deriving theory from data (e.g. codes are derived from the data), while ***deductive research*** is to propose theories/hypotheses and then use the data to test these hypotheses whether they should be rejected or not (propose certain codes and then compare the real data with these pre-contemplated codes). We use rejecting a hypotheses instead of accepting, because no amount of evidence can fully prove accepting, while one opposing evidence would be sufficient to reject. ***Abductive research*** focuses on iterative appearance of key concepts and codes.

When conducting ***constant comparative analysis***, the researcher reads the whole documents, divides the document into groups of passages, and then puts labeling codes to these chunks of messages (Leech & Onsuegbuzie, 2007). Similar passages are then grouped together and a theme is identified and documented for each group.

For example, below is a brief excerpt of data from an interview about "the meaning of home".

> For me, home is a place that I can relax and seek shelter physically and psychologically. This is especially important since I work a lot. I leave home at 7 in the morning, dropping the kids to school, and then head to work. I pick up the kids at 5:30 from the after-school care in the afternoon and won't be home until 6pm. We then have to rush to get dinner ready and eat. Then when I am cleaning up, my husband checks my children's homework that they worked on in the aftercare center. After the kids go to bed, I don't have much time left for myself before going to bed. I do some art work or knitting, but sometimes just too tired to do anything. Weekend is busy too and I got so little time spent at home and I cherish every moment of being at home. When I don't have to tend the kids nor do chores, the rare moments, I sit on our front porch, read a book or simply relax and enjoy everything. We don't have much money to decorate our house, but I try my best to make it cozy, neat, and inviting. This is very important for me and the family.

The passage is broken into chunks of passages, and code for each chunk is assigned (Table 8.1). Then the theme of the passages can be retrieved and summarized as in Table 8.2.

Table 8.1 Emergent codes from constant comparison analysis

Passages	Code
I can relax and seek shelter physically and psychologically	Relax, shelter
I leave home at 7 in the morning, dropping the kids to school, and then head to work. I pick up the kids at 5:30 from the after-school care in the afternoon and won't be home until 6 pm. We then have to rush to get dinner ready and eat. Then when I am cleaning up, my husband checks my children's homework that they worked on in the aftercare center. After the kids go to bed, I don't have much time left for myself before going to bed. I do some art work or knitting, but sometimes just too tired to do anything.	Don't have much time left for myself, too tired
This is especially important since I work a lot… Weekend is busy too and I got so little time spent at home and I cherish every moment of being at home.	Work a lot, busy, important, cherish
When I don't have to tend the kids nor do chores, the rare moments, I sit on our front porch, read a book or simply relax and enjoy everything.	Relax, enjoy
We don't have much money to decorate our house, but I try my best to make it cozy, neat, and inviting. This is very important for me and the family.	Cozy, neat, inviting, important

Table 8.2 Summarizing themes from codes

Code	Theme
"relax", "shelter", "enjoy", "cozy", "neat", "inviting"	Make the home cozy, neat, and inviting, so that the person can relax and enjoy.
"work a lot", "busy", "Don't have much time left for myself, too tired", "important", "cherish"	The person is very busy every day, so it makes sense that she values every moment spending at home, especially when she has time to relax.

Keywords-in-Context Analysis

Keywords-in-context method uses words appearing before and after the key words as context to compare how key words are used by different research subjects (Fielding & Lee, 1998). Through this type of analysis, researchers are able to identify patterns, relationships, and reveal interesting facts from less interesting conversations or speeches. One drawback of this approach is that if the context words are not sufficient, it may be difficult to reveal the real meaning of the key words.

Conversation Analysis

Conversation analysis meticulously analyzes the complete transcript of conversations, including the implications of pauses, hems, haws, and even gestures. Sociologists, instead of planners, may analyze these conversations in detail. However, planners may find it useful by understanding the conversation between public meeting attendees and a planning commissioner, so that the planner may incorporate the key meaning from these conversations in planning recommendations. The purpose of conversation analysis is to uncover the implicit assumptions and structure in social life. The first fundamental assumption of conducting conversation analysis is that conversation is a socially structured activity and has rules of behavior. Another assumption is that conversations must be understood contextually. The third assumption is that structure and meaning of conversations must be transcribed.

Word Count Analysis

Linguistics believe that when people talk, they often follow certain habits and use certain words more or less than other people or in different contexts. Therefore, qualitative researchers sometimes count words with significant meanings in a transcript or document. How frequent of certain words appear in the study materials will then determine the overall implications and meanings of the materials. One weakness of word count, however, is that the subjects may not need to use certain words to convey their opinions or perception, and word count without contexts may be meaningless to the researcher. Word count is especially a very useful technique in focus group research by identifying the most vocal, the least vocal, and words frequently expressed by the participants (Leech & Onsuegbuzie, 2007).

Classical Content Analysis

Classical content analysis, similar to constant comparative analysis, is used frequently in qualitative research. Constant comparative analysis mostly focuses on themes, while content analysis counts the frequency of codes and identifies the most important and the least important codes in a document. Usually these codes are generated deductively to be used testing a hypothesis and these codes can be analyzed using quantitative procedures or a mixed method (Kelle, 1996; Onsuegbuzie & Teddlie, 2003).

Semiotics and Domain Analysis

Semiotics studies the science of signs and symbols, and any specific meanings associated with them. Signs and symbols are highly related to cultures, ideologies, background, and other characteristics. They are the commonly agreed conventions within certain groups of people or within a specific place. Symbols have three layers of identities: the symbol itself (cover term), what the symbol refers (terms), and the relationship between the symbol and the referent (semantic relationship) (Spadley, 1979). For example, carved pumpkins and the orange color are often associated with harvesting and the Halloween. Victory can be expressed with the index and middle finger standing showing the letter "V". In English culture, when asking people to smile when taking a picture, "say cheese" is often used. There are many other types of signs and symbols representing specific meanings or even attached feelings of specific groups of people or places. Various architectural design can indicate different cultural backgrounds, particularly the differences between different civilizations.

Understanding the meaning of signs and symbols can better help planners understand the diverse needs of the residents. Some researchers, particularly in sociology and archeology, are devoted solely in studying signs and symbols, by searching for the "larger units of cultural knowledge" (domains) (Leech & Onsuegbuzie, 2007, p. 570). ***Domain analysis*** is not limited to analyzing strict symbols; it can be used in analyzing any type of text features in a document.

Taxonomic Analysis

Taxonomic analysis is the second step after a domain analysis, although domain analysis can be used alone in a research project. Taxonomic analysis attempts to understand how research participants use specific words and the attributes of each domain. In a domain analysis, once the domains have been identified, each of the domains will be probed further to refine the domains into a taxonomic system, to

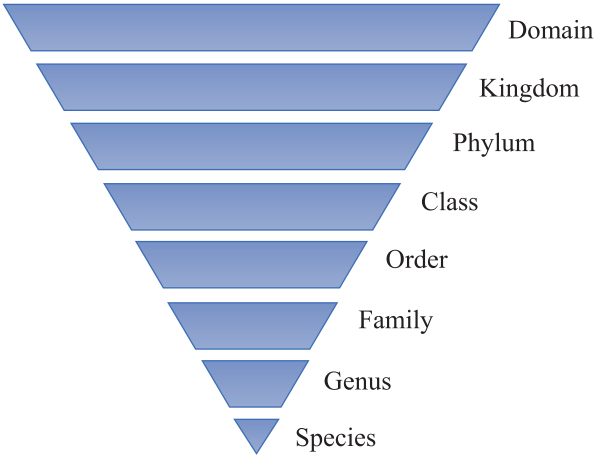


Fig. 8.1 Taxonomy in biology

define how the domains are related to each other, and to discover similarities and differences among domains. For example, the following example of a taxonomy hierarchical system in biology may help us understand the position of domains in the taxonomic system (Fig. 8.1). Domain is the highest rank in the hierarchical biological classification system; for example, there are three domains of life, the Archaea, the Bacteria, and the Eucarya.

Componential Analysis

Componential analysis is another procedure after the domain analysis. Once the domains are identified, the attributes of each domain is explored further to discover how each attribute contributes or relates to the research questions. For example, in traffic safety analysis, there may be a few domains, such as pedestrian safety, bicyclist safety, motorist safety, and safety signals and design. The researcher is interested in the attributes under each domain, such as how pedestrian safety is different or similar from bicyclist safety and what factors determine the differences and similarities. The componential analysis, therefore, provides more refined details of the domains and how attributes under each domain relate to each other.

Qualitative Data Processing

Qualitative data often include complex texts or graphs which cannot be easily processed or explained by contemporary computer software technologies. The focus of computing is often in quantitative data and the most sophisticated software or applications are often used to process and analyze quantitative data. However, there are ways using software applications to process qualitative data before analyzing. Much of the qualitative software is able to generate codes (key words or concepts), summarize codes, and provide basic analysis in patterns and relationships. Qualitative data processing usually involves the following procedures.

Coding

Coding is the first step of qualitative data processing. When facing complex textual data the first step is to classify and categorize individual concepts within a retrieval system, such as a computer software called NVivo or Atlas.ti. There are three types of coding systems, open coding, axial coding, and selective coding. ***Open coding*** is the initial classification and labeling of concepts in qualitative data analysis. For example, a planner is interested in how local transportation plans integrate smart and artificial intelligence technologies in future planning. The initial concepts that the planner wishes to focus on may be smart, artificial, intelligence, sensing, autonomous, automation, autopilot, etc. These key words are then labeled in the documents, with the help of computer software.

Axial coding involves around a reanalysis of the results of open coding in Grounded Theory Methods, in order to identify the important, general concepts. The process of axial coding may involve regrouping of data, or to look for more analytical concepts. For example, when looking at smart transportation technologies, the planner can regroup the planning documents based on the frequency of each of the open codes appears in the documents. The planner may also propose new concepts, such as new materials, to identify if any future plans will use new materials in constructing transportation infrastructure. The planner may also expand the key concepts to others, such as "high speed rail", "hyper loop", "maglev train", "alternative modes", etc.

Selective coding builds on the results of open coding and axial coding to identify the central/general concepts that organizes the other concepts as identified in the textual materials. For example, based on open coding and axial coding, the planner can then categorize the findings to smart roadways, high-speed rail, autonomous vehicles, or other categories to summarize the findings.

Memoing

In qualitative research, ***memoing*** involves the process of writing notes about the project. When writing the notes, one should record the time and date details of the interview or data collection methods, pay attention to things that stick out, and highlights important details relating to the research question from the qualitative interaction. Memoing should be completed during the data collection process, or right after the process when memories are still fresh. The memos, which are not formal memoranda as mentioned in Chap. 6, are also part of the data that can be used in analysis. Memos can describe and define concepts, deal with methodological issues, or offer initial theoretical formulations. There are three types of memos in qualitative research. One is *code notes*, describing and identifying code labels and their meanings. Code notes is similar to the metadata in database analysis. Metadata is data about data. Variable codebooks, explaining the type, meaning, and attributes of the variables, is one type of metadata. Another type of memo is *theoretical notes*, reflecting the meanings of concepts and theories, and their relationships. The third type is *operational notes*, often addressing methodological issues when conducting qualitative data collection and analysis.

Concept Mapping

After the coding process, it is imperative to organize the main concepts derived from open, axial, and selective coding. Putting concepts in a graphic format will help the researcher or planner, or the targeted audience to visualize the classification of individual concepts. The following example is a simple graphical ***concept mapping*** indicating the relationship between different concepts for the GIS data of the transportation planning example mentioned in the coding section.

Graphs like Fig. 8.2 is often the theoretical foundation of a statistical model called Structural Equation Model, where the influences and structures of variables are revealed through a flow chart and the flow chart can then be represented with a series of structural equations to denote the relationships.

Spreadsheets are also another effective way to map the concepts. In a spreadsheet, key concepts can be listed and checked for each of the text passages, which will be included in the spreadsheet as well.

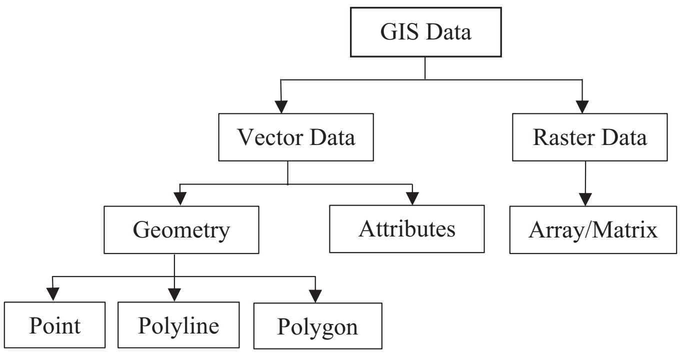


Fig. 8.2 An example of concept mapping

Common Software for Qualitative Data Analysis

Before the advancement of technology, interview and other qualitative data transcriptions are analyzed through tedious and clerical work such as manual coding, memoing, colored pens, "cut and paste", and other approaches. The invention of software in qualitative data analysis made the tasks much easier and faster (St. Johnson & Johnson, 2004; Wong, 2008). Software use also increased validity and auditability of qualitative research. Generally speaking, commonly used software in qualitative data analysis should include the following functions (Predictiveanalyticstoday.com):

- Annotation, enabling annotating within the texts or the materials being analyzed.
- Data visualization such as heat maps, clustering, or multidimensional scaling.
- Media analytics to analyze various media forms.
- Mixed methods research integrating quantitative and qualitative research.
- Qualitative comparative analysis by comparing and testing hypotheses in the data.
- Quantitative content analysis by quantifying key attributes of the study materials.
- Sentiment analysis by identifying and classifying opinions expressed in the study materials.
- Statistical analysis including descriptive statistics.
- Text analytics analyzing texts and providing codes and categories for easy process and analysis.
- Providing content search tools.
- Providing coding tools.
- Providing linking ability.
- Providing tools for mapping or network analysis.
- Providing tools for query and writing.

Common software analyzing qualitative data include NVivo, Atlas.ti, HyperTRANSCIBE, HyperRESEARCH, and a few others. Even though computerized programs have aided qualitative researchers in transcribing, coding, and categorization, the researchers need to decide on how to categorize and code the data, and which patterns or themes can be drawn from the codes and categories.

NVivo is one of the leading computer-assisted qualitative data analysis software (CAQDAS) which not only helps generating coding, but also integrates coding into qualitative reasoning, linking, shaping, and modeling processes (Wong, 2008). The codes are saved by the application as nodes and can be used by the researchers in the analysis process. NVivo is also able to generate models delineating key concepts and relationships.

Atlas.ti is another popular software used in qualitative data analysis. It includes similar functions as other software such as NVivo and HyperRESEARCH. It also allows directly importing data from Twitter, Endnote, and Evernote data. This ability expanded the use of this software not only in qualitative data analysis, but also in literature review and other related tasks.

HyperTRANSCRIBE is a transcription application helping with transcribing interviews in the format of AVI, QuickTime, MPEG-2, or MPEG-4 (Hesse-Biber, 2010). This application allows the researcher to conduct other tasks, such as memoing while listening to the recordings during the transcribing process. HyperRESEARCH is a more complex software emphasizing coding, analyzing frequencies, structures, proposing hypotheses, importing quantitative data matrixes, and other more comprehensive and advanced analysis of qualitative and mixed methods of both qualitative and quantitative analyses. Both of these applications are created by Researchware, Inc., which specializes in developing qualitative research applications.

Regardless of widespread use of qualitative software in data processing, there are concerns using these types of software, such as the increasingly rigid processes and retrieval methods of software, the pressure on researchers to use large amounts of data rather than on depth and meaning, the time spent on learning how to use the software, increased commercialism, and distraction and deviation from the core value of conducting qualitative research (St. Johnson & Johnson, 2004). Computer-Assisted Qualitative Data Analysis Software (CAQDAS) should be only treated as data management packages to aid in data analysis process (Zamawe, 2015). Therefore, researchers need to balance the needs of using these types of software with the research question, research design, and analytic methods, critically considering the cost and benefit of using certain software.

References

Fielding, N. G., & Lee, R. M. (1998). *Computer analysis and qualitative research*. Sage.

Frechtling, J., & Sharp, L. (1997). *User-friendly handbook for mixed method evaluations*. National Science Foundation. https://www.nsf.gov/pubs/1997/nsf97153/start.htm

Hesse-Biber, S. (2010). *Analyzing qualitative data: With or without software*. (Lecture notes). https://www.bumc.bu.edu/crro/files/2010/07/Hesse-Bieber-4-10.pdf.

Kelle, U. (Ed.). (1996). *Computer-aided qualitative data analysis*. Sage.

Leech, N. L., & Onsuegbuzie, A. J. (2007). An array of qualitative data analysis tools: A call for data analysis triangulation. *School Psychology Quarterly, 22*(4), 557–584. https://doi.org/10.1037/1045-3830.22.4.557

Mahoney, C. (1997). Common qualitative methods. In J. Frechtling & L. Sharp (Eds.), *User-friendly handbook for mixed method evaluations*. National Science Foundation (NSF).

Neuman, W. L. (2011). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Pearson Education.

Onsuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook for mixed methods in social and behavioral research* (pp. 351–383). Sage.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage.

Spadley, J. P. (1979). *The ethnographic interview*. Holt, Rinehart and Winston.

St. Johnson, W., & Johnson, P. (2004). The pros and cons of data analysis software for qualitative research. *Journal of Nursing Scholarship, 32*(4), 393–397. https://doi.org/10.1111/j.1547-5069.2000.00393.x

Wong, L. P. (2008). Data analysis in qualitative research: A brief guide to using Nvivo. *Malaysia Family Physician, 3*(1), 14–20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4267019/

Zamawe, F. C. (2015). The implication of using NVivo software in qualitative data analysis: Evidence-based reflections. *Malawi Medical Journal, 27*(1). https://doi.org/10.4314/mmj.v27i1.4

Chapter 9
Demographic Analysis

Demography refers to a scientific study of population (Shryock et al., 1973). It is the basis of urban planning. Population distribution determines the provision of public services across a city. Population level and growth determines the current and future distribution of land use activities. Population make-up is the foundation of adopting and promoting many social policies. The objectives of demographic analyses are to (1) understand the demographic attributes, such as age, gender, and race; (2) uncover the dynamics of a population and the driving forces behind them; and (3) project future demographic activity levels. This chapter explores demographic data sources and introduces demographic analysis methods, including descriptive analysis, extrapolation methods, and the cohort component model.

Demographic Data

In the U.S. the Census Bureau is one of the major data providers for urban planning and public administrative activities. It operates more than 130 programs and surveys to collect data (The U.S. Census Bureau, 2020).

Resident-based Versus Employment-based Data

Census data can be divided into two broad categories, depending on the type of individuals or institutions who are surveyed. One type of survey respondents are people who reside in an area, i.e. residents. These data are resident-based data. The other type of survey respondents are workers or firms. These data are, consequently,

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_9].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_9

employment-based data. For instance, both the American Community Survey (ACS) and the Economic Census collect employment by industry data. However, ACS Survey data are produced by surveying households, while the economic census surveys workers. Therefore, the ACS data describe population attributes and are demographic data. While the economic census data are economic data, describing economic activities taking place in a community. This section introduces the two most used datasets, the Decennial Census and the American Community Survey.

The Decennial Census

The decennial census has been conducted every 10 years since 1790, as required by the U.S. Constitution to support the reconfiguration of political representation among states (Prewitt, 2000). From 1940, the Census Bureau began to publish data for small geographies, such as census tracts, for selected urbanized areas. From 1990, data for small geographies became available for the whole nation. The Census Bureau distributed two different questionnaires, the short-form and the long-from, for household surveys. All households must return the short-form questionnaires, based on which the Census Bureau developed its 100% survey data. These data are called summary tape file 1 (STF1) in 1980 and 1990, and summary file 1 (SF1) in 2000. STF1 and SF1 contain limited amount of information, including gender, age, relationship, race, and ethnicity information for population, and tenure information for housing units. The long form questionnaires were mailed to a sample of U.S. households. The sampling rate is 16.67%. One out of six households were sampled. Based on these long-form surveys, the Census Bureau produced summary tape file 3 (STF3) in 1980 and 1990, and summary file 3 (SF3) in 2000. These datasets include additional population and housing data. For instance, 2000 census SF3 data consists of 484 population tables and 329 housing tables, covering subjects as presented in Table 9.1.

The American Community Survey

The American Community Survey (ACS) is an ongoing survey. The Census Bureau sends out questionnaires to about 250,000 randomly selected households or group quarters every month (The National Research Council, 2007). Based on this survey, the Census Bureau produces various ACS products, including 1-year estimates, 3-year estimates, and 5-year estimates.

Table 9.2 provides a summary of ACS datasets. 1-year estimate data has been released annually since 2005. These data are produced based on the survey conducted throughout a year. For instance, 2018 ACS 1-year estimate is created based on the survey collected between January 1, 2018 and December 31, 2018. 2016 ACS surveyed 2,143,000 housing units in the U.S., which is approximately 1.68%

Table 9.1 Additional survey subjects in 2000 long-from questionnaire

Population	Housing
Marital status	Units in structure
Place of birth, citizenship, and year of entry	Number of rooms
School enrollment and educational attainment	Number of bedrooms
Ancestry	Plumbing and kitchen facilities
Residence 5 years ago	Year structure built
Language spoken at home	Year moved into unit
Veteran status	Housing heating fuel
Disability	Telephone
Grandparents as caregivers	Vehicles available
Labor force status	Farm residence
Place of work and journey to work	Value of home
Work status last year	Monthly rent
Industry, occupation and class of workers	Shelter costs
Income	

From the 2010 decennial census, the Census Bureau replaced long-form questionnaire with the American Community Survey. The 2010 decennial census only provides a limited amount of demographic information, but it covers the whole nation for small geographies. The 2020 decennial census will offer a similar amount of information.

Table 9.2 ACS datasets

	1-year estimate	1-year supplemental estimate	3-year estimate	5-year estimate
Geographic area population size				
65,000+	X	X	X	X
20,000-64,999		X	X	X
Less than 20,000				X
Sample period	12 months		36 months	60 months
Sample size	Smallest		Larger than 1-year estimate	Largest
Release information	Annually released since 2005 to present	Annually released since 2014 to present	Annually released from 2005–2007 3-year data to 2011–2013 3-year data	Annually released since 2005–2009 5-year data to present
Best used when	Currency is more important		–	Precision is more important and/or must study small geographies (e.g. census tracts, block groups)

of the total housing units in the nation. With this small sample size, 1-year estimate data are less accurate and have large margins of errors. This data is the best choice in a situation that demands for timely data. Further, 1-year estimates are only available for a limited number of counties with population larger than 65,000 and are not available at small geographies. Please see Fig. 9.1 for the availability of 1-year estimates. From 2014, the Census Bureau began to issue 1-year supplemental estimates for areas with population over 20,000.

Five-year estimates are created based on surveys collected over a 5-year span. For instance, 2014–2018 5-year estimates are created based on surveys collected between January 1, 2014 and December 31, 2018. There are two advantages of this data. First, the sample size is much larger than 1-year estimates, and therefore, these estimates are more precise. Second, the data is available for all geographies, including small geographies such as census tracts and block groups. Nevertheless, the fact that these are estimates over a long period of time (5 years) presents challenge of using these data. This lack of time accuracy makes these data less useful when currency is important.

Selecting between 1-year estimates and 5-year estimates involves a trade-off decision between currency and reliability. Three-year estimates are created based on surveys over a 3 year time frame, and these estimates' currency and reliability stay between 1-year and 5-year estimates.

ACS estimates provide a similar amount of population and housing information as decennial census SF3 data, as shown in Table 9.1. In addition, ACS also provides information related to fertility, veteran status, food stamp usage, health insurance, and computer and internet use.

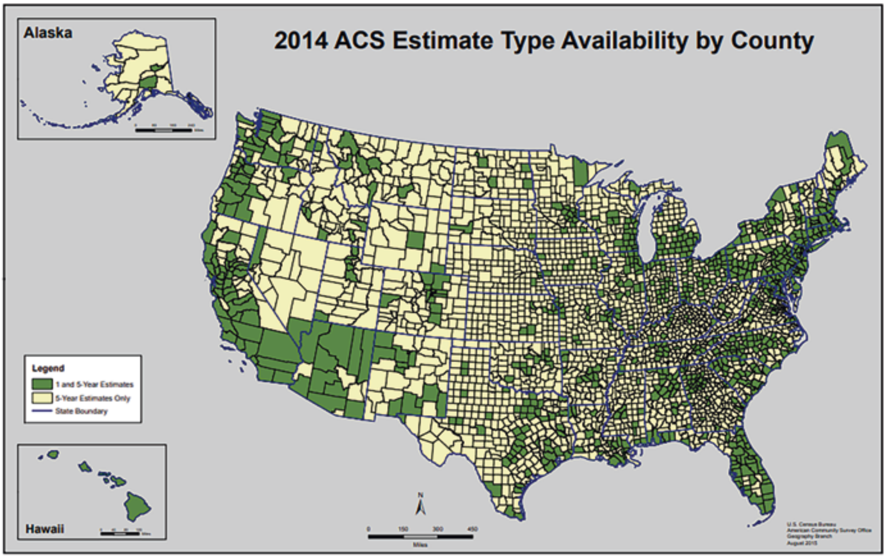


Fig. 9.1 ACS estimate type availability by county. (Source: The Census Bureau)

Describing the Population

Size

Population size is the most fundamental demographic attributes of a community. It is the number of people residing in an area at a time. For planners, the population size of a community provides fundamental information of planning activities in all sub-areas, such as education, health care, housing, and economic development. For instance, the 2018 populations of Kentucky and Florida are 4.461 million and 21.24 million, respectively. Consequently, Florida has a higher demand for public services, and a higher level of economic activities.

Composition and Characteristics

Composition describes characteristics of a population in terms of factors of interest, such as gender, age, race, occupation and so on. There are several common characteristics and measures for studying a population.

- Race and ethnicity

Race and ethnicity are considered separate and distinct identities. A ***race*** is a group of people with shared physical or social qualities and grouped into a distinct category by the society. Table 9.3 presents the region of origin of different races. An ***ethnicity*** reflects cultural factors, including nationality, regional culture, ancestry, and language. In the census data, there are two ethnicities: Hispanic or Latino and Non-Hispanic or Latino. A Hispanic or Latino is a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.

- Gender and Sex ratio

A ***sex ratio*** is the ratio of males to females. This is one of the key elements determining population natural growth, due to birth and death. For most communities, this ratio should be around 1, indicating similar amounts of males and females in the

Table 9.3 Regions of origin of races

White	Black or African American	American Indian or Alaska Native	Asian	Native Hawaiian or Other Pacific Islander
Europe	Africa	North America	Far East	Hawaii
Middle East		South America	Southeast Asia	Guam
North Africa		Central America	India	Samoa
				Pacific Islands

population. Table 9.4 presents 2018 sex ratios by age cohorts in the U.S. For a young population, the ratio tends to be over 1. This reflects the fact that the sex ratio at birth is male biased (Mathews & Hamilton, 2005). For older populations, the value of this ratio declines quickly, due to a higher mortality among males than among females. For age cohort 85 years+, males only count as 55% of females. There are many reasons for an abnormal sex ratio, which is a ratio much higher or lower than 1, for a community. Many of these reasons are about policies or activities what could attract gender- and age-biased migrations. For instance, a community that is attractive for elderly population (such as Marion County, FL) could have a lower sex ratio. A community with a specific type of economic opportunities may be more attractive to males, such as Santa Clara County, CA, where Silicon Valley is located. Such a community could have a higher than 1 sex ratio. An abnormal sex ratio could also reflect social problems, such as the increasing amount of single motherhood and teen pregnancy rates (Barber, 2000; Dollar, 2017).

- Dependency Ratios

Dependency ratios are important demographic concepts for studying the sustainability of an economy. According to people's economic productivity, a population can be divided into two groups: people who are in working age vs. who are not in. A person's working age sits between her adolescence and retirement. There is no consensus about the actual age range, as legal working age and recommended

Table 9.4 2018 sex ratios by age cohorts in the U.S.

	Male	Female	Sex ratio
Total population	161,118,151	166,049,288	0.97
Under 5 years	10,047,966	9,598,349	1.05
5–9 years	10,095,699	9,710,201	1.04
10–14 years	11,000,943	10,391,979	1.06
15–19 years	10,972,722	10,472,771	1.05
20–24 years	11,121,699	10,596,263	1.05
25–29 years	11,869,477	11,451,225	1.04
30–34 years	11,111,452	10,912,520	1.02
35–39 years	10,790,784	10,780,518	1.00
40–44 years	9,900,261	10,026,890	0.99
45–49 years	10,253,794	10,479,646	0.98
50–54 years	10,266,496	10,605,308	0.97
55–59 years	10,515,549	11,108,992	0.95
60–64 years	9,880,411	10,782,410	0.92
65–69 years	8,055,881	9,051,407	0.89
70–74 years	6,221,547	7,242,478	0.86
75–79 years	4,195,815	5,182,697	0.81
80–84 years	2,591,562	3,577,879	0.72
85 years+	2,226,093	4,077,755	0.55

Data source: Calculated based on 2018 ACS 1-year estimate retirement age varies across countries and over time. The World Bank uses 15–64 as the working age. Consequently, people aged 0–14 are children, aged 15–64 are in working age, and aged 65+ are retired. Based on these two groups of population, the following dependency ratios are defined,

$$Total\ Dependency\ Ratio = \frac{\begin{array}{c}Number\ of\ People\ aged\ 0\ to\ 14\\ +Number\ of\ people\ aged\ 65+\end{array}}{Number\ of\ people\ aged\ 15\ to\ 64} ?100$$

$$Children\ Dependency\ Ratio = \frac{Number\ of\ People\ aged\ 0\ to\ 14}{Number\ of\ people\ aged\ 15\ to\ 64} ?100$$

$$Aged\ Dependency\ Ratio = \frac{Number\ of\ people\ aged\ 65+}{Number\ of\ people\ aged\ 15\ to\ 64} ?100$$

A lower total dependency ratio indicates more working people to support the dependent population, allowing for better pension, health care, and education services. An increasing total dependency ratio implies a growing financial stress on the productive population for supporting the other two groups who are economically unproductive. Children and Aged Dependency Ratios further provide specific information about young and elderly population.

In the U.S., the Census Bureau uses 18–64 as the working age, because 18 is the typical age when high school students graduate and leave their parents' home and uses 65 as the full retirement age for all people. The projection of File and Kominski (2012) (Fig. 9.2) suggests that the total dependency ratio will increase from 62% in 1990 to 76% in 2060. As children dependency ratio (i.e. youth dependency ratio) declines from 42% to 37% and aged dependency ratio (i.e. old-age dependency) increase from 20% to 39%, the projection indicates that elderly population will be the source of the increasing dependency ratio.

- Population Pyramid

A ***population pyramid*** is a graphic presentation of the population of a region by age cohort and gender. It is also called age-sex pyramid. At the national level, with the presence of regulations on international migration, the shape of a population pyramid is influenced more by birth and death, and the population pyramid can provide useful information about future population growth trend. There are three types of population pyramids, expansive, constrictive, and stationary. As shown in Fig. 9.3, Kenya has an expansive population pyramid. With a large young population base, a researcher can conclude that this population is overall young and there are high birth rates. With these trends continuing, the population size will expand quickly. The U.S. has a stationary population pyramid. The distribution of population over age cohorts is basically equal. This population is stable, not growing and not declining. Germany offers an example of constrictive population pyramids. It

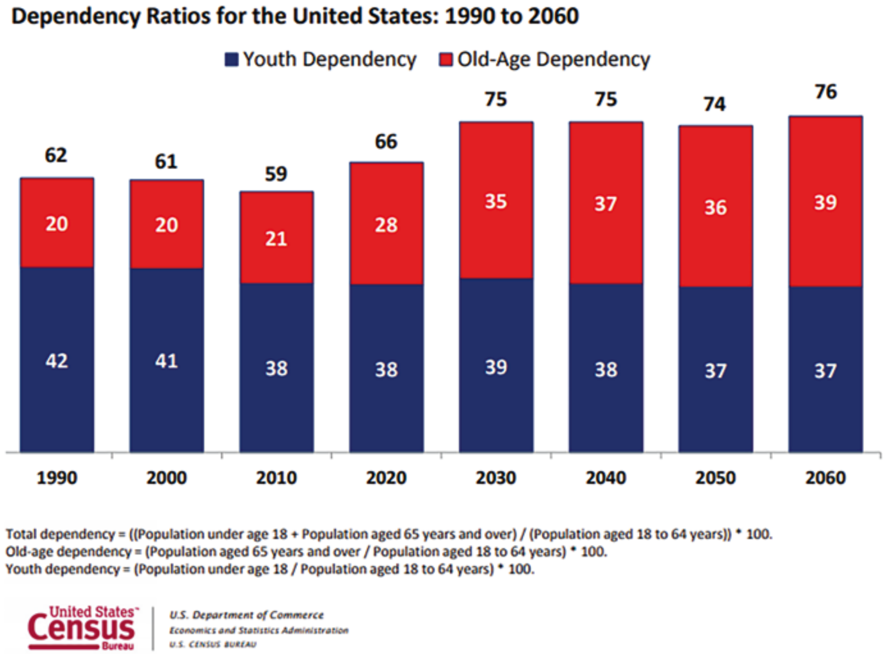


Fig. 9.2 U.S. dependency ratio projections. (Source: File and Kominski (2012))

has smaller percentages of people in the younger age cohorts. If there is no migration, this population will be elderly and shrinking.

For a county or a city, a population pyramid does not talk much about the future growth, because migration plays a more important role than birth and death activities in population dynamics. A population pyramid can inform birth and death information, which has more impacts on young and elderly age cohorts, but not migration information.

Nevertheless, a county's population pyramid is a useful tool for understanding the structure and composition of populations. Figure 9.4 presents four examples. Tompkins County, NY is the county where the Cornell University is located in. The population pyramid illustrates that the population sizes of age groups 15–19 and 20–24 are much larger than other age cohorts. This reflects two demographic attributes: the university population dominates the county population and most of the university population are 17–24 years old. New York County, NY contains Manhattan. For its population pyramid, age groups 25–29 and 30–34 stand out as the two groups with the most people. This is related to the nature of Manhattan's demographic mobility and economic structure, which greatly attracts young college graduates. Cook County, IL is the Central Business District (CBD) county of Chicago Metropolitan Statistical Area. Its population pyramid is more like the U.S. population pyramid (Fig. 9.3). Many CBD counties in the U.S. may take this shape, with a comparatively small young population base and an even population

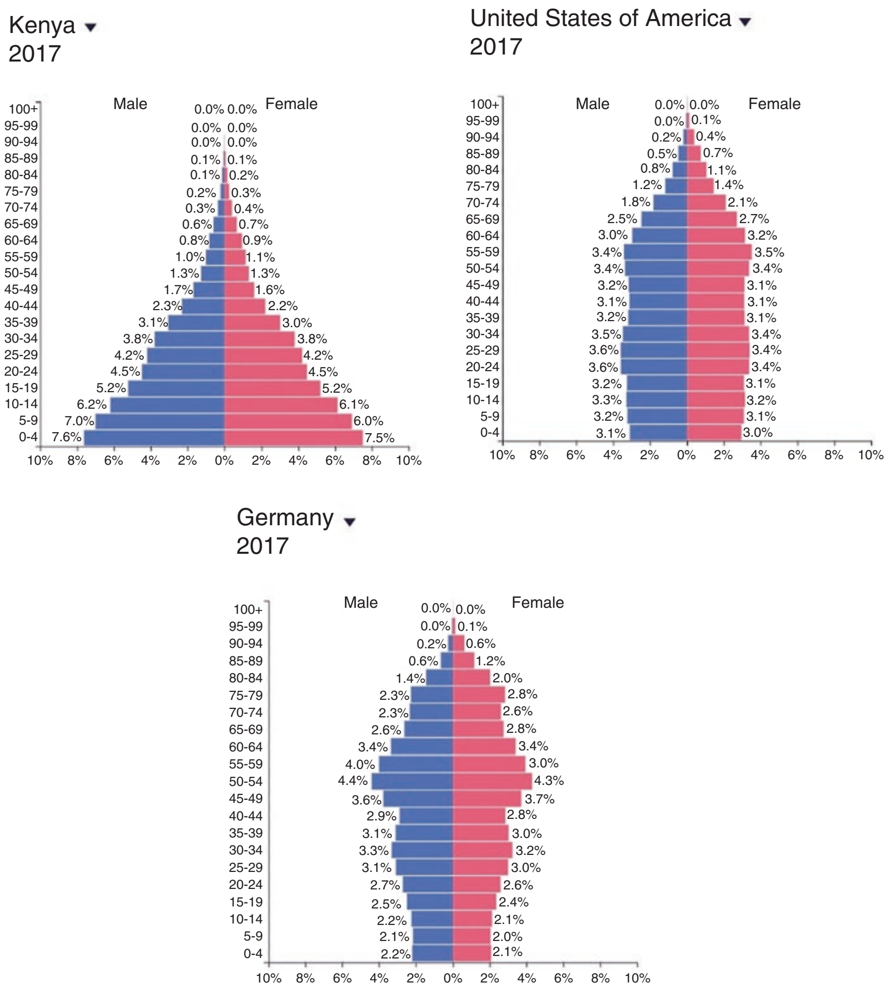


Fig. 9.3 Examples of population pyramids. (Source: https://www.populationpyramid.net)

distribution across young and mid-age age cohorts. McHenry, IL is a suburban county of Chicago Metropolitan Statistical Area. The population pyramid has a unique shape with two sets of dominate age groups: a set of young dependents and a set of mid-age working population. This reflects the demographic attributes of a suburban bedroom county that attracts wealthy traditional households with parents and children.

- Marriage

Marriage status provides useful information for analyzing a population. It determines the structure of a family. For instance, according to 2017 American Community Survey 5-year estimates, the marriage rate in Oldham County, KY, is

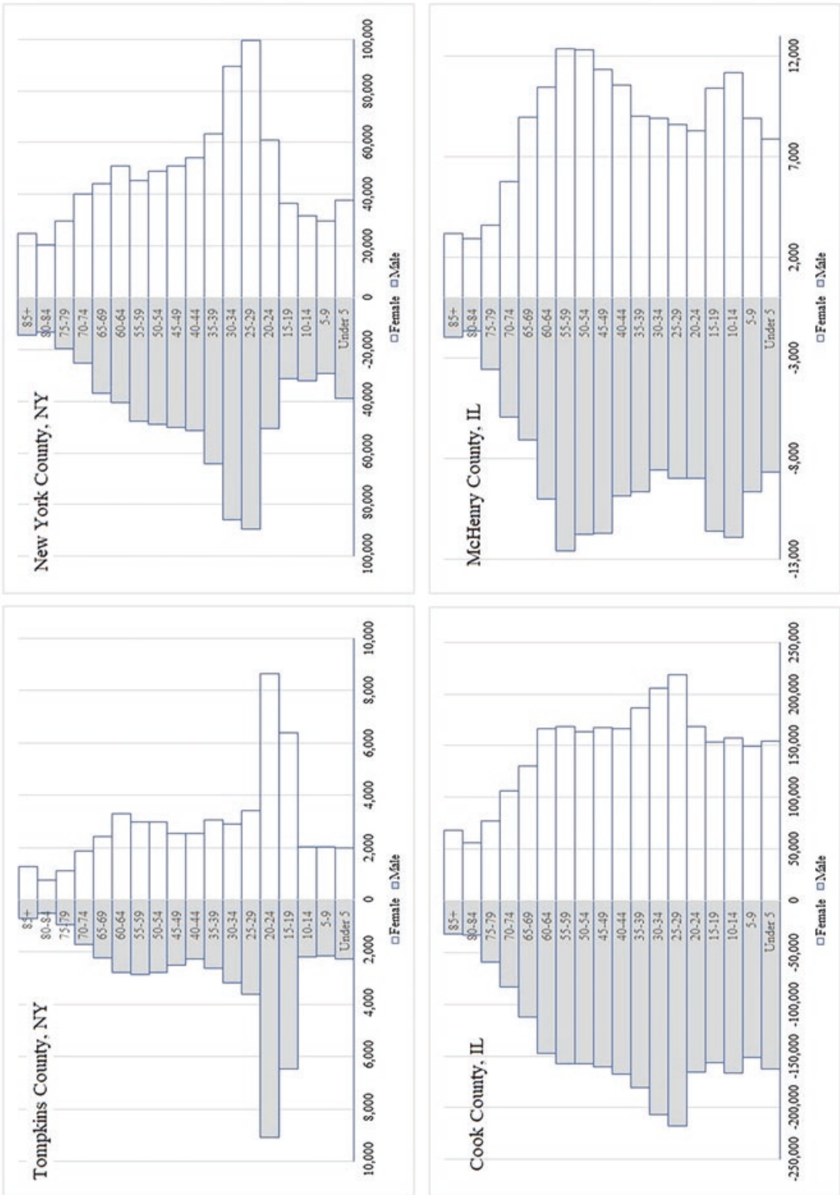


Fig. 9.4 County population pyramid examples. (Source: prepared by the authors based on 2014–2018 ACS 5-year estimates)

58%, and in Jefferson County, KY, is 43.6%. This is consistent with people's understanding of these communities. Oldham county is a suburban county for the Louisville Metropolitan Statistical area, and has more traditional families (i.e., parents with children). The decision about when to get married directly affects population growth that is related to birth. In the United States, men's age at first marriage has increased from 26.8 in 2000 to 29.8 in 2018. Women's age at first marriage increased from 25.3 to 27.8.

Population Changes

Population changes can be assessed by directly comparing population levels of two time periods. This is ***population growth*** and is defined as:

$$a = P_t \,?\, P_{t?1}$$

where P_t is population level at time t, P_{t-1} is population level at time t-1, and a is population growth, or population increment, from t-1 to t. Local officials may be more interested in population annual changes or ***annual increment***, which is calculated as:

$$aa = \frac{P_t \,?\, P_{t?1}}{y}$$

where y is number of years between time t-1 and t. A population growth or an annual growth can be a negative number, indicating that a decline occurred in the community. The absolute value of a population growth tends to be larger for an area with a sizeable population and smaller for an area with a small population. ***Growth rate*** and ***annual growth rate*** can account for this scale effect and are calculated as:

$$r = \frac{P_t \,?\, P_{t?1}}{P_{t?1}}$$

$$ar = \left(\frac{P_t}{P_{t?1}}\right)^{1/y} ?1$$

where r is the population growth rate from t-1 to t and ar is the annual growth rate. Let us work on an example. According to the 2010 Decennial Census and the 2018 American Community Survey, the U.S. population is 308,745,538 million in 2010 and 327,167,439 in 2018. What are the 2010–2018 population growth and growth rate, and what are annual growth and growth rate? 2010–2018 population growth is the difference between 2018 and 2010 population, i.e. a = 327,167,439 – 308,745,538 = 18,421,910. 2010–2018 growth rate can be calculated as:

$r = \frac{327167439 ? 308745538}{308745538} = 0.0597$, i.e., 5.97%. Annual population growth is: aa = 18,421,910/8 = 2,302,738. Annual population growth rate is: $ar = \left(\frac{327167439}{308745538}\right)^{1/8} ?1 = 0.0073$, i.e., 0.73%.

Distribution

Population distributions refers to spatial patterns about where people live. There are several common purposes of exploring a population distribution. First, activities are not evenly distributed, and planners need to identify concentration patterns such that infrastructure and service investments can follow up. A mapping analysis of population distribution or population density distribution can serve this purpose. As promoting equity being an important task for planning practitioners, it is important to explore whether population distribution patterns differ across different racial or income groups. A comparative analysis of population distributions of different groups helps address the equity consideration. At last, planners with specific interests, for example, health care or education, may be interested in a specific segmentation of a population. For them, the mapping analysis of population distribution should focus on the population segmentation of interest.

Technically, the analysis of a population distribution is about making a thematic map in a mapping software that can deal with geographic information. A ***thematic map*** displays information of a single subject or theme of all geographies within a region. For instance, Fig. 5.4 in Chap. 5 presents county level population density distribution of the U.S. At the national level, spatial patterns emerge, with higher levels of density in coastal areas and urbanized areas.

Selecting an appropriate geographical scale matters for a population distribution analysis. Popular geographic areas include state, county, census tracts, block groups, and blocks. An appropriate geographic scale is expected to be able to effectively uncover and present spatial patterns. For instance, for the U.S. population density distribution, a state level analysis will not uncover population concentrations as well as a county level analysis. A block group level analysis will encounter the issue of effective presentation for the whole nation on a piece of letter size paper.

Choosing a Benchmark

Choosing appropriate benchmark communities can help planners better assess the demographics of a community. The choice of benchmark communities should be based on similarity and/or the relations between the subject community and benchmark communities and depends on the purpose of the research. If the purpose of a demographic analysis is to understand the performance of the study area in a larger region, the larger region (e.g. a state or the nation) is a good choice for the benchmark comparison. Table 9.5 presents an example. Just presenting Jefferson County, KY data makes it difficult to assess the county. After bringing the State of Kentucky as a comparison, it is easier for the audience to understand the role of Jefferson County in Kentucky. It has a higher median income, a higher workers' wage, and a higher level of educational attainment. However, if the purpose of the research is to understand the competitiveness of Jefferson county in the country, it is a better choice to use a similar county. Hamilton, Ohio is the CBD county for Cincinnati Metropolitan Statistical Area (MSA), which is the closest MSA to the Louisville MSA and has a similar size of population and economy. By bringing Hamilton, OH as a benchmark, one will conclude that the growth of population and employment in Jefferson county is higher than Hamilton, OH. However, average wage and educational attainment in Jefferson County is behind Hamilton, OH. In practice, it is better to bring in more benchmark comparisons. For instance, when Jefferson County, KY, was preparing its 2040 Comprehensive plan, the planning department used 13 cities as comparison, including Austin, TX; Boston, MA, Boulder, CO, Cincinnati, OH, Denver, CO, Indianapolis, IN; Lexington, KY, Nashville, TN, Miami, FL, Oklahoma City, OK, Philadelphia, PA, Portland, OR, and San Antonio, TX.

Extrapolation Analysis

Population projections are the basis of planning and policy-making activities. Should a community expand the current sewage system, add a new school, or expand the roadways? How can a city government guide the city's future housing and economic development? All these rely on the projection of the future population.

Table 9.5 Selecting a benchmark comparison

	Jefferson, KY	Kentucky	Hamilton, OH
Median household income (2017 ACS 5-year average)	52,237	46,535	52,389
Bachelor's degree or higher (2017 ACS 5-year average)	33.90%	24.67%	38.11%
2010–2017 population growth	4.10%	3%	−0.60%
2016 Average workers' wage (County Business Patterns)	$49,167	$40,696	$56,651
2010–2016 employment growth (County Business Patterns)	10.22%	10.05%	4.93%

Estimates Versus Projections

Estimate and projection are two concepts that planners constantly deal with in their daily activities. An ***estimate*** is a value that is inferred for a population based on a set of sampling data. For instance, American Community Survey data is produced by sampling and provides population estimates for communities. Estimates provide information about current or past conditions of a population.

A ***projection*** focuses on the future level of activity. It is a conditional statement about the future based on assumptions. For instance, for a city, one can assume that the past growth pattern will continue in the future, and as a result, the future population can be derived. One also can assume that the future growth will slow down to only half of the past growth and provides another population projection. Klosterman (1990) further differentiates between a forecast and a projection. A forecast is based on personal judgements about what is likely to happen in the future, while a projection is based on an objective assumption. For instance, for a city's future population, the mayor may provide his forecast based on his knowledge about the city, while a planner may provide various projections based on different assumptions of growth trends, as introduced in following sections.

Extrapolation Versus Interpolation

Both extrapolation and interpolation are methods used to guess the value of a data point that does not exist. ***Interpolation*** is an estimation of a value within a set of values. For instance, before the American Community Survey became available, the Census Bureau only provided decennial population census data every 10 years. Figure 9.5 shows the decennial population for Jefferson County, KY over 1950–2010. If the county wants to know 2005 population, the county can make a reasonable guess by averaging 1990 and 2000 population. This is an example of interpolation to estimate a value based on neighboring values that provide the upper and lower bounds.

A policymaker is more likely to be interested in the level of future population. For instance, how to project 2020 and 2030 Jefferson county population based on historical population data. This is an example of extrapolation, i.e. guessing a value based on beyond an existing range of data.

Figure 9.6 offers a better explanation for the rationale of an extrapolation analysis. Given 1950–2010 U.S. historical information is available, what is the 2020 population? A simple line graph of the 1950–2010 population will be persuasive to argue that there was a clear line trend in decennial population levels. One can assume that this line trend will continue, and the 2020 population can be easily derived.

An ***extrapolation*** starts with an analysis of historical data to uncover trends or patterns. This is a curve fitting stage. With a stable trend confirmed, an analyst can

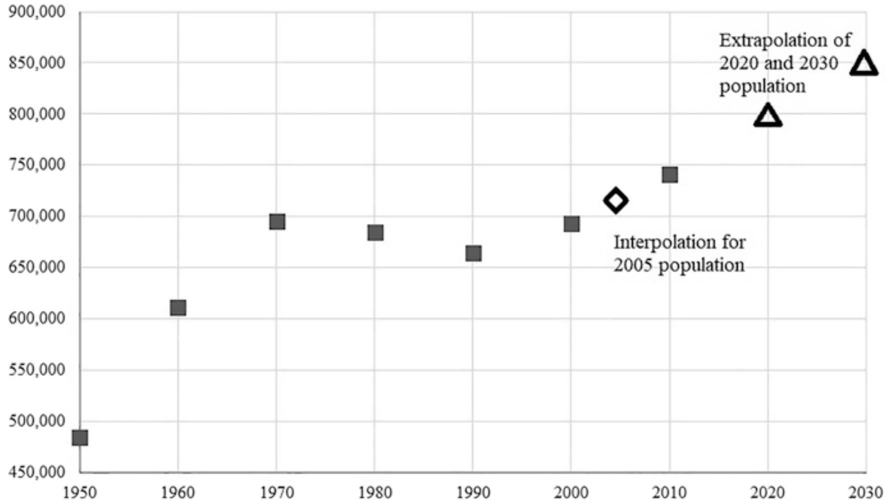


Fig. 9.5 Interpolation vs. extrapolation of Jefferson County, KY, population

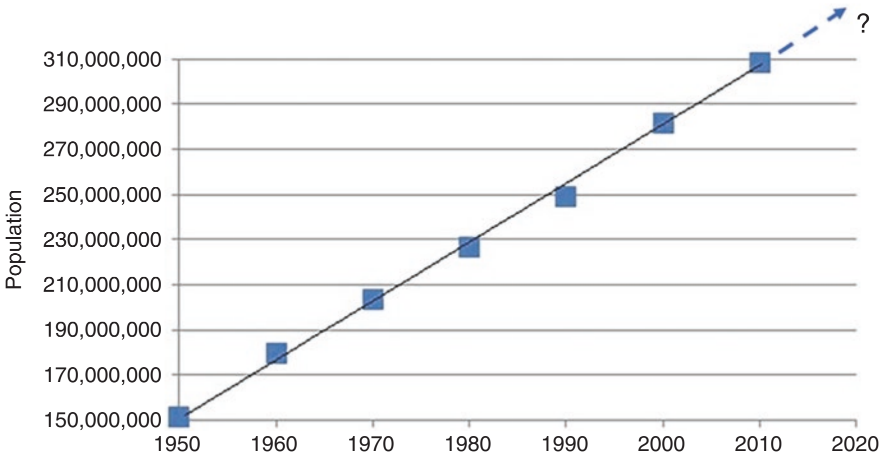


Fig. 9.6 1950–2010 U.S. population

proceed with the second step that is to assume the trend will continue in the future and to project the future population accordingly. Before we explore specific population growth trends, we first need to explain several basic concepts in population growth modelling.

Period and year are the first set of concepts of interest. Population data is often provided by year. However, it will be more convenient to replace year with period in a trend analysis for projecting future population. Table 9.6 presents 1950–2010 Decennial population data for Jefferson County, Kentucky. The first column presents year information and the second column presents period information. 1950 is

Table 9.6 Jefferson County, KY, Decennial population

Year	Period	Population
1950	0	484,615
1960	1	610,947
1970	2	695,055
1980	3	685,004
1990	4	664,937
2000	5	693,604
2010	6	741,069
2020	7	?

the earliest observation, and therefore, is period 0, i.e. the initial period. 1960 is period 1. Using the logic, 2020 is period 7. Time periods are set up by an analyst or a researcher. These are continuous integer numbers. Further, every period covers the same length of time. For instance, in Table 9.6, every period covers 10 years.

There are several notations to clarify. The tradition is to use P to indicate population, and a subscript number to indicate period. For Table 9.6, P_0= 484,615. This is the initial population, as suggested by the data, for this Jefferson County study, or the population at period 0. P_1= 610,947. This is the population at period 1. The objective of this study is to project P_7, which is the population at period 7, i.e., 2020 population. A more generalized notation P_t indicates population at period t.

Popular Growth Trends

The existing empirical studies of population growth uncover several popular growth trends, including linear growth, exponential, modified exponential, and logistic trends.

- The Linear Growth Trend

A ***linear growth model*** assumes that a population growth follows a linear growth pattern, i.e. a straight line population trend as shown in Fig. 9.8. For this growth trend, the amount of growth for each period is a fixed amount. As shown in the hypothetical example of Fig. 9.7, the population increases by 1000 over every period.

The linear growth model is to formulate the population at time t (P_t) as:

$$P_t = P_0 + t ? a$$

where a is the parameter of the model, which is the amount of population increment over a period. With a known, this model allows for the projection of population of any period (t). Let us use a hypothetical example to illustrate the linear growth model. Given that the current population for a city is 10,000 and the population will

t	P	a (growth)
0	10000	
1	11000	1000
2	12000	1000
3	13000	1000
4	14000	1000
5	15000	1000
6	16000	1000
7	17000	1000
8	18000	1000
9	19000	1000
10	20000	1000

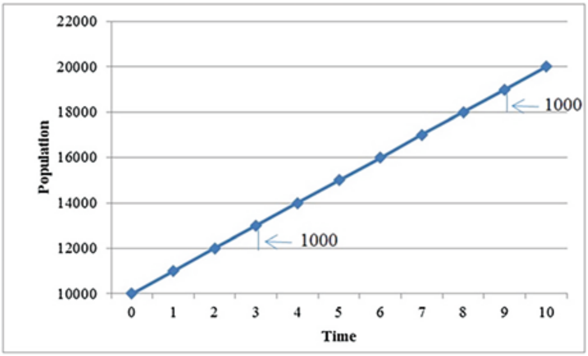


Fig. 9.7 A hypothetical linear growth example

t	P	Growth	Growth Rate
0	10000		
1	12000	2,000	20%
2	14400	2,400	20%
3	17280	2,880	20%
4	20736	3,456	20%
5	24883.2	4,147	20%
6	29859.84	4,977	20%
7	35831.81	5,972	20%
8	42998.17	7,166	20%
9	51597.8	8,600	20%
10	61917.36	10,320	20%

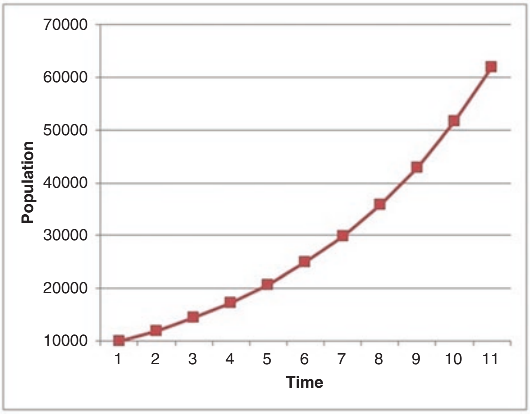


Fig. 9.8 A hypothetical exponential growth

increase by 1000 per year, project the population for 10 years from now. This example states that population increases by 1000 every year. This fixed increment suggests a use of the linear growth model with $a = 1000$. If we use current population as the initial population, with P_0=10,000, the population in 10 years is P_{10}, which can be calculated according to the linear growth model as: $P_{10} = P_0 + 10 * a = 10000 + 10 * 1000 = 20000$.

- The Exponential Growth Trend

If a population growth has a fixed growth rate over time, this growth follows an ***exponential growth model***. Figure 9.9 presents an ideal exponential growth, with a constant growth rate of 20% over every period. Exponential growth is not just in textbook. Many cities followed exponential growth during their early urbanization processes. Cook County, IL, the CBD county of Chicago, followed an exponential

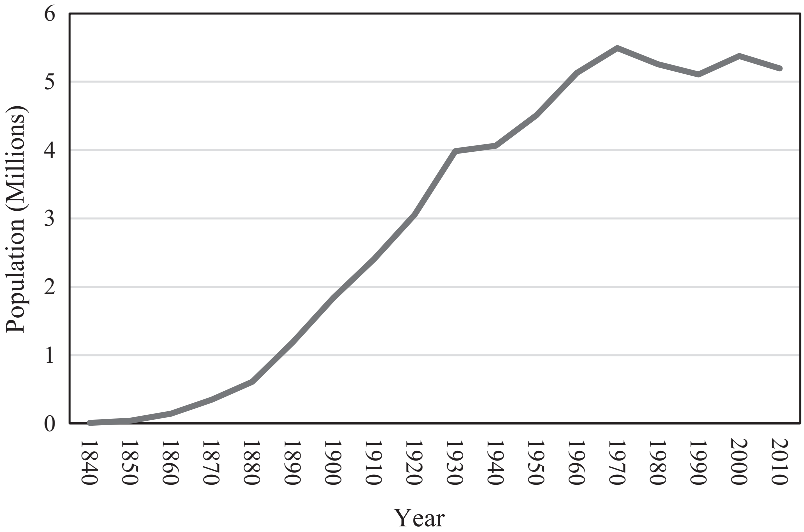


Fig. 9.9 1840–2010 population growth, Cook County, IL

growth over 1900–1930, with a constant 10-year growth rate around 30%. Harris County, TX (which is part of Houston) followed an exponential growth over 1930–1980, with a constant 10-year growth rate around 40–50%. The exponential growth model depicts an increasing level of growth over time. As shown in Fig. 9.8, 20% growth in period 10 is 10,320, as compared with the 20% growth in period 1, which is 2000.

The exponential model projects the population at time t (P_t) as:

$$P_t = P_0\, ?(1+r)^t$$

where r is the only parameter of the model. It is the growth rate over a period. P_0 is initial population at period 0. Let us illustrate the use of the exponential model through a hypothetical example. Given the current population for a city is 10,000 and the 5-year growth rate is 0.1, please use the exponential model to project the population in 10 years. In this example, the growth rate is given for 5 years, with $r = 0.1$. It is easier to use 5 years as one period. If the initial population is 10,000, $P_0 = 10{,}000$, then population in 10 years is the population of period 2, P_2. According to the exponential model, $P_2 = P_0 * (1 + r)^2 = 10000 * (1 + 0.1)^2 = 12100$.

- Modified Exponential Growth

Both linear and exponential growth models are unlimited population growth models. These models assume that the population of a county/state/country can continue to expand at a fixed growth or a fixed growth rate. They both ignore the accommodation capacity of a region. When the land use inventory of a county is running out, development activities will slow down and cease at the capacity. With an increasing level of population and population density, many urban issues, such as congestion, pollution, and higher level of competition for limited resources, emerge as deterrent factors that prevent further growth. Population of a region may stabilize around the capacity. For instance, Fig. 9.9 presents Cook County, IL population growth over 1840–2010. It seems that the growth after 1970 fluctuated around 5,200,000. This is very likely to be the accommodation capacity of Cook county given the current development intensity.

The ***modified exponential growth model*** introduces a growth capacity to accommodate the reality. Figure 9.10 illustrates an ideal modified exponential growth model, in which the growth capacity is known and is denote as C. There is a hidden characteristic for a modified exponential growth that is related to the remaining capacity of growth. For any period, the difference between the capacity and the population at that period (C - P_t) is the remaining capacity. If a growth follows a modified exponential model, the ratio of remaining capacities of any two consecutive periods is a constant, which is denoted as v. In Fig. 9.10, v is the ratio of B divided by A.

Following the modified exponential growth model, the population at time t can be projected as:

$$P_t = C ? v^t \left(C ? P_0\right)$$

where v is the ratio of remaining capacity of two consecutive periods, C is the growth capacity, and P_0 is population at time 0, i.e., the initial population. Let us have an example. The current population of a city is 10,000. It is known that the population will stabilize at a level of 30,000. The population in a year is expected to

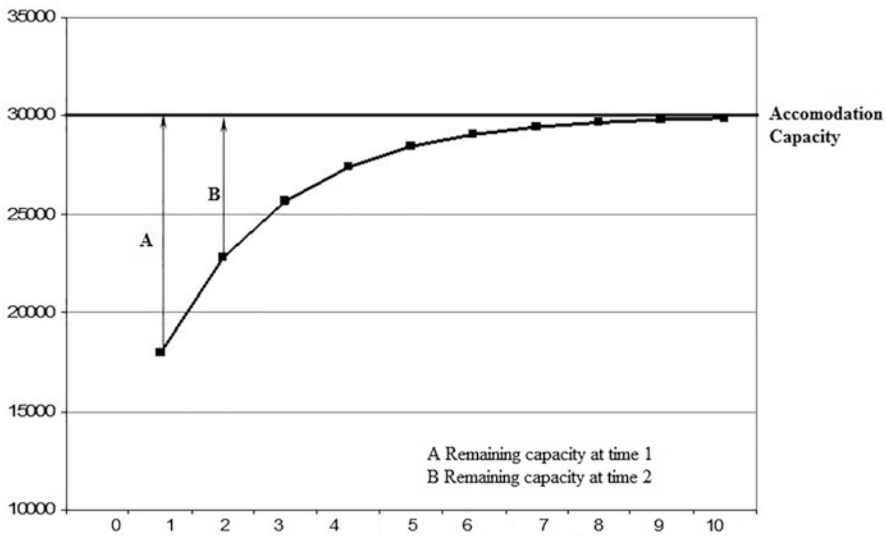


Fig. 9.10 An illustration of the modified exponential model

be 12,000. Please use the modified exponential model to project the population in 10 years. In this example, one period is a year. The initial population is 10,000, with P_0=10,000. The population in period 1 is 12,000, with P_1=12,000. The growth capacity is 30,000, with C = 30,000. With P_0, P_1, and C ready, the remaining capacities of periods 0 and 1 are 20,000 and 18,000, respectively. The parameter v is the ratio of these two remaining capacities, with $v = 18{,}000/20{,}000 = 0.9$. Population in 10 years is the population in period 10. It can be calculated as $P_{10} = C - v^{10}(C - P_0) = 30000 - 0.9^{10}(30000 - 10000) = 23026$.

A challenge of using the modified exponential model is to determine the growth capacity of a study area. There are several methods. One method is to study the land use inventory (see Chap. 12 for land use inventory analysis) in relation to the population level of a city. Assuming that the per capital land consumption remains stable, one can estimate the additional population growth that the region can accommodate based on the availability of developable land supply. Another method is to use an analog. Find a region that already reached its growth capacity and use that information for the study region.

- The Logistic Growth

The ***logistic growth*** accounts for the dynamics between population growth and resource limits during a three-stage urbanization process of initial, acceleration, and terminal phases (Mulligan, 2013). The trend displays a sigmoid curve, as shown in Fig. 9.11. In the initial stage of the urbanization, the population base is low, and this limits the level of population growth. With the expanding population base and plenty of resources, population grows and enters the acceleration phase. In this phase, the population expands rapidly. This increasing body of population exerts a pressure on limited physical and institutional resources, and the population growth

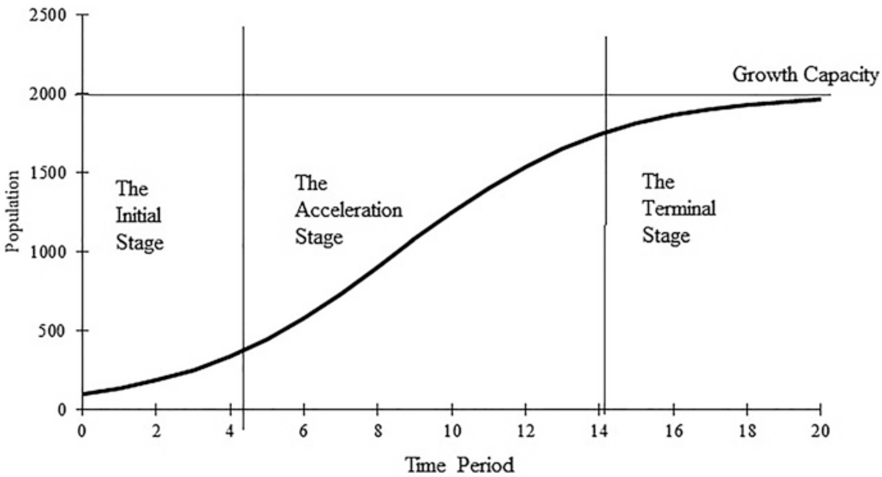


Fig. 9.11 The logistic growth curve

slows down and enters the terminal phase when the population level becomes stabilized around the holding capacity or the growth capacity.

There are various forms of the Logistic growth function (Klosterman, 1990; Oppenheim, 1980; Wang & Hofe, 2008). This book presents the Oppenheim version because it requires less parameters. The Logistic growth function is:

$$P_t = \frac{1}{\left(\frac{1}{P_0} ? \frac{b}{a}\right) e^{?at} + \frac{b}{a}}$$

where P_0 is the initial population, a and b are parameters. a controls the steepness of the curve, i.e., how fast the population grows over time. With a and b known, the population at time t (P_t) can be projected accordingly. There are two nice properties of this Logistic growth function. First, the ratio of a/b is the holding capacity. This function does not require figuring out the holding capacity, which is difficult to identify. Another property is that the population growth rate has a linear relation with the previous population level, with an intercept equal to a and a slope equal to b:

$$\frac{P_t ? P_{t?1}}{P_{t?1}} = a ? bP_{t?1}$$

As the population growing and approaching the growth capacity, the growth rate declines. This relation yields another function for projecting population of time t (P_t) based on the previous population, with

$$P_t = \left(1 + a ? bP_{t?1}\right) ? P_{t?1}$$

Using the second property, one can assess whether a population growth follows the Logistic trend, find the parameter values, and project future population. Figure 9.12 illustrates an example of the New York County, NY. Historical decennial population data are collected. The scatter plot suggests a linear relationship between growth rates and previous population levels. A regression analysis further uncovers that.

$$\text{Growth Rate} = 0.6137 ? 3\text{E} ? 07^{?}\ \text{Previous Population},$$

and the R-square is 0.7714. This regression analysis justifies the use of the Logistic growth trend for analyzing New York County population. Further, the two parameter values are a = 0.6137 and b = 3E-07. With these two parameters available, the growth capacity can be easily calculated as a/b = 0.6137/3E-07 = 2,045,667. 2020 population can be projected based on 2010 population, as:

$$\text{P}_{2020} = \left(1 + 0.6137 ? 3\text{E} ? 07^{?} 1{,}537{,}195\right)^{?} 1{,}537{,}195 = 1{,}771{,}681.$$

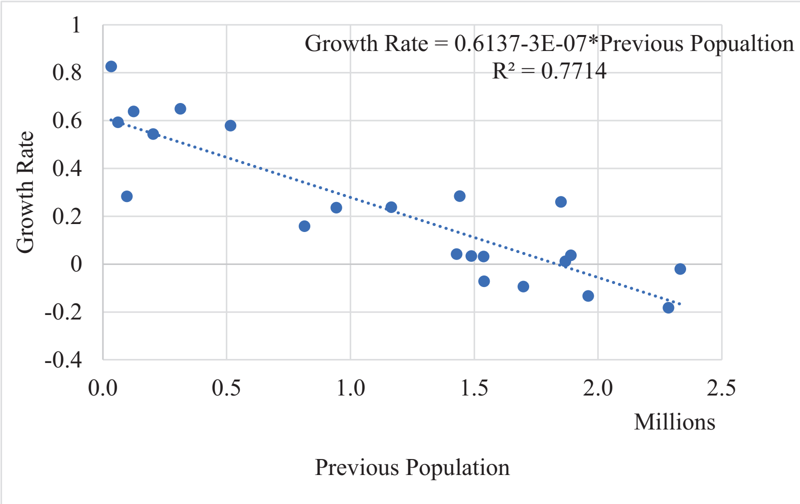


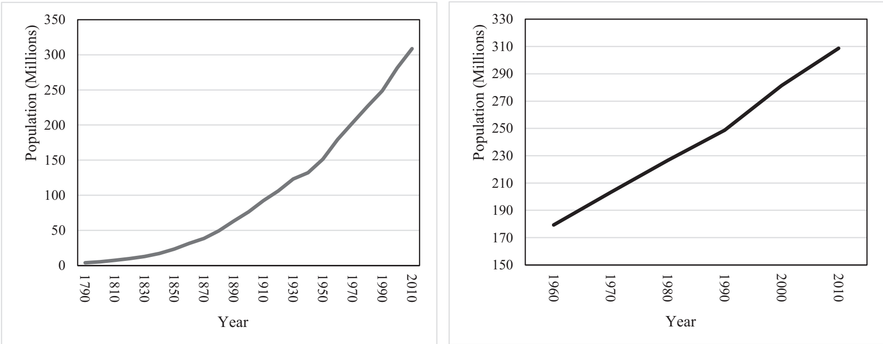
Fig. 9.12 Growth rate and previous population, New York County, NY

The 2030 population can be projected based on the 2020 population.

- Choosing the Right Growth Trend

The reliability of a population projection by using an extrapolation method relies on the ability of a planner to choose an appropriate growth model. To do so, an exploration of the historical trend is necessary. With the assumption that this historical growth will continue in the future, one can project future population. With a more stable historical trend, the population projection tends to be more reliable. Graphic analysis in Excel is a tool to uncover a historical growth trend. An appropriate growth model should be able to explain a region's historical trend.

It is critical to choose an appropriate time coverage for a historical trend analysis. Figure 9.13 presents trend analysis for U.S. population. If one decides to study the 1790–2010 trend, she may conclude that U.S. population growth has followed an exponential growth, and this exponential growth may continue in the future. However, if the decision to study the 1960–2010 trend, the linear growth model will be the one for explaining the population growth. The rationale of extrapolation analyses is related to the continuities in socio-economic attributes and policies of a region. A historical population growth is an outcome of the interactions of many local and national factors, such as economic policies, infrastructure conditions, amenities, and the quality of local labor. If there are no dramatic changes in these factors, the past growth trend may continue in the future. As for the example of projecting 2020 or 2030 U.S. population, 1960–2010 is a better choice than 1790–2010. Factors that influenced population growth in the nineteenth century and the early twentieth century play litter roles in the future growth. Determinants for urban growth have changed dramatically with the introduction of a national highway network in the 1950s.



1790-2010 1960-2010

Fig. 9.13 U.S. population growth trend analysis (1790–2010 vs. 1960–2010)

The reliability of a population projection also relies on local stability, i.e., no dramatic changes in the economic and political-administrative environments of a community. For instance, Arlington County, VA, contains Crystal city, which was selected for Amazon's second headquarter in 2017. It will be very challenging to use extrapolation methods to project the future populations of Arlington County and Crystal City. The reason is that because of the Amazon investment, which changes the local economic environment, the county's and the city's growths are going to depart from their historical trends. A good planner should bring their local knowledge (for example, possible policies in the future, and negotiations between local governments and Amazon) into analyses, and this will increase the reliability of the projection.

The Cohort-component Method

When analyzing or projecting the population of a community, one may explore different factors contributing towards the growth or decline of a local population, such as local policies that can attract people, people's lifestyle and preference, the local medical system, or local amenities. Regardless, all possible factors related to population change are influencing future population through three mechanisms: birth, death, and migration. In a short period, birth, death, and migration activities may remain stable. For instance, the New York city is dominated by young working population (Fig. 9.4). The existing demographic and economic structures may persist in the next 5 years. Therefore, future migration patterns and families' reproductive decisions may be very similar to nowadays. Given the same healthcare system, the current death pattern may continue in the future. If one knows how many new births there will be, how many people will die, how many people will move into and move out of the region in the future period, one can know the future population size. Cohort-component model is a demographic method that projects population by decomposing future growth into three components, the mortality (death) component, the fertility (birth) component, and the migration component.

Age Cohort

To conduct a *cohort-component analysis*, one first needs to divide a population into age cohorts. Table 9.7 presents several examples of age cohort designs. There are three criteria for setting up age cohorts. First, all age cohorts must have the same size, except for the eldest age cohort, which has an open age range. In design 1, the size of all age cohorts is 20 years, except for the 80+ age cohort. The size is 10 years in design 2 and 5 years in design 3. Second, age cohorts should be mutually inclusive and collectively exhaustive. Age cohorts should cover all possible ages. A person belongs to one and only one age cohort. At last, the age cohort size should be consistent with the study period of an analysis. For instance, if one adopts age cohort design 1 to analyze the U.S. population growth, she needs to study 20-year growth. She will study the birth, death, and migration activities in the past 20 years, make assumptions about the birth, death, and migration activities in next 20 years, and project the population in 20 years. However, if one uses age cohort design 3, consistent with the age cohort size of 5 years, she must study activities by 5 years, and make projections accordingly. The reason will be explained in the next section.

The Mortality Component

Mortality is death. It is a comparatively independent demographic event influenced by individual biological attributes and exogenous factors such as the quality of the medical system, general condition of nutrition, sanitation of the built environment, war, outbreak of new diseases, and so on. For a time when there are no significant changes in the external environment, we can argue that mortality is more influenced by individual attributes. The mortality rate of an age cohort tends to be stable. For instance, the death rate of the 65–69 age cohort in the U.S. was 1.49% in 2016. The annual death rate is expected to remain stable in the following years for this age cohort.

Table 9.7 Age cohort design examples

Design 1	Design 2		Design 3			
0–19	0–9	50–59	0–4	25–29	50–54	75–79
20–39	10–19	60–69	5–9	30–34	55–59	80+
40–59	20–29	70–79	10–14	35–39	60–64	
60–79	30–39	80+	15–19	40–44	65–69	
80+	40–49		20–24	45–49	70–74	

- Mortality Rates

There are several mortality rates of interest for planners. ***Crude mortality rate*** is the number of deaths during a year for every 1000 people. It includes deaths from all causes. Crude mortality rate is an indicator of the general health condition of a population and the quality of the region's healthcare system.

Infant mortality rate is another general health and social well-being indicator. It is the number of infant deaths during a year for every 1000 live births. An infant death occurs to an infant before the first birthday. With improving healthcare system and quality of life, the infant death rate will decrease. The U.S. infant mortality rate was 5.7 infant death per 1000 live births in 2018.

The risk of death varies with demographic attributes, such as race, age, and gender. A crude mortality rate or an infant mortality rate cannot address such variations. ***Age specific mortality rates*** (ASMRs) consider differences across population age cohorts and provide critical information for identifying population segmentations at high risk and in need of health services. Figure 9.14 presents U.S. ASMRs between February 1, 2020 and October 31, 2020, the COVID-19 pandemic time. The high mortality rates of elderly population suggest identifying elderly population groups as high risk groups. For an age cohort, the mortality rate can be calculated as number of deaths within the age cohort divided by the number of people in this age cohort.

- Mortality and Population Projection

The impact of mortality on a population is to decrease the size of the population. Mortality is related to the survival during the aging process of a population. 2018 U.S. mortality rate is 8.58 per 1000 people. This implies that the 2019 survival rate is 991.42 (1000–8.58) per 1000 people. The event of surviving links age cohorts over time. Table 9.8 uses Jefferson County, KY to illustrate how the mortality

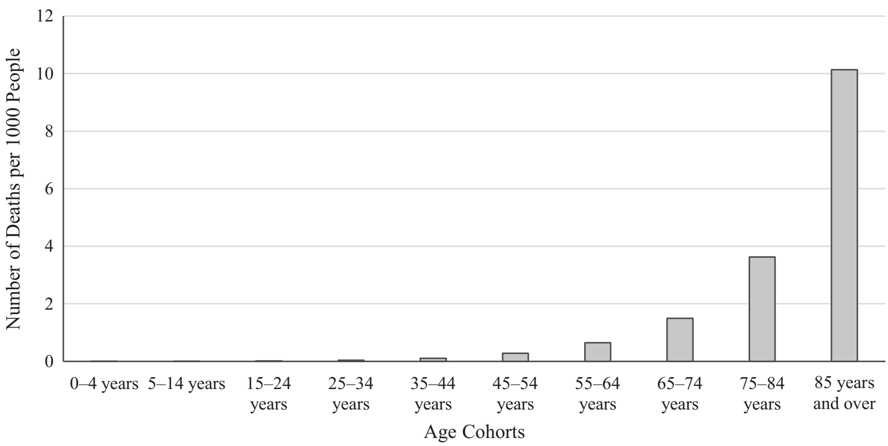


Fig. 9.14 Age specific mortality rates (ASMRs) in United States, 2/1/2020–10/31/2020. (Data source: The CDC)

Table 9.8 The impacts of mortality on population dynamics – Jefferson County, KY

Age Cohort	2010 Population	# of Death 2010-2014	# of Survivors	2015 Population Related to the Mortality Component
Under 5	48,634	411	48,223	0
5-9	47,238	29	47,209	48,223
10-14	47,444	44	47,400	47,209
15-19	47,427	136	47,291	47,400
20-24	48,887	307	48,580	47,291
25-29	53,848	330	53,518	48,580
30-34	50,435	466	49,969	53,518
35-39	48,529	486	48,043	49,969
40-44	47,401	677	46,724	48,043
45-49	55,008	1116	53,892	46,724
50-54	55,749	1806	53,943	53,892
55-59	49,937	2392	47,545	53,943
60-64	41,464	2649	38,815	47,545
65-69	28,926	2833	26,093	38,815
70-74	22,136	3063	19,073	26,093
75-79	18,527	3811	14,716	19,073
80-84	15,452	5125	10,327	14,716
85+	14,054	10667	3,387	13,714

component affects the population in the next period. The survived population of an age cohort moves to the next age cohort over time. For instance, 48,223 people with age younger than 5 years old survived over 2010–2014. In 2015, all these people became 5–9 years old, and therefore, completely moved to the next age cohort of 5–9. Similarly, people in other age cohorts were moving to the next older age cohort, except for the people in the oldest age group, 85+. In 5 years, if they survived, they remained in the 85+ age cohort. The youngest population age cohort, under 5, did not benefit from the mortality component. With all population moving to the next age cohort during the aging process, the mortality component zeroed out the population of age cohort under 5. These dynamics related to mortality enable an analysis of the expected population sizes of all age cohorts due to the mortality component.

For a successful analysis of the impacts of mortality on population dynamics, age cohort size must be consistent with population dynamics. In the example of Table 9.6, the age cohort size is 5 years. Therefore, analyses should focus on every 5 years, i.e., 2010, 2015, 2020, 2025 and so on. If age cohort size and population dynamics are inconsistent with each other, population of an age cohort may not completely move into the next age cohort.

The Fertility Component

Fertility is related to birth activities. A high level of births directly causes a population to grow. Childbearing decisions can be influenced by many factors, such as marital status, employment status, women's education level, cultural background, and so on.

- Birth Rates

Birth rates are important measures in urban demography. They directly contribute toward population growths and affect public programs and budgets for education and health systems. There are several common birth rate measures.

The ***crude birth rate*** (CBR) is the number of live births every 1000 population during a specific period which is often a year. It is a common indicator for fertility.

To address different fertility across different population segmentations by age, ***age specific birth rates*** (ASBRs) are calculated as the number of live births from an age cohort divided by the number of people in this age cohort. Figure 9.15 presents 2007, 2014, and 2016 U.S. ASBRs. ASBRs vary significantly across age cohorts but tend to be stable within the same age cohort. Over 2007–2014, ASBRs declined for young age cohorts (15–19, 20–24, and 25–29) and slightly increased for the other older age cohorts.

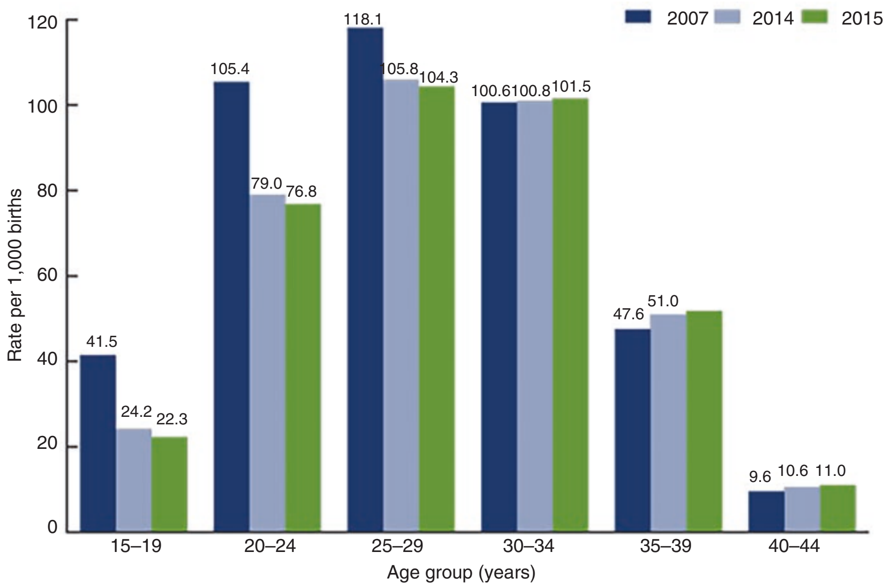


Fig. 9.15 Age-specific birth rates, by age of mother: United States, 2007, 2014, and 2015. (Source: The CDC. https://www.cdc.gov/nchs/products/databriefs/db258.htm. Visited on December 17, 2020)

The ***total fertility rate*** (TFR) is the number of children that a woman would have during her lifetime. It is an indicator for the sustainability for a population. A TFR of 2.1, i.e., every woman having 2.1 children, represents a replacement fertility for stabilizing a population. For a developing country with high mortality, the replacement fertility rate can be higher.

- Fertility and Population Dynamics

How does the fertility component contribute towards the future population? As for the Jefferson County, KY, for example, as shown in Table 9.9, this is equivalent to asking which age cohort those babies born over 2010–2014 will join in 2015? All 2010–2014 births constitute the youngest age group in 2015. Fertility determines the size of the youngest population cohort in the next period, and indirectly affects the sizes of older age cohorts in future periods through the surviving process.

The Migration Component

Migration refers to human movements from one location to another. ***In-migration*** is the amount of people moving from outside into a region/city and it increases the local population. ***Out-migration*** is the amount of people moving out of a region/city, and the local population decreases accordingly. ***Net migration*** is simply in-migration minus out-migration. Migration rate for an age cohort is the size of net

Table 9.9 The impacts of fertility on population dynamics, Jefferson County, KY

Age Cohort	2010 Population	# of Birth 2010-2014	2015 Population Related to Birth
Under 5	48,634	0	49538
5-9	47,238	0	0
10-14	47,444	49	0
15-19	47,427	4022	0
20-24	48,887	12335	0
25-29	53,848	14353	0
30-34	50,435	12413	0
35-39	48,529	5093	0
40-44	47,401	1119	0
45-49	55,008	115	0
50-54	55,749	39	0
55-59	49,937	0	0
60-64	41,464	0	0
65-69	28,926	0	0
70-74	22,136	0	0
75-79	18,527	0	0
80-84	15,452	0	0
85+	14,054	0	0

migration of this cohort divided by number of people in this age cohort. The growth of a county/city largely depends on whether the city/county can attract people. A city's economic and social development depends on what type of population the city can attract.

The migration of an age cohort directly affects the same age group and causes migratory growth or decline. Table 9.10 illustrates an example about how the migration component affects the age-specific population cohorts in the upcoming period. Assuming that in the last day of 2014, one counts the size of migration over 2010–2014 for every age cohort (as shown in column Migration 2010–2014). These migrations directly affect their corresponding age cohorts. For instance, for age cohort 5–9, because of the migration component, the population of this age cohort declined by 1470 in 2015.

Figure 9.16 summarizes the idea of a cohort component analysis. It divides a city's population into *n* equal-sized age cohorts and studies population growth over a period that is consistent with the age cohort size. Age cohort 1 is the youngest age group and age cohort *n* is the oldest. With the mortality component, the analysis captures the aging process, during which people in a younger age cohort is surviving into the older age cohort, except for the people in the oldest age cohort. The fertility component rejuvenates the youngest age cohort and determines the population size of the youngest age cohort in the next period, with the births from all age cohorts constituting the youngest age group. Migrants directly join or leave their corresponding age cohorts, causing the population cohorts to increase or decline.

Table 9.10 The impacts of migration on population dynamics, Jefferson County, KY

Age cohort	2010 population	Migration 2010–2014	2015 population related to migration
Under 5	48,634	−797	−797
5–9	47,238	−1470	−1470
10–14	47,444	270	270
15–19	47,427	−497	−497
20–24	48,887	497	497
25–29	53,848	10,335	10,335
30–34	50,435	30	30
35–39	48,529	−3013	−3013
40–44	47,401	314	314
45–49	55,008	2072	2072
50–54	55,749	−3233	−3233
55–59	49,937	268	268
60–64	41,464	1884	1884
65–69	28,926	1096	1096
70–74	22,136	1123	1123
75–79	18,527	−295	−295
80–84	15,452	−361	−361
85+	14,054	2843	2843

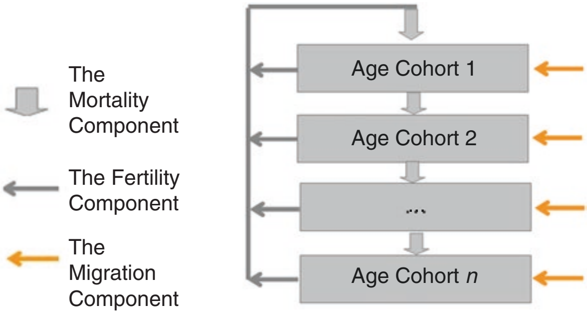


Fig. 9.16 Impacts of fertility, mortality, and migration on population dynamics

Set Up a Cohort-Component Model

To use the Cohort-Component method to project a future population, one needs to determine what are age-specific survival, birth, and migration rates for a projection period. With this rate information available, one can calculate the expected amount of people who will survive, the new births, and the level of migration that will take place. Future population can be projected accordingly. Generally, a cohort-component analysis has four steps.

Step 1: Collect available population by age data, as well as birth, death, and migration by age data for the most recent years. Age specific population data is available from the Decennial Census and the American Community Survey. Birth and death data are vital statistics. All births and deaths have to be registered with state governments. All states operate health statistics programs and birth and death information is available through these programs. The American Community Survey asks whether a respondent lived in the same residence 1 year ago. Based on this information, the Census Bureau compiles county to county migration flows, which enables the calculation of the in-, out-, and net-migration of a county.

Step 2: Decide on an age cohort size. Table 9.5 provides three examples of 5 years, 10 years, and 20 years. Birth, death, and migration dynamics tend to be less volatile for a smaller age cohort. For instance, the variance of the probability of death is lower in the 50–54 age cohort than in the 50–59 age cohort. In this sense, population dynamics can be more accurately captured by using a smaller-sized age cohort. The most used age cohort size is 5 years. The reason is that the Census Bureau provides population by 5-year age cohorts. The age cohort size determines the size of a study period. If a study adopts 5-year age cohorts and if 2015 is the base year, the projections will be for years 2020, 2025, 2030, and so on.

Step 3: Calculate birth, survival, and migration rates over the most recent study period for every age cohort. As shown in Table 9.11, given the age cohort size is

Table 9.11 Calculating birth, survival, and migration rates, Jefferson County, KY

Age cohort	2010 population	# of death 2010–2014	# of birth 2010–2014	Migration 2010–2014	Five year survival rates	Five year birth rates	Five year migration rates
Under 5	48,634	411	0	−797	0.9915	0.0000	−0.0164
5–9	47,238	29	0	−1470	0.9994	0.0000	−0.0311
10–14	47,444	44	49	270	0.9991	0.0010	0.0057
15–19	47,427	136	4022	−497	0.9971	0.0848	−0.0105
20–24	48,887	307	12,335	497	0.9937	0.2523	0.0102
25–29	53,848	330	14,353	10,335	0.9939	0.2665	0.1919
30–34	50,435	466	12,413	30	0.9908	0.2461	0.0006
35–39	48,529	486	5093	−3013	0.9900	0.1049	−0.0621
40–44	47,401	677	1119	314	0.9857	0.0236	0.0066
45–49	55,008	1116	115	2072	0.9797	0.0021	0.0377
50–54	55,749	1806	39	−3233	0.9676	0.0007	−0.0580
55–59	49,937	2392	0	268	0.9521	0.0000	0.0054
60–64	41,464	2649	0	1884	0.9361	0.0000	0.0454
65–69	28,926	2833	0	1096	0.9021	0.0000	0.0379
70–74	22,136	3063	0	1123	0.8616	0.0000	0.0507
75–79	18,527	3811	0	−295	0.7943	0.0000	−0.0159
80–84	15,452	5125	0	−361	0.6683	0.0000	−0.0234
85+	14,054	10,667	0	2843	0.2410	0.0000	0.2023

5 years, one should find 5-year death, birth, and migration data, and calculated rates accordingly.

Step 4: Introduce the base population for the projection, assume that future birth, survival, and migration rates are the same as the past rates, and calculate the expected number of survivors, birth, and migration for every age cohort. Project age-specific population for the next period, following the following rules:

- All the expected births go to the youngest age cohort;
- The expected number of survivors join the next older age cohort, except for the survivors of the oldest age cohort, who remain in the oldest age cohort; and
- The expected migration of an age cohort directly affects the population level of this age cohort.

Table 9.12 provides an example of projecting 2020 population for Jefferson County, KY, based on 5-year survival, birth, and migration rates that have been calculated previously. To project 2020 population with 5-year age cohort set up, one needs to use 2015 population as the base period. Expected number of survivors, births, and migration are calculated accordingly for each age cohort. With all the births joining the youngest age cohort in 2020 and the migration of this youngest age cohort directly affecting the population level, 2020 population of under 5 age cohort is projected as 50,375. The oldest age group is affected by its expected migration and the expected number of survivors from this age cohort and the previous younger age

Table 9.12 2020 population projection, Jefferson County, KY

Age cohort	2015 population	Five year survival rates	Five year birth rates	Five year migration rates	Expected # of survivors 2015–2019	Expected # of births 2015–2019	Expected migration 2015–2019	2020 population
Under 5	48,741	0.9915	0.0000	−0.0164	48,329	0	−799	50,375
5–9	46,753	0.9994	0.0000	−0.0311	46,724	0	−1455	46,874
10–14	47,479	0.9991	0.0010	0.0057	47,435	49	270	46,994
15–19	46,903	0.9971	0.0848	−0.0105	46,769	3978	−492	46,943
20–24	47,788	0.9937	0.2523	0.0102	47,488	12,058	486	47,254
25–29	58,915	0.9939	0.2665	0.1919	58,554	15,704	11,308	58,796
30–34	53,548	0.9908	0.2461	0.0006	53,053	13,179	32	58,586
35–39	46,956	0.9900	0.1049	−0.0621	46,486	4928	−2915	50,138
40–44	48,357	0.9857	0.0236	0.0066	47,666	1142	320	46,806
45–49	48,796	0.9797	0.0021	0.0377	47,806	102	1838	49,504
50–54	50,659	0.9676	0.0007	−0.0580	49,018	35	−2938	44,868
55–59	54,211	0.9521	0.0000	0.0054	51,614	0	291	49,309
60–64	49,429	0.9361	0.0000	0.0454	46,271	0	2246	53,860
65–69	39,911	0.9021	0.0000	0.0379	36,002	0	1512	47,783
70–74	27,216	0.8616	0.0000	0.0507	23,450	0	1381	37,383
75–79	18,778	0.7943	0.0000	−0.0159	14,915	0	−299	23,151
80–84	14,355	0.6683	0.0000	−0.0234	9594	0	−335	14,580
85+	16,557	0.2410	0.0000	0.2023	3990	0	3349	16,933

cohort. Consequently, 2020 population for 85+ age cohort is 16,933 (3349 + 9594 + 3990). For all other age cohorts, their populations are determined by the expected number of survivors from the younger population cohort and the expected number of migration of the same age cohort. For instance, 2020 population of 25–29 is projected as 58,796, which includes the expected number of survivors from the 20–24 age cohort (47,488) and the expected number of migration of age cohort 25–29 (11,308).

Population Projection Quality Assessment

All population projection methods have assumptions. The linear growth model assumes a constant growth. The exponential model assumes a constant growth rate. The modified exponential model assumes a constant ratio of the remaining growth capacity of two time periods. The logistic model assumes a relationship between growth rate and previous population level. The cohort-component model assumes that past birth, survival, and migration rates continues in the future. The actual population growth always departs from these ideal patterns. Projection errors exist. It is important to understand how to measure projection errors and what factors influence projection errors.

One can assess a projection error (E) by comparing a population projection (PP) with the actual population (AP), as $E = PP - AP$. This projection error can be a positive or a negative number, with a positive number indicating a situation of over estimation and a negative number indicating under estimation. One also can assess a percent error (PE). It is the error (E) divided by the actual population (AP), as $PE = (E / AP) * 100\%$. A percent error is a relative error. It normalizes different units into the same scale and enables an assessment of projection quality of population projections for regions with different population sizes.

Rayer (2008) applied different extrapolation methods and the cohort-component methods to project the 1960–2000 population for 2482 U.S. counties with no boundary changes in the past century. The projections were compared with their actual observations to assess quality. This research uncovered several factors that could affect the quality of a population projection. First, for all methods, projection error increases with projection horizon. A projection horizon is the time interval between the base year and the target year for a projection. For an extrapolation analysis, the base year is the initial year (P_0). For the cohort component analysis, the base year is the year for the base population, which is used for making projections, i.e. column 2015 population in Table 9.12. Second, for all methods, projection error tends to be lower when a population is more stable, i.e. with the growth rate close to 0. Third, for all methods, projection error decreases with an increasing population size. Population projections for large cities tend to be more reliable. At last, as for which method yields the most reliable projection, Rayer suggested applying different methods and using the average of the projections.

References

Barber, N. (2000). On the relationship between country sex ratios and teen pregnancy rates: A replication. *Cross-Cultural Research, 34*(1), 26–37.

Dollar, C. B. (2017). Sex ratios, single motherhood, and gendered structural relations: Examining female-headed families across racial-ethnic populations. *Sociological Focus, 50*(4), 375–390.

File, T., & Kominski, R. (2012). *Dependency ratios in the United States: A state and metropolitan area analysis*. Data from the 2009 American Community Survey.

Klosterman, R. E. (1990). *Community analysis and planning techniques*. Rowman & Littlefield.

Mathews, T. J., & Hamilton, B. E. (2005). Trend analysis of the sex ratio at birth in the United States. *National Vital Statistics Reports, 53*(20), 1–17.

Mulligan, G. F. (2013). Reprint of: Revisiting the urbanization curve. *Cities, 32*, S58–S67.

Oppenheim, N. (1980). *Applied Models in Urban and Regional Analysis*. Prentice-Hall, Inc. Englewood Cliffs: New Jersey.

Prewitt, K. (2000). The US decennial census: Political questions, scientific answers. *Population and Development Review, 26*(1), 1–16.

Rayer, S. (2008). Population forecast errors: A primer for planners. *Journal of Planning Education and Research, 27*(4), 417–430.

Shryock, H. S., Siegel, J. S., & Larmon, E. A. (1973). *The methods and materials of demography* (Vol. 2). US Bureau of the Census.

The National Research Council. (2007). *Using the American community survey: Benefits and challenges*. National Academies Press.

The U.S. Census Bureau. (2020). Please check https://www.census.gov/programs-surveys/surveys-programs.html, visited on August, 2020.

Wang, X., & Hofe, R. (2008). *Research methods in urban and regional planning*. Springer Science & Business Media.

Chapter 10
Analyzing the Local Economy

Understanding and promoting economic growth has been a consistent core concern of cities (Leitner, 1990; Stansel, 2005; Blakely & Leigh, 2013; Malizia et al., 2020). Local and Federal economic policies and programs have been initiated to reinvest in urban cores, to improve local competitiveness, and to market places. Local governments are especially motivated to promoting economic growth. For instance, in 2017, when the Amazon announced its search for a second headquarter, which would create 50,000 jobs, almost all U.S. cities with population close or beyond one million extended their invitations. Why are cities eager for such investment? What are the consequences of such investments on the local economy? How to analyze a local economy? This chapter explores theories and methods for analyzing local economic composition, exploring local economic strengths and weaknesses, and understanding current and future economic activities.

The Concept of Economy

Classical economic theories and discussions center around supply and demand and the relations between them (Case & Fair, 2007). An ***economy*** is a complex system with supply-demand interactions. Figure 10.1 provides a simplified illustration. The consumption activities of consumers, who can be local, regional, or global, generate demand for the supply side of an economy. Actors on the supply side include farms and factories that produce goods, and businesses that provide services. Theoretically, supply and demand are related by price through supply-demand equilibrium. Practically, these two sides of an economy are linked by distribution and trade, i.e., retail and wholesale industries.

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_10].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_10

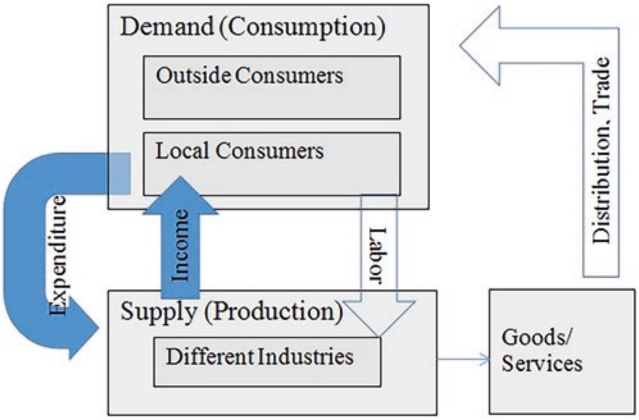


Fig. 10.1 The concept of economy

There are two circular flows in an economy, with one centering around the exchange of products/services and the other around the exchange of labor. Industries on the supply side provide goods and services to consumers through distribution and trade. In return, consumers make spending, which becomes revenue for industries to sustain their productions. At the same time, consumers are workers. They supply labors to the production section, and in return, they obtain income in the form of wage to support their consumptions. Any disruption to these two economic circular flows can cause an economy to decline.

Many government economic policies solely focus on job creation. For instance, state Enterprise Zone programs attempt to use financial incentives to encourage job creations to revitalize local economic conditions in declining neighborhoods (Zhang, 2019). Such a program ignores other elements of an economy, such as labor quality and local consumption abilities. Is there a group of qualified local labor to support an industry expansion in these neighborhoods? Will there be enough expenditure to sustain the production? A proper understanding of the structural integrity within an economy helps uncovering a set of economic programs with desired impacts.

Economic Activity Classification

It is important to group economic activities into categories. For different purposes, there are different classification methods. Two popular methods to group economic activities are by industry and by occupation.

An ***industry*** refers to a group of activities with similar operation processes or production cycles. For instance, factories could produce different products. But their production cycles are similar. They organize inputs for the production, produce products in factories, and ship products to retailers or wholesales for distributing. There are different types of retail stores, but they have similar business operations of selling goods in relatively small quantities to consumers.

Occupations are defined based on the nature of individual workers' activities. The differences among occupations are related to their knowledge bases, skill sets, and consequently salary differences. Planning is an occupation. All planners with the AICP (American Institute of Certified Planners) certification may be equipped with similar knowledge and with similar analytical and mapping skills. Workers within the same occupation can work with different industries. For example, lawyers can work with a law firm, a university, a manufacturing firm, or a government.

The ***North American Industry Classification System*** (NAICS) is the standard that is currently used by U.S. Federal agencies when collecting economic data such as the County Business Patterns and the Economic Census. The NAICS classifies economic activities according to production differences. It is an industry-specific classification system. Within the system, every industry is represented as a code within a hierarchy. As shown in Fig. 10.2, the total employment is divided into twenty 2-digit sectors, each of which consists of a varying number of 3-digit sectors. This nested disaggregation continues until the 6-digit level is reached. As an illustration, Fig. 10.2 only presents all the 2-digit sectors and the complete hierarchy attached to utilities (NAICS code 22). There is only one 3-digit sector (sector 221), which consists of three 4-digit sectors (sectors 2211, 2212, and 2213). Each of these 4-digit sectors contain 5-digit and 6-digit sectors.

The ***U.S. Standard Occupational Classification*** (SOC) system is published by the Federal Office of Management and Budget to classify economic activities based on occupations. This system has a hierarchical structure of jobs differentiated based on skills and qualifications. A six-digit code indicate an occupation. There are 23 major occupations, with their codes ending with 0000, as shown in Fig. 10.3. For instance, 23-0000 refers to legal occupations. There are further hierarchies nested within this 23-0000 branch.

U.S. Economic Data and Limitations

Economic data describes the past and current conditions of a community's economy, such as the employment level, the average household income, labor force attributes, production activities and so on. The Decennial Census and the American Community Survey contain resident-based data for exploring an economy, as introduced in Chap. 9. There are several employment-based economic databases that exclusively focus on employment, unemployment, production, workers' payroll, and other business attributes. The Census Bureau is the primary provider of such economic data and regularly releases the Economic Census and the County Business Patterns.

The Economic Census is available every 5 years. It is constructed based on surveys of all multi-establishment firms, all large single-establishment firms, and a

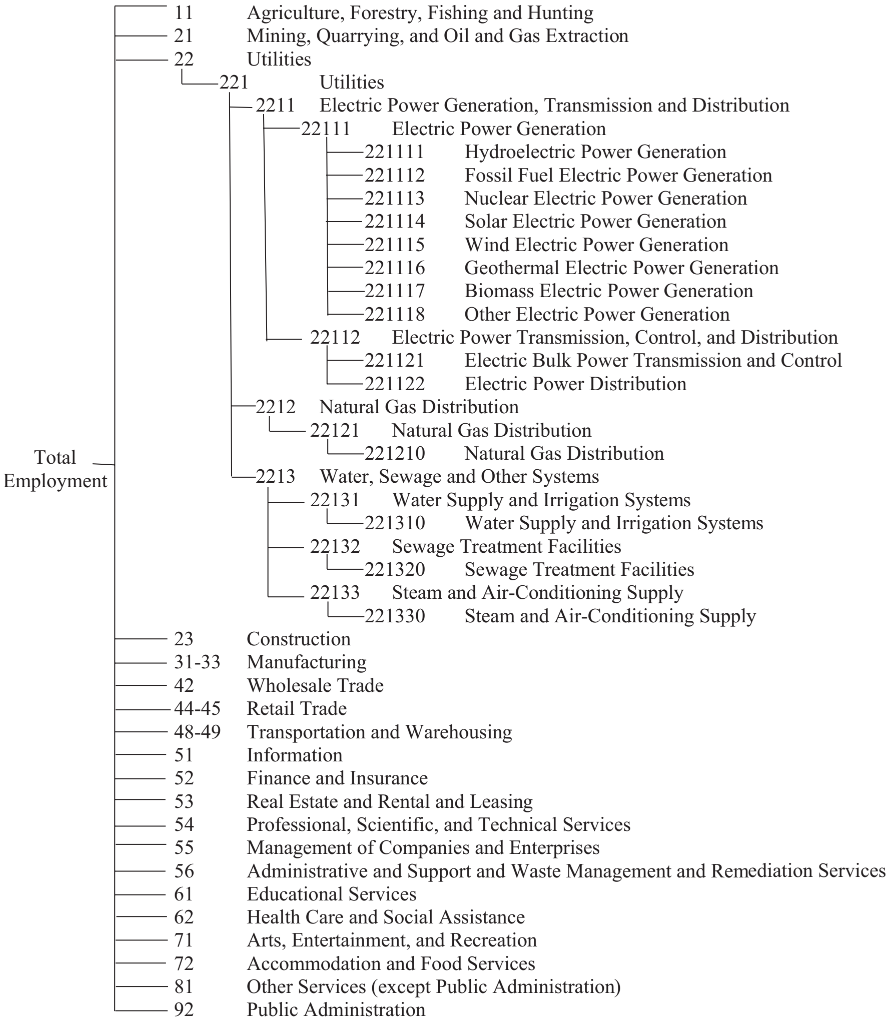


Fig. 10.2 The NAICS hierarchy (the complete branch of sectors related to utilities)

sample of small single-establishment firms. Available information from this database includes number of establishments, value of sales, shipments, receipts, and revenue, number of employment, annual payroll and first quarter payroll. Data is available for the U.S. states, commonwealths, territories, counties, and municipalities.

The County Business Patterns database contains annual economic information for U.S. counties and zip codes. It is extracted from the Business Register, which is a database containing individual administrative records of establishments and with restricted access. The main data source of the Business Register is the business master file from the Internal Revenue Service and the Census Bureau consolidates additional economic information from the Bureau of Labor Statistics, the Social

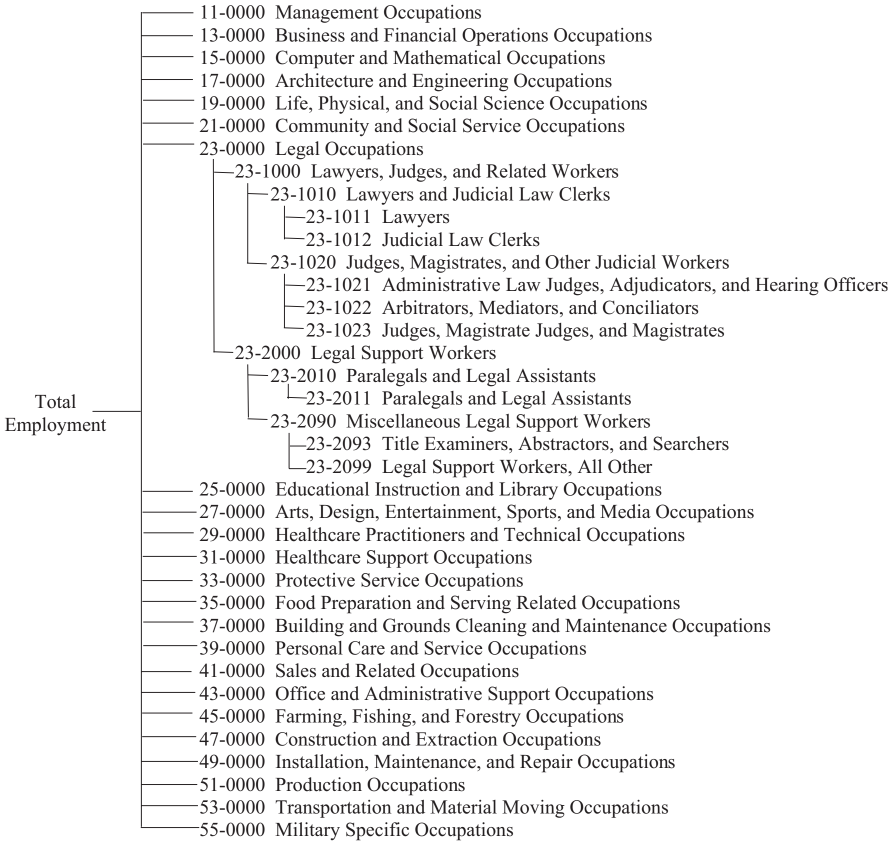


Fig. 10.3 The SOC hierarchy (major occupations and the complete branch of legal occupations)

Security Administration, and the Census Bureau itself (DeSalvo et al., 2016). The County Business Patterns provide industry specific information about the number of establishments, the number of employments, annual payroll, and first quarter payroll for U.S. counties. At the zip code level, the Census Bureau provides number of establishment information and total employment and payroll information.

When using the Economic Census or the County Business Patterns, one must deal with data suppression, which is a strategy widely used by governmental agencies to protect private information when releasing economic datasets. For instance, if a suburban county has only one manufacturing firm, statistical agencies cannot release the information of the manufacturing industry for this county because this will disclose the private information of this firm. In this situation, the economic data of this industry will be flagged. A common practice is to provide the range information instead of the actual employment. In the County Business Patterns databases, there are 12 employment ranges or flags: A 0–19, B 20–99, C 100–249, E 250–499, F 500–999, G 1000–2499, H 2500–4999, I 5000–9999, J 10,000–24,999, K

25,000–49,999, L 50,000–99,999, and M 100,000 or more. If the manufacturing employment of this suburban county is 750, it will be suppressed as Zhang and Guldmann (2010) explored 2000 County Business Patterns and found that 34.63% of U.S. state data and 66.61% of county data are suppressed in the database.

There are several methods to estimate these suppressed data. The Census Bureau recommends using midpoint of the lowest and the highest values of an employment range. For instance, if a record is flagged as B, the estimate of this record will be 60, the midpoint of range 20–99.

Gardocki and Baj (1985) and Kreahling et al. (1996) introduce establishment size information to refine employment estimates. Sechrist (1986) proposes to use the economic information at a higher geographic or economic hierarchy to adjust employment estimates by using the midpoint approach or by using establishment size information. Zhang and Guldmann (Zhang & Guldmann, 2009; Zhang & Guldmann, 2015) propose a programming approach to produce employment estimates for suppressed records.

Assessing an Economy

An assessment of the economic landscape of a city or a region provides essential inputs for all types of planning activities. Such an assessment often consists of a set of demographic and economic indicators to describe different aspects of an economy, as illustrated in Fig. 10.1.

Demographic and housing attributes provide information to assess local wealth, consumer preferences, and labor quality. Common indicators for local wealth include average/median household income, wage, average/median per capita income, average/median housing price, poverty rate and so on. With these indicators, a planner can assess the spending power of an economy. Examples of consumer preference indicators are average family size, median age, average number of children in a family, percent of females, and so on. With such information, a planner can make of a preliminary assessment of consumer behavior. As for labor quality, educational attainment is one of the common indicators.

Economic data, such as the Economic Census, the County Business Patterns, and employment data collected by the Bureau of Labor Statistics (BLS), provide information to construct indictors that directly measure the production side activities and the overall economy. The BLS regularly monitors unemployment rate for U.S. counties, which is calculated as the number of unemployed people divided by the size of the labor force. Unemployment rate serves as an important economic health measure. When the unemployment rate is increasing for a city, it means more people are losing jobs and losing their income sources. Consequently, the declining spending power of the consumption side will lead to a diminishing level of economic production.

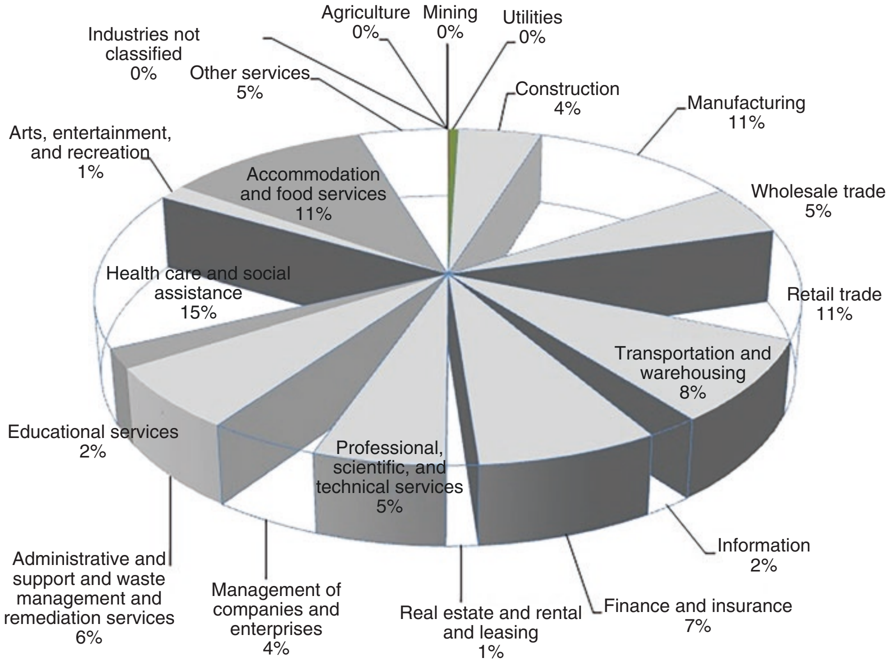


Fig. 10.4 Jefferson county 2016 employment by industry

Industry-specific economic data enables a study of a city's economic structure, which presents the composition of various components of an economy. Pie graph is a good choice to present an economic structure. With that planners can address such questions as what types of activities exist in the economy, what is the leading industry, and what are the weaknesses and strengths of a local economy? Figure 10.4 presents 2016 Jefferson County (KY) employment by industry. It provides a lot of information. First, with the presence of 20 industries, we can conclude that Jefferson County has great diversity in its economy. Second, three industries (health care, manufacturing, and retail) capture 37% of the total employment, and these three industries are the leading industries in the local economy because of employment sizes.

The Economic Base Theory

The ***Economic Base Theory*** explains why an economy grows or re-grows. It argues that investments that target at serving external demand stimulate a local economy to grow. A city's economic prosperity is built upon such basic economic activities. To understand the theory, one needs to understand the difference between basic and non-basic employment.

Basic vs. Non-basic Employment

Let's look at two hypothetical scenarios. In scenario 1, a local restaurant is operated in a city with a population of 100,000. The restaurant hires 20 workers with annual salary $40,000. If everyone in the city buys a one-dollar hamburger from the restaurant once a week, the annual income of the restaurant is $5,200,000 (assuming every year has 52 weeks). To operate the restaurant, the owner must pay the workers' wages, local taxes, utility costs, and rent, as shown in Fig. 10.5. What is left ($2,800,000) is the pure profit for the owner.

In scenario 2, Honda opens a factory in the same city, and produces 2000 cars every year. All those cars will be shipped out and sold to people outside at the price of $18,000. The factory hires 150 workers with an average annual salary $80,000. This production activity generates $36,000,000 income every year. At the same time, there are costs for hiring local workers ($12,000,000), paying taxes and utilities, and rent its factory, as show in Fig. 10.6. What is left ($5000,000) is the profit that will go to the headquarter in Japan.

How do these two different economic activities in the two scenarios impact the local economy? In scenario 1, local workers and the restaurant owner serve local demand from local people. This economic operation causes a redistribution of wealth of one dollar per week from everyone to the restaurant, and then, to the 20 workers, the government, the utility companies, and the restaurant owner. The total local wealth does not change. As a comparison, in scenario 2, the demand comes from the outside. People outside of the city pay $36 million for the production. Five million of this income goes back to Japan, the rest stays in the city as workers' wage, local governments' tax revenues, and local utility companies' and local property owners' incomes. The total local wealth will increase.

Employments created in scenario 1 are ***non-basic employments***. They serve local demand. Transactions associated with this type of business practice are monetary flows within the city, i.e., wealth flowing from one local individual's pocket to another's pocket. This type of activities certainly improves quality of life for residents but does not increase the total local wealth.

Employments created in scenario 2 are ***basic employments*** that serve external demand. They involve activities that produce goods and services sold to people and

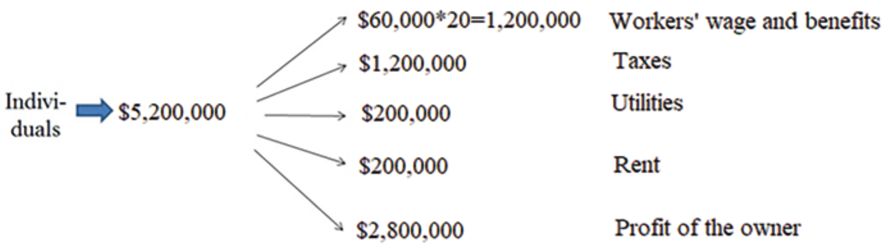


Fig. 10.5 Scenario 1

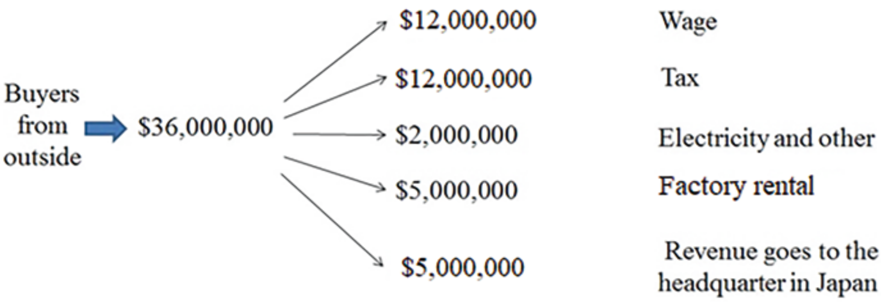
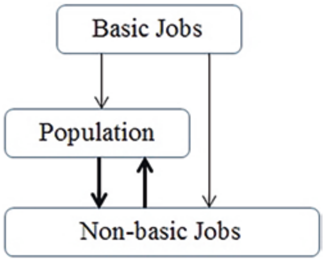


Fig. 10.6 Scenario 2

Fig. 10.7 The economic base model



businesses outside of the city. Transactions associated with the business practice are associated with monetary flows from outside into the city, i.e., wealth flowing into the city from outside. This type of activities bring income into a city and increase the total local wealth.

Base Multiplier

The economic base theory presents the role of export-oriented industries in stimulating the growth of a local economy. As shown in Fig. 10.7, a region's growth starts with a large size investment of basic jobs that serve external demand. This investment brings workers into the city, and they have a demand for service activities, such as retailing and personal service activities. This generates non-basic jobs. Workers also bring their families in. The local population grows. This growing population generate more demand for non-basic jobs. The increasing non-basic job opportunities attract more people to move into the region, which create additional demand for non-basic jobs. Non-basic jobs attract people and people have demand for more non-basic jobs in the service industry. This cycle continues until it gradually slows down and reaches its equilibrium.

Let us use the Amazon's second headquarter investment as an example. The Amazon announced it in 2017 and stated that this investment would bring 50,000 jobs. These 50,000 jobs serve amazon customers all over the country and are basic jobs. These job opportunities would attract people to move into the city. The growing population generate demand for different types of services. As a result, there would be consistent growth in retail, real-estate, finance, and other service industries. The total job growth will be much more than 50,000 jobs. There would be a multiplier effect in job creation from the Amazon investment. This multiplier effect is measured as the Base Multiplier.

The ***base multiplier*** (BM) of a city is calculated as the ratio between the total local employment (T) and total local basic employment (B), as

$$BM = T / B$$

Let us look at an example before we further interpret the meaning of a base multiplier. In a rural county, GE (General Electric) made an initial investment that created 7000 job opportunities. As a result, 5000 other jobs (retail, personal service, K-12 education, transportation, entertainment etc) were stimulated, and the total employment of the county stabilized around 12,000. What is the Base Multiplier for this county? In this county, the total basic job is the 7000 GE jobs, with B = 7000. The total job is 12,000, with T = 12,000. The base multiplier can be calculated as BM = T/B = 12,000/7000 = 1.71.

What does a base multiplier mean? In the above example, a total of 12,000 jobs, which include 7000 basic jobs and 5000 non-basic jobs, have been created because of the 7000 basic jobs investment. The base multiplier takes the total employment normalized by total basic employment. It informs that 1.71 unit of total jobs is related to one unit of basic job, or 0.71 unit of basic job has been created because of one unit of basic job.

Base multiplier provides useful information for analyzing the potential impacts of an external investment and for projecting the overall employment growth related to this investment. In the case of the GE example, based on what has happened in the city, local administrators know that one unit of basic job led to 1.71 unit of growth in the total employment. If the government know that GE is going to expand the factory and add 2000 additional jobs, it will be able to project that the total employment will grow by 3420 (2000*1.71).

Calculating Basic and Non-basic Employment – Quick Approaches

Government agencies regularly publish employment data by industry, by occupation, by race, by gender, but never by basic and non-basic employment. To find out the base multiplier for a city, local administrators or planners need to figure out the size of basic employment in their city. There are several quick and informal approaches.

The first approach is to allocate all the employment of an industry into the basic or non-basic sector based on the nature of the industry in a city. Table 10.1 provides

Table 10.1 2016 employment by industry

	Florida	Kentucky	Ohio
Total for all sectors	8,169,642	1,603,173	4,790,178
Agriculture, forestry, fishing and hunting	9051	1489	1207
Mining, quarrying, and oil and gas extraction	3747	10,134	10,575
Utilities	26,487	8530	23,488
Construction	397,243	69,808	188,864
Manufacturing	306,700	235,487	662,428
Wholesale trade	318,689	71,545	232,886
Retail trade	1,088,564	214,356	573,837
Transportation and warehousing	248,639	96,476	176,312
Information	167,490	32,140	84,760
Finance and insurance	355,395	74,742	255,131
Real estate and rental and leasing	164,836	18,186	66,391
Professional, scientific, and technical services	499,518	74,749	250,766
Management of companies and enterprises	153,650	31,161	150,837
Administrative and support and waste management and remediation services	1,707,780	121,236	397,162
Educational services	174,121	28,807	121,226
Health care and social assistance	1,086,559	255,345	840,716
Arts, entertainment, and recreation	198,688	17,667	68,762
Accommodation and food services	935,554	173,991	481,956
Other services (except public administration)	326,030	66,783	202,431
Industries not classified	901	541	443
Total basic jobs (in grey cells)	518,186	247,110	674,210
Base multiplier	15.77	6.49	7.10

an illustration for analyzing the economies of Florida, Kentucky, and Ohio. The three-sector model in economics divides economic activities into primary, secondary, and tertiary sectors. Primary economic activities extract raw materials from the environment and include agriculture and mining. These activities support the other two types of economic activities. Secondary economic activities focus on production, such as manufacturing. Tertiary activities are service oriented, directly serving people. Primary and secondary economic activities are assumed to meet external demand, while tertiary activities focus on local demand. Therefore, we can argue that agriculture, mining, and manufacturing employments are basic. The argument should also consider local conditions. For instance, the recreational industry in Florida is different from the recreational industry from the other two states. Florida is an important recreational destination in the country. Its recreational industry is serving customers from other states, while in the other two state, this industry is serving local demand. Therefore, we can argue that the employment in Arts, Entertainment, and Recreation is basic in Florida. Total basic jobs can be calculated, and base multipliers can be derived as the ratio of total jobs and total basic jobs for the three states.

The second approach is to calculate total basic employment based on top employers in the community. Top employers hire a large number of workers. It is unlikely that all these workers serve local demand. For instance, Jefferson County, KY, is the home of UPS' Worldport air hub, which handles packages from more than 220 countries and territories around the world. In 2019, the UPS hired 23,533 workers in Jefferson County. Majority of these workers are serving parcel mailing demand from outside of the region. Therefore, it is safe to argue that all the UPS employments are basic. We can extend the same arguments to other top employers, but not for public schools and the local government, which are larger employers but provide service to local population. Total basic employment and base multiplier can be calculated accordingly.

The above two approaches ignore the fact that every local industry is serving both local and external demand simultaneously. For instance, for the UPS in Jefferson County, a small portion of the UPS workers are delivering parcels within Jefferson County and another big group is working to serve the external demand. The third approach makes up for that. Top employers will be contacted to obtain information about what portion of their workers is serving local people. The basic employment calculation is based on a refined list of top employers. For instance, a local administrator can make a quick phone call to the manager of the UPS and ask for his/her opinion about what proportion of their workers focus on local delivery. Based on the information, the local administrator can obtain more accurate estimates about basic and non-basic employment within the UPS.

Location Quotients

Location quotients (LOQs) provide information for industry-specific assessments and enable local officials and planners to assess the strength and weakness of a local economy. To use the approach, one must (1) collect employment by industry data for the study region and a larger region inclusive of the study area (in this example, the nation). As shown in Table 10.2, the study region is the state of Kentucky; (2) calculate the employment percentage distribution for both the study region and the larger region. For instance, in Kentucky and in 2016, manufacturing employment were 14.69% of the total employment, and in the nation, 9.14% of the total employment was manufacturing; and (3) for every industry, calculate its location quotient (LOQ) as the study region's percentage divided by the corresponding nation's percentage.

Location quotients compare industries in a local economy to their corresponding ones in a larger region, usually the national economy. What do these values mean? Take a further look at the manufacturing industry in Table 10.2. In both the Kentucky and the national economy, the proportion of an industry's employment represents its importance in the economy. In Kentucky, the manufacturing proportion is 14.69%, which is higher than the manufacturing proportion in the nation (9.14%). As a result, the location quotient for Kentucky's manufacturing industry is 1.61 (14.69% divided by 9.14%). This higher-than-one location quotient indicates a relative strength of

Table 10.2 2016 location quotient analyses for Kentucky

	Kentucky	U.S.	% of Total		Location Quotient
			Kentucky	U.S.	
Total for all sectors	1,603,173	126,752,238			
Agriculture, forestry, fishing and hunting	1489	160,411	0.09%	0.13%	0.73
Mining, quarrying, and oil and gas extraction	10,134	587,017	0.63%	0.46%	1.36
Utilities	8530	638,917	0.53%	0.50%	1.06
Construction	69,808	6,311,264	4.35%	4.98%	0.87
Manufacturing	235,487	11,590,420	14.69%	9.14%	1.61
Wholesale trade	71,545	6,110,748	4.46%	4.82%	0.93
Retail trade	214,356	15,967,893	13.37%	12.60%	1.06
Transportation and warehousing	96,476	4,729,709	6.02%	3.73%	1.61
Information	32,140	3,447,950	2.00%	2.72%	0.74
Finance and insurance	74,742	6,336,795	4.66%	5.00%	0.93
Real estate and rental and leasing	18,186	2,111,418	1.13%	1.67%	0.68
Professional, scientific, and technical services	74,749	8,799,893	4.66%	6.94%	0.67
Management of companies and enterprises	31,161	3,380,437	1.94%	2.67%	0.73
Administrative and support and waste management and remediation services	121,236	11,628,509	7.56%	9.17%	0.82
Educational services	28,807	3,677,275	1.80%	2.90%	0.62
Health care and social assistance	255,345	19,735,708	15.93%	15.57%	1.02
Arts, entertainment, and recreation	17,667	2,311,437	1.10%	1.82%	0.60
Accommodation and food services	173,991	13,704,017	10.85%	10.81%	1.00
Other services (except public administration)	66,783	5,499,244	4.17%	4.34%	0.96
Industries not classified	541	23,176	0.03%	0.02%	1.85

the corresponding industry in the local economy. Kentucky has strengths, or is more competitive, in mining, manufacturing, and transportation and warehousing industries.

Pay attention to the Educational Services industry, 1.8% of Kentucky workers work in this industry, while 2.9% of workers belong to this industry in the nation. As a result, the location quotient of this industry is 0.62 (1.8% divided by 2.9%), which is lower than 1, indicating a weakness of the Kentucky economy. Look across industries, Kentucky economy's weaknesses are in Real Estate and Rental and Leasing, Professional, Scientific, and Technical Services, Educational Services, and Arts, Entertainment, and Recreation industries.

In summary, a larger-than-one location quotient indicates a strength in this specific industry of the local economy. A location quotient smaller than one indicates a weakness. A location quotient equaling one indicates that the performance of this local industry meets the national average.

Calculating Base Multiplier: The Location Quotient Approach

One can reasonably assume that the national economy is self-sufficient and is producing to meet the demand of people within the country, i.e., local demand. Economic theories suggest that producers will not produce more than the demand. As a result, the distribution of employment in the nation represents a set of benchmarks to meet local demand. For instance, in Table 10.2, with 9.14% of workers working in the manufacturing industry in the U.S., this industry can produce the right amount of goods to fully meet the manufacturing demand in the country. With 12.6% of workers working in the retail industry, this industry can provide the right amount of service to fully meet the demand for retail service in the country.

A less-than-one location quotient indicates that the proportion of this industry in the local employment is smaller than the corresponding proportion in the national economy. The national proportion represents the benchmark percentage of employment required to fully satisfy local demand. Therefore, this local industry does not have enough capacity to meet local demand. All the existing employment is serving local demand, and therefore, is non-basic employment. At the same time, the local economy may import this type of goods from outside of the region. Using the Agriculture, Forestry, Fishing, and Hunting industry as an example, the location quotient is 0.73, which is less than one. It suggests that this industry in Kentucky cannot fully meet the local demand. Therefore, all the existing 1489 jobs are non-basic employment and there is 0 basic employment in this industry in Kentucky.

If a location quotient equals one, it indicates that the proportion of this industry in the local employment is the same as the corresponding industry in the national economy. This local industry has the right capacity to meet local demand. All the existing employment is serving local demand, and therefore, is non-basic employment. The local demand is fully served. There is no need to import, and there is no capacity to export as well. The basic employment is 0.

If a location quotient is greater than one, it indicates that the proportion of this industry in the local employment is higher than the corresponding industry in the national economy. This local industry has more than enough capacity to meet local demand. The industry first meets local demand, and then uses the extra capacity for exporting activities. There are both basic and non-basic employments in this industry.

The following are the steps for calculating the base multiplier of a region based on location quotients information, using Kentucky as a case study, as shown in Table 10.3. First, for the industries with location quotients less than or equal to 1 (greyed cells in Table 10.3), these industries do not have any capacity to serve external demand. All the existing jobs are non-basic. There is 0 basic job. Second, for an industry with location quotient more than 1, look at the national percentage benchmarks, and calculate the amount of employment needed to serve local demand. In the case of the mining industry, to fully meet local demand, Kentucky needs to have 0.46% of its total employment, i.e., 7425 workers, work in the mining industry for serving local demand. Therefore, these 7425 jobs are non-basic employment. Third, for an industry with location quotient more than 1, calculate the basic employment as the difference between the industry's employment and the non-basic employment as calculated in the previous step. In the case of mining, it has 10,134 employment, 7425 of which are non-basic. Therefore, 2709 (10,134-2709) are basic. Fourth, add up the basic employment of every industry to get the total basic employment. At last, calculate the base multiplier as the total employment divided by the total basic employment.

Shift-Share Analysis

Many factors affect a region's economic growth. First, national programs, policies, and infrastructure investments can affect local growth. A good example is how the national highway investment has influenced growth in many U.S. cities (Li & Whitaker, 2018). Second, different industries have different growth and decline patterns, for instance, a quick expansion of the information industry in the 1990s (Jorgenson, 2001) and a sharp manufacturing decline in the 2000s (Atkinson et al., 2012; Pierce & Schott, 2016). Such industry specific growth or decline influences the growth of all businesses within the industry, regardless of their locations. At last, a city itself may have its own influences an industry's growth due to its advantages in location, natural resources, human capital, local policy, and existing economic strength (Mathur, 1999; McCann & Van Oort, 2019; Stijns, 2005). Shift-Share analysis is an approach for analyzing a city's economic growth in relations to these three different factors, national, industrial, or local.

Table 10.3 Calculating base multiples using location quotients

	Kentucky	U.S.	% of total		Location quotient	Basic employment	Non-basic employment
			Kentucky	U.S.			
Total for all sectors	1,603,173	126,752,238					
Agriculture, forestry, fishing and hunting	1489	160,411	0.09%	0.13%	0.73	0	1489
Mining, quarrying, and oil and gas extraction	10,134	587,017	0.63%	0.46%	1.36	2709	7425
Utilities	8530	638,917	0.53%	0.50%	1.06	449	8081
Construction	69,808	6,311,264	4.35%	4.98%	0.87	0	69,808
Manufacturing	235,487	11,590,420	14.69%	9.14%	1.61	88,890	146,597
Wholesale trade	71,545	6,110,748	4.46%	4.82%	0.93	0	71,545
Retail trade	214,356	15,967,893	13.37%	12.60%	1.06	12,393	201,963
Transportation and warehousing	96,476	4,729,709	6.02%	3.73%	1.61	36,654	59,822
Information	32,140	3,447,950	2.00%	2.72%	0.74	0	32,140
Finance and insurance	74,742	6,336,795	4.66%	5.00%	0.93	0	74,742
Real estate and rental and leasing	18,186	2,111,418	1.13%	1.67%	0.68	0	18,186
Professional, scientific, and technical services	74,749	8,799,893	4.66%	6.94%	0.67	0	74,749
Management of companies and enterprises	31,161	3,380,437	1.94%	2.67%	0.73	0	31,161
Administrative and support and waste management and remediation services	121,236	11,628,509	7.56%	9.17%	0.82	0	121,236
Educational services	28,807	3,677,275	1.80%	2.90%	0.62	0	28,807
Health care and social assistance	255,345	19,735,708	15.93%	15.57%	1.02	5726	249,619
Arts, entertainment, and recreation	17,667	2,311,437	1.10%	1.82%	0.60	0	17,667
Accommodation and food services	173,991	13,704,017	10.85%	10.81%	1.00	661	173,330
Other services (except public administration)	66,783	5,499,244	4.17%	4.34%	0.96	0	66,783
Industries not classified	541	23,176	0.03%	0.02%	1.85	248	293
Total basic employment in Kentucky						147,731	
Total employment in Kentucky						1,603,173	
2016 base multiplier						10.85	

Set Up a Shift-Share Analysis

A ***shift-share analysis*** compares a region's industry-specific employment growth in relation to the nation across time. One must collect past and current employment by industry data for both the nation and the study area, which can be a city, a county, or a state. Figure 10.8 illustrates the basic idea, using manufacturing industry as an example. With data collected, the first thing is to identify the past and current manufacturing employment in the nation and in the study area. Then the analysis proceeds with uncovering relations that can be used to explains the current local manufacturing employment based on past and current national employment and past local manufacturing employment. The uncovered relations can be used to project the local manufacturing employment in the future. When an area is bigger, it is easier to detect stable growth trend. Therefore, future national manufacturing employment is comparatively easier to project. With this information available, together with current local manufacturing employment, one can project future manufacturing employment.

There are several things to clarify for using the shift-share analysis method. First, study period is an important concept. Industry-specific employment data are available annually in many databases. One needs to define a study period considering trends within the data and the target year of the projection. For instance, if one use 10-year as one period in exploring job growth, the person must project employment for every 10 years. If one has analyzed how industries in a local economy has grown from 2005 to 2015, the person will continue to study employment by every 10 years, i.e., 2025, 2035, and so on. If one uses 5-year as one period and analyzes historical growth over 2010–2015, she will project future employment by 5 years, i.e., 2020, 2025, 2030, etc.

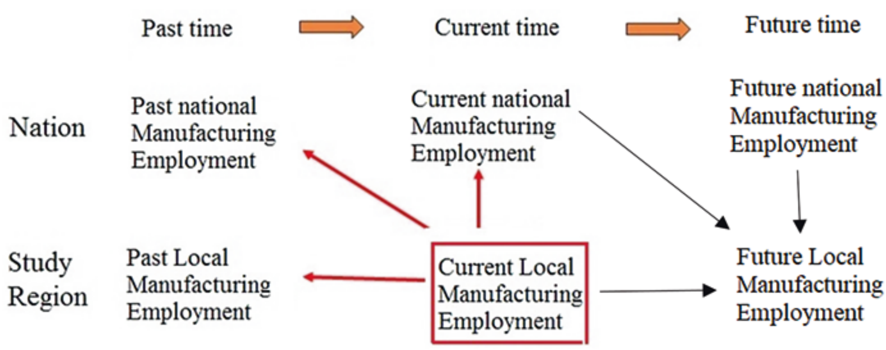


Fig. 10.8 Basic ideas of shift-share analysis

National Share

The national economic environment affects all local economies within it. When a national economy is growing, all local economies are expected to grow. Figure 10.9 presents manufacturing employment growth trends in 3 U.S. counties, as well as the overall U.S. employment growth. The manufacturing growth of all the three counties accelerated their declining trends over 2008–2010, when the national economy recessed, and began to increase in Jefferson County, KY, and Miami-Dade County, FL, over 2010–2015, when the national economy regrew. ***National share*** captures the impact from the national economic environment and represents the expected amount of employment of a local industry if it follows the national economic growth. The national economic growth is measured by the total employment growth rate for the country.

To calculate a national share for a local industry, first calculate the total employment growth rate for the nation, and then calculate the expected amount of employment if this industry follows this overall national growth rate. Table 10.4 provides an example of calculating 2016 national shares for all industries in Miami-Dade County, FL. Over 2012–2016, the total employment growth rate of the nation is 9.33%. The 2016 national share of an industry is the expected amount of employment if this industry follows the overall national growth rate of 9.33%. In the case of the manufacturing industry, its 2012 employment is 32,528. If it followed the overall national trend and increased by 9.33% over 2012–2016, its expected amount of employment in 2016 would be 32,528*(1 + 9.33%) = 35,562. The 2016 National Share column presents the amount of employment that every industry would have had in 2016 if this industry has had followed the overall national growth rate of 9.33%.

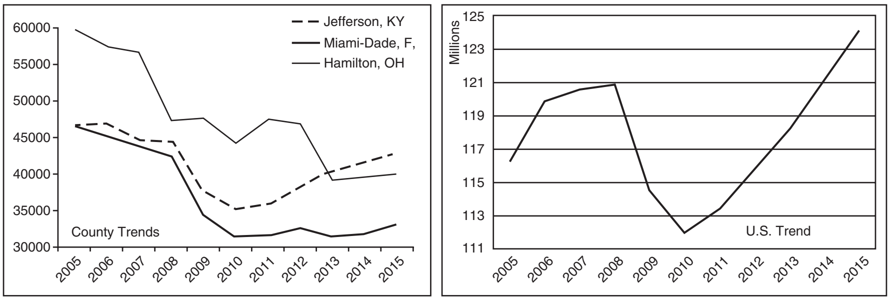


Fig. 10.9 Manufacturing employment trends, 2005–2015

Table 10.4 Calculation of national share, industry mix, and regional shift

	2012		2016								
	U.S.	Employment Miami-Dade, FL	U.S.	Employment Miami-Dade, FL	2012–16 U.S. growth rate (B)	2016 national share	B – A	2016 industry mix	2012–16 local growth rate (C)	C – B	2016 regional shift
Total for all sectors	115,938,468	861,535	126,752,238	972,716	9.33% (A)						
Agriculture, forestry, fishing and hunting	161,077	296	160,411	227	−0.41%	324	−9.74%	−29	−23.31%	−22.90%	−68
Mining, quarrying, and oil and gas extraction	727,626	404	587,017	594	−19.32%	442	−28.65%	−116	47.03%	66.35%	268
Utilities	641,063	4122	638,917	3608	−0.33%	4506	−9.66%	−398	−12.47%	−12.13%	−500
Construction	5,260,942	26,201	6,311,264	37,494	19.96%	28,645	10.64%	2787	43.10%	23.14%	6062
Manufacturing	11,192,043	32,528	11,590,420	35,960	3.56%	35,562	−5.77%	−1876	10.55%	6.99%	2274
Wholesale trade	5,776,243	66,405	6,110,748	72,906	5.79%	72,599	−3.54%	−2348	9.79%	4.00%	2655
Retail trade	14,807,958	125,248	15,967,893	142,824	7.83%	136,930	−1.49%	−1871	14.03%	6.20%	7765
Transportation and warehousing	4,233,381	51,595	4,729,709	62,766	11.72%	56,407	2.40%	1237	21.65%	9.93%	5122
Information	3,136,025	18,813	3,447,950	22,175	9.95%	20,568	0.62%	117	17.87%	7.92%	1491
Finance and insurance	5,979,661	43,975	6,336,795	49,973	5.97%	48,077	−3.35%	−1475	13.64%	7.67%	3372
Real estate and rental and leasing	1,940,681	21,205	2,111,418	25,048	8.80%	23,183	−0.53%	−112	18.12%	9.33%	1977
Professional, scientific, and technical services	8,016,181	60,293	8,799,893	69,970	9.78%	65,917	0.45%	271	16.05%	6.27%	3782

(continued)

Table 10.4 (continued)

	2012		2016		2012–16 U.S. growth rate (B)	2016 national share	B – A	2016 industry mix	2012–16 local growth rate (C)	C – B	2016 regional shift
	U.S.	Employment Miami-Dade, FL	U.S.	Employment Miami-Dade, FL							
Management of companies and enterprises	3,037,299	16,293	3,380,437	23,134	11.30%	17,813	1.97%	321	41.99%	30.69%	5000
Administrative and support and waste management and remediation services	9,866,296	72,969	11,628,509	65,369	17.86%	79,775	8.53%	6227	−10.42%	−28.28%	−20,633
Educational services	3,477,047	33,067	3,677,275	36,382	5.76%	36,151	−3.57%	−1180	10.03%	4.27%	1411
Health care and social assistance	18,378,342	132,670	19,735,708	138,060	7.39%	145,044	−1.94%	−2576	4.06%	−3.32%	−4409
Arts, entertainment, and recreation	2,057,290	12,415	2,311,437	15,688	12.35%	13,573	3.03%	376	26.36%	14.01%	1739
Accommodation and food services	11,985,274	106,323	13,704,017	128,799	14.34%	116,240	5.01%	5330	21.14%	6.80%	7229
Other services (except public administration)	5,256,250	36,683	5,499,244	41,610	4.62%	40,104	−4.70%	−1726	13.43%	8.81%	3231
Industries not classified	7789	30	23,176	129	197.55%	33	188.22%	56	330.00%	132.45%	40

Industry Mix

Different industries have different histories and are in different growth stages. They have different growth trends. As shown in the 2012–2016 U.S. Growth Rate column in Table 10.4, no industries exactly followed the overall national trend of a 9.33% growth. Some industries grew faster. For instance, the Construction industry grew by 19.96% across the country, which is much higher than the overall growth rate, while the Manufacturing industry only grew by 3.56%, which is lower than 9.33%. ***Industry mix*** reflects the differences between an industry's growth and the overall total employment growth in the country. It exams how to modify the national share to reflect the influence from an industry on a local growth.

To calculate the industry mix for a local industry, one first needs to calculate growth rates for all industries in the nation, as shown in the 2012–16 U.S. Growth Rate column. Then calculate the difference between the industry-specific growth rate and the total employment growth rate, i.e., the (U.S. Industry Growth Rate) – (U.S. Total Growth Rate) column in Table 10.4. At last, calculate the industry mix of a local industry as the previous employment of this industry multiplied by the growth rate difference that has just been calculated.

The 2016 Industry Mix column in Table 10.4 presents industry mixes for all industries in Miami-Dade County, FL. A negative industry mix indicates a comparatively slow-growing or declining industry in the nation. This industry-specific feature has a negative impact on all the firms within this industry across the country, regardless of their locations. For instance, the industry mix is −1876 for Miami-Dade's manufactory industry. This suggests that because of the under-performance of the manufacturing industry across the country, Miami-Dade's manufacturing industry lost 1876 employments in 2016. A positive industry mix indicates a fast-growing industry in the national economic system. For instance, the industry mix of the accommodation and food service industry in Miami-Dade is 5330. This suggests that this industry is a fast-growing industry, and because of this, Miami-Dade gained 5333 employments in this industry in 2016.

Regional Shift

The introduction of industry mix allows different industries to have different growth rates. The same industry could have different growth trends in different counties or cities. As shown in Fig. 10.9, the manufacturing industries in Miami-Dade County, FL, Jefferson County, KY, and Hamilton, OH have different growth trends. They all declined over 2005–2010. However, over 2010–2015, the three counties had different manufacturing growths. This can only be explained by local factors, such as the size of existing manufacturing activities and infrastructure support, local political environment, and/or local land use regulatory system. ***Regional Shift*** is to reflect the influence of local factors on industry growth.

To find out the regional shift for a local industry, first calculate the growth rate of every local industry over the study period, as presented in the 2012–2016 Local Growth Rate column in Table 10.4. Then assess the difference between the local growth rate of an industry and the national growth rate of the same industry. For instance, for the manufacturing industry, the 2012–16 national growth rate is 3.56% and the local growth rate is 10.55%. The difference is 6.99%. Then, calculate the regional shift as the previous employment of this industry multiplied by this rate difference. In Miami-Dade County, the manufacturing industry grew 6.99% faster than the national manufacturing growth. A 6.99% growth based on 2012 manufacturing employment is 2274 (6.99% * 32,528) jobs. This is the 2016 regional shift of the manufacturing industry in Miami-Dade County.

Regional shift is especially useful for a local government to assess the weaknesses and strengths of its economy, with a positive regional shift indicating a strength and a negative one for a weakness. With the regional shift being 2274 for the manufacturing industry and -4409 for the health care and social assistance industry, these results suggest that the local environment in Miami-Dade County caused an increase of 2274 manufacturing jobs in Miami-Dade County, and a loss of 4409 health care and social assistance jobs in 2016.

Employment Decomposition

Check the 2016 National Share, 2016 Industry Mix, and 2016 Regional Shift columns in Table 10.4, for any industry, if one adds up the national share, the industry mix, and the regional shift of an industry, it is exactly the 2016 employment for this industry. For instance, for the manufacturing industry, if one adds up the national share (35,562), the industry mix (−1876), and the regional shift (2274), she gets 35,960, the 2016 manufacturing employment in Miami-Dade County. The Shift-Share analysis decomposes the local employment of an industry into three components, national share (NS), industry mix (IM), and regional shifts (RS). These three components are mathematically represented as:

$$NS_{i,t} = e_{i,t-1} \times \left(\frac{E_t}{E_{t-1}} \right)$$

$$IM_{i,t} = e_{i,t-1} \left(\frac{E_{i,t}}{E_{i,t-1}} - \frac{E_t}{E_{t-1}} \right)$$

$$RS_{i,t} = e_{i,t-1} \left(\frac{e_{i,t}}{e_{i,t-1}} - \frac{E_{i,t}}{E_{i,t-1}} \right)$$

where $NS_{i,\,t}$ is the national share at time t for industry i; $IM_{i,\,t}$ is the industry mix at time t for industry i; $RS_{i,\,t}$ is the regional shift at time t for industry i; E_{t-1} and E_t are total employment of the nation at time t-1 and t; $E_{i,\,t-1}$ and $E_{i,\,t}$ are the national employment for industry i at time t-1 and t; and $e_{i,\,t-1}$ and $e_{i,\,t}$ are the local employment for industry i at time t-1 and t. For any time t, $e_{i,\,t} = NS_{i,\,t} + IM_{i,\,t} + RS_{i,\,t}$.

The national share and the industry mix components both consider the influence of national level factors on a local economy. Let us define a ***regional proportion*** of industry i at time t ($RP_{i,\,t}$) by adding its corresponding national share and industry mix, as:

$$RP_{i,t} = NS_{i,t} + IM_{i,t} = e_{i,t-1} \cdot \frac{E_{i,t}}{E_{i,t-1}}$$

Thus, the local employment of industry i at time t consists with two components, the regional proportion and the regional shift, with $e_{i,\,t} = RP_{i,\,t} + RS_{i,\,t}$, where the regional proportion directly reflects this industry's growth at the national level and the regional shift accounts for locality. In Table 10.5, the 2016 Regional Proportion column is simply the addition of the 2016 national share and 2016 industry mix columns.

Employment Projection

For a future time (t + 1), if we know the future regional proportion for industry i ($RP_{i,\,t+1}$) and the future regional shift ($RS_{i,\,t+1}$), we will be able to project the future employment level of this industry, as $e_{i,\,t+1} = RP_{i,\,t+1} + RS_{i,\,t+1}$.

A regional proportion is affected by the growth of this industry in the nation. At the national level, if an industry has been growing in the past time, it is very likely that it will continue to grow at a similar speed in next period. For instance, the national manufacturing employment has grown 3.56% over 2012–2016. It is reasonable to assume that this industry will continue to grow 3.56% in the nation over 2016–2020. Therefore, 2020 regional proportion for manufacturing is 34,885, which is 3.56% more than 2016 manufacturing regional proportion.

A regional shift indicates a local strength or a weakness and is affected by local factors. If a set of industries are the strengths in the local economy, they may continue to be the strengths in the next period. Therefore, with no additional information, one can assume that future regional shifts are the same as the current ones, i.e., future local strengths/weaknesses are the same as current strengths/weaknesses. In Table 10.5, for every industry, 2020 regional shift equals 2016 regional shift for every industry. With 2020 regional proportion and regional shift ready, one can directly add them up to get 2020 employment for every industry.

In summary, the shift-share approach is indeed a ratio approach to compare local industry growth with national industry growth and decompose the employment of every local industry into three components: national share, industry mix, and regional shift. National share and industry mix can be grouped as regional

Table 10.5 Employment projection

	2012 employment Miami-Dade, FL	2016 employment Miami-Dade, FL	2012–16 U.S. growth rate	2016 national share	2016 industry mix	2016 regional shift	2016 regional proportion	2020 regional proportion	2020 regional shift	2020 employment, Miami-Dade, FL
Total for all sectors	861,535	972,716	9.33%							
Agriculture, forestry, fishing and hunting	296	227	−0.41%	324	−29	−68	295	294	−68	226
Mining, quarrying, and oil and gas extraction	404	594	−19.32%	442	−116	268	326	263	268	531
Utilities	4122	3608	−0.33%	4506	−398	−500	4108	4094	−500	3594
Construction	26,201	37,494	19.96%	28,645	2787	6062	31,432	37,707	6062	43,769
Manufacturing	32,528	35,960	3.56%	35,562	−1876	2274	33,686	34,885	2274	37,159
Wholesale trade	66,405	72,906	5.79%	72,599	−2348	2655	70,251	74,319	2655	76,974
Retail trade	125,248	142,824	7.83%	136,930	−1871	7765	135,059	145,638	7765	153,403
Transportation and warehousing	51,595	62,766	11.72%	56,407	1237	5122	57,644	64,402	5122	69,524
Information	18,813	22,175	9.95%	20,568	117	1491	20,684	22,742	1491	24,232
Finance and insurance	43,975	49,973	5.97%	48,077	−1475	3372	46,601	49,385	3372	52,756
Real estate and rental and leasing	21,205	25,048	8.80%	23,183	−112	1977	23,071	25,100	1977	27,078
Professional, scientific, and technical services	60,293	69,970	9.78%	65,917	271	3782	66,188	72,659	3782	76,441
Management of companies and enterprises	16,293	23,134	11.30%	17,813	321	5000	18,134	20,182	5000	25,183

Administrative and support and waste management and remediation services	72,969	65,369	17.86%	79,775	6227	−20,633	86,002	101,363	−20,633	80,730
Educational services	33,067	36,382	5.76%	36,151	−1180	1411	34,971	36,985	1411	38,396
Health care and social assistance	132,670	138,060	7.39%	145,044	−2576	−4409	142,469	152,991	−4409	148,582
Arts, entertainment, and recreation	12,415	15,688	12.35%	13,573	376	1739	13,949	15,672	1739	17,411
Accommodation and food services	106,323	128,799	14.34%	116,240	5330	7229	121,570	139,004	7229	146,233
Other services (except public administration)	36,683	41,610	4.62%	40,104	−1726	3231	38,379	40,153	3231	43,384
Industries not classified	30	129	197.55%	33	56	40	89	266	40	305

proportion. The projection of the future employment relies on the ability to identify future regional proportion and regional shift.

Input-Output Analysis

The supply-demand relation is one of the fundamental elements in economic theories. When the price of a service or a product is increasing, the supply will increase, but the demand will decrease, and vice versa. The market will stabilize at an equilibrium state when the supply meets the demand. This is theoretically sound, but practically difficult to use in daily decision-makings. For instance, a city projects that its population will grow and the demand for different services/products will grow accordingly. On the supply side, how many products will be produced to meet this increasing demand? How many jobs will be created? The input-output model addresses these practical questions. It was proposed by Wassily Leontief, and he won the Nobel prize in Economics because of this work. To introduce the method, let us first explore the concept of a transaction table.

The Transaction Table

The central element of an input-output analysis is the ***transaction table*** of a region's economy. Table 10.6 presents a simplified example. In this economy, there are only three industries: agriculture, mining, and manufacturing. Final demand comes from final consumers, such as households or exporting destinations, for pure consumption purposes. This transaction table traces flows of goods in monetary form among producers and purchasers (consumers). Rows represent sellers/producers, and columns represent buyers/consumers. A number in a cell indicates the amount of goods sold or transacted from a seller to a buyer. For instance, there are 100 monetary units of agriculture goods sold from the agriculture industry to the agriculture industry,

Table 10.6 A simple example of a transaction table

		To Buyer				
		Agriculture	**Mining**	**Manufacturing**	**Final Demand**	**Total Production**
From Seller	**Agriculture**	100	0	500	1000	1600
	Mining	0	20	1000	200	1200
	Manufacturing	200	500	1000	1500	3200
		Block A			Block Y	Block X

and 500 monetary units of manufacturing goods sold from the manufacturing industry to the mining industry.

The transaction table can be divided into several blocks. Block Y corresponds to the demand side, i.e., consumers, within an economy. Because of the existence of these demands, the economic machine, i.e., the supply side as presented by Block A, is producing to meet them. To produce, these industries must purchase from each other, and this generates additional demand. As a result, the amount of goods that the economic machine produces will be more than the amount that is consumed by final consumers. For instance, the pure consumption demand for manufacturing goods is 1500 monetary units. To have this amount of output to be sold to final consumers, the manufacturing industry needs to purchase from other industries. Similarly, other industries need to purchase from the manufacturing industry. As a result, to satisfy the final demands of 1000 units of agriculture, 200 units of mining, and 1500 units of manufacturing goods, the economic machine will produce 1600 units of agriculture, 1200 units of mining, and 3200 units of manufacturing goods. Block X represents the total production from every industry.

Input-Output Coefficients

Let us focus on the data on the first row and first column of the hypothetical example presented in Table 10.7. The first row presents the amount of goods sold from the agriculture industry to all industries and the final consumers, with the last number indicating the total agriculture production. These are output information for the agriculture industry. Every row represents the output activities of a row industry. The first column represents the amount of goods that the agriculture industry purchases from all industries for the purpose of producing to satisfy the final agriculture demand. These are inputs for the agriculture industry. Every column represents

Table 10.7 An illustration of the input-output activities for the agriculture industry

		To Buyer				
		Agriculture	**Mining**	**Manufacturing**	**Final Demand**	**Total Production**
From	**Agriculture**	100	0	500	1000	1600
Seller	**Mining**	0	20	1000	200	1200
	Manufacturing	200	500	1000	1500	3200
	Total Intermediate Inputs	300				

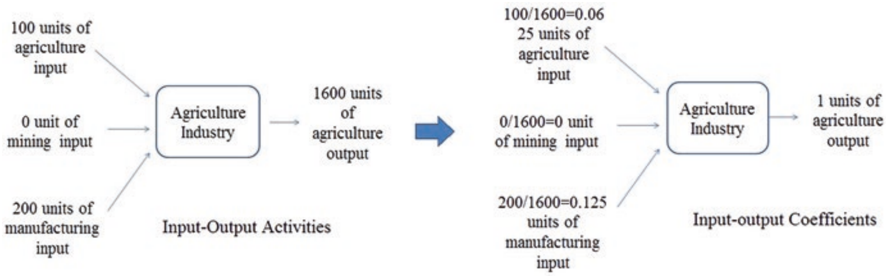


Fig. 10.10 Input output coefficients

the input activities for the column industry. For the agriculture industry, the total intermediate inputs are 300 and the total output is 1600. Through the process of the agriculture production, which involves adding labor and land, the ***value added*** to the 300 units of material inputs is 1300 units. This added value covers labor costs and taxes, and the rest will be a pure profit.

The transaction table contains input-output relations for all industries. The left side figure of Fig. 10.10 presents the input-output activities for the agricultural industry. To produce 1600 units of agriculture goods, the agriculture industry needs 100 units of agricultural inputs and 200 units of manufacturing inputs. On the right side of Fig. 10.10, these input-output relations are normalized to provide information about the inputs requirement for producing one unit of agriculture goods. These are ***input-output coefficients***. Unless there is a technology innovation, these normalized input-output relations are expected to be stable over time. For instance, if nowadays for producing 1 unit of agriculture goods, we need 0.0625 unit of agriculture good and 0.125 unit of manufacturing good as inputs, the same input-output relations should hold for the agriculture industry in the future, unless there are some revolutionary changes, such as invention of chemicals or new genetic modification technology.

Input-Output Matrix

An ***input-output (IO) matrix*** is a square matrix, containing all the input-output coefficients. Box 10.1 provides a quick refreshment of matrix. You can follow the following steps to construct such an IO matrix.

1. The dimension of the IO matrix depends on the number of industries in the economy. For instance, in the example of Table 10.6, there are three industries. Therefore, the input-output matrix is a three by three matrix, with three rows and three columns.
2. Every column contains the input-output coefficients for every industry. For instance, the following matrix is the input-output matrix for the example in

Box 10.1 The Concept of Matrix

A **matrix** is a rectangular array of numbers consisting of *n* horizontal rows and *m* vertical columns. A matrix is often presented in a set of brackets. The dimension of a matrix is described as *n***m* (# of rows * # of columns). The following are several examples of matrices.

$$\begin{bmatrix} 2 & 4 & 2 \\ 3 & 6 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 9 \\ ?10 & 14 \\ 9 & 8 \end{bmatrix} \quad : \begin{bmatrix} 3 & 1 & 1 \\ 2 & 5 & 2 \\ 4 & 4 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$2^?3$ matrix $3^?2$ matrix $3^?3$ matrix $3^?3$ matrix Identity matrix

A **square matrix** has the same number of rows and columns.

For an **identity matrix** (I), all elements off the diagonal are zero and all elements on the diagonal are 1

Table 10.6. The first column presents the input-output coefficients for the first industry (agriculture). To produce one unit of agriculture goods, the industry needs to buy 0.0625 unit of input from the agriculture industry, 0 unit from the mining industry, and 0.125 unit from the manufacturing industry. The second column presents the input-output coefficients for the second industry (mining). To produce one unit of mining goods, the industry needs 0 unit of input from the agriculture industry, 0.0167 unit from the mining industry, and 0.4167 unit from the manufacturing industry. The third column contains the input-output coefficients for the third industry (manufacturing). To produce one unit of manufacturing goods, the industry needs 0.1563 unit of input from the agriculture industry, 0.3125 unit of input from the mining industry, and 0.3125 unit of input from the manufacturing industry (Fig. 10.11).

The Leontief Matrix

The input-output coefficients link inputs to outputs within the economic machine, i.e., linking Block A to Block X in Table 10.6. This whole production is to meet the final demands (Block Y). How to find out the relations between the final demands and the production? With these relations available, one can estimate the level of economic activities for a set of future final demand, which can be projected based on the future population level. The ***Leontief matrix*** addresses this issue. In the following, we will first give instructions about constructing such a matrix, and then explain the meanings of different elements in the matrix.

The Leontief matrix is constructed based on the input-output coefficients. To construct it:

1. Check the dimension of the input-output coefficient matrix. In the above example, it is a 3*3 matrix.

Fig. 10.11 An IO matrix

$$\begin{bmatrix} 0.0625 & 0.0000 & 0.1563 \\ 0.0000 & 0.0167 & 0.3125 \\ 0.1250 & 0.4167 & 0.3125 \end{bmatrix}$$

2. Write out an identity matrix of the same dimension. An identity matrix has equal row and column numbers. Cells on the top-left to the lower-right diagonal line take values of one, and all other cells take values of zero. In the example, we need to write an identify matrix with three rows and columns, as:

$$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

3. Take the identity matrix minus the input-output coefficient matrix. For the matrix subtraction, do a cell-by-cell subtraction.

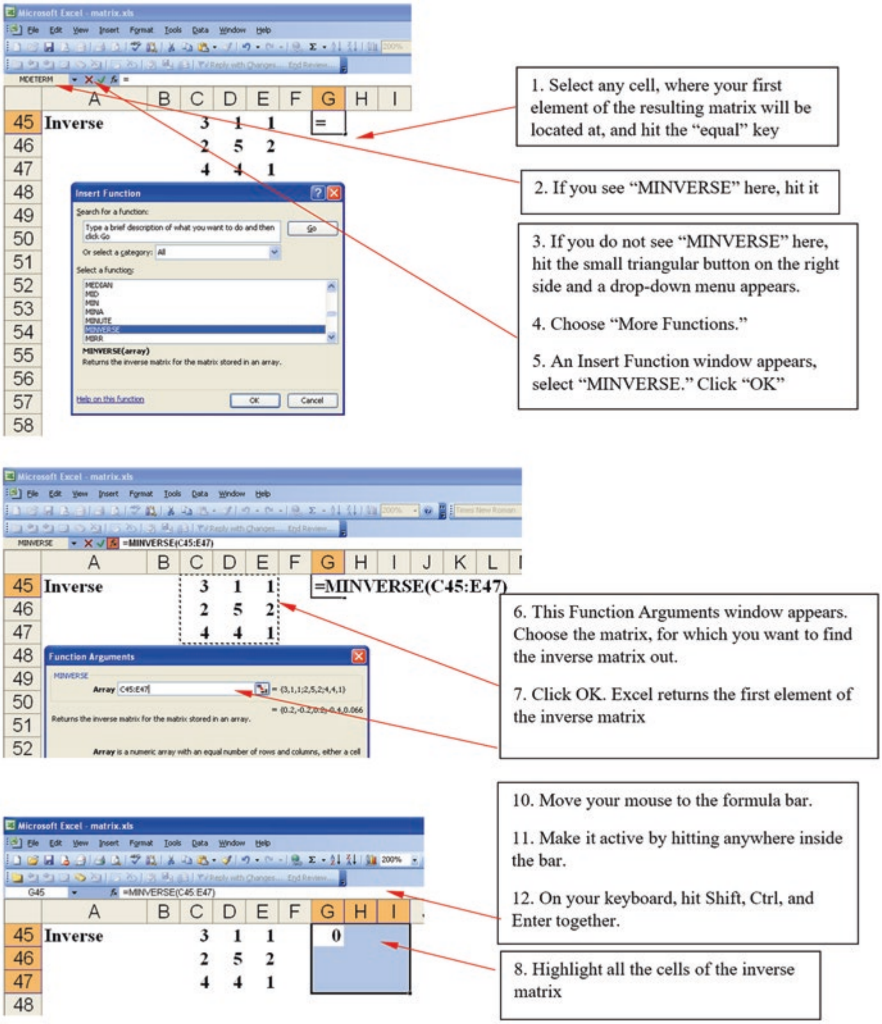
$$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} ? \begin{bmatrix} 0.0625 & 0.0000 & 0.1563 \\ 0.0000 & 0.0167 & 0.3125 \\ 0.1250 & 0.4167 & 0.3125 \end{bmatrix} = \begin{bmatrix} 0.9375 & 0.0000 & ?0.1563 \\ 0.0000 & 0.9833 & ?0.3125 \\ ?0.1250 & ?0.4167 & 0.6875 \end{bmatrix}$$

4. Take the inverse of the resulting matrix above. Please see Box 10.2 for how to find the inverse of a matrix. This inverse matrix is the Leontief matrix.

$$\begin{bmatrix} 0.9375 & 0.0000 & ?0.1563 \\ 0.0000 & 0.9833 & ?0.3125 \\ ?0.1250 & ?0.4167 & 0.6875 \end{bmatrix}^{?1} = \begin{bmatrix} 1.1083 & 0.1322 & 0.3119 \\ 0.0792 & 1.2689 & 0.5947 \\ 0.2495 & 0.7931 & 1.8717 \end{bmatrix}$$

The Leontief matrix contains ***Leontief coefficients***, which link final demands to outputs. Every column of the matrix represents how one unit increase of this type of final demand affects the production. For instance, in the Leontief matrix listed in step 4 above, the first column presents the Leontief coefficients for the first type of demand (of agriculture goods). With one unit increase of final demand of agriculture goods, there will be 1.1083 units increase of output from the agriculture industry (the first industry), 0.0792 unit increase of output from the mining industry (the second industry), and 0.2495 unit increase of output from the manufacturing industry (the third industry). The second column corresponds to the impacts of one unit increase of the second type of final demand (mining) on outputs. With 1 unit increase of mining final demand from final consumers, there will be 0.1322 unit increase of

Box 10.2 Use MINVERSE Function in Excel



output from the agriculture industry, 1.2689 units increase of mining output, and 0.7931 unit increase of manufacturing output. Similarly, the third column presents the impacts of one unit increase of the third type of final demand (manufacturing) on economic outputs.

Direct, Indirect, and Induced Effects

We are in a consumption economy. The economic machine is producing to meet the demand of consumers. Demand is constantly changing, depending on the population size, demographics, and people's preferences. The Leontief coefficients describe the direct and indirect effects of a unit change in a type of final demand on industries. Figure 10.12 presents these effects of the example.

The ***direct effect*** from a final demand is on its corresponding industry. For instance, 1 unit increase of agriculture final demand has a direct effect on the agriculture industry, causing an increase of 1.1083 units of agriculture output. One unit increase of mining final demand has a direct effect on the mining industry, causing an increase of 1.2689 units of mining output. One unit increase of manufacturing final demand has a direct effect on the manufacturing industry, causing an increase of 1.8717 units of manufacturing output.

Indirect effects are the results of industry-to-industry transactions that are indirectly caused by the increase of the final demand. For instance, in Fig. 10.12, if there is a one unit increase of agriculture final demand, the agriculture industry will be directly affected, and it will produce the right amount to satisfy this final demand. To produce, the agriculture industry needs to buy inputs from other industries. Therefore, the other industries will be indirectly affected. One unit increase of the agriculture final demand has indirect impacts on mining and manufacturing industries, causing a 0.0792-unit increase of mining production and a 0.2495-unit increase of manufacturing production.

Induced effects are related to the increased personal income due to the direct and indirect effects of the increase of a final demand. Industries experiencing increased revenue from the direct and indirect effects will subsequently increase payroll expenditures, by hiring more employees, increasing payroll hours, raising salaries and so on. Households will, in turn, increase spending in the economy, and this stimulates the economy to produce more.

Multiplier Analyses

- Output Multiplier

		Final Demand		
		Agriculture	Mining	Manufacturing
Industry	Agriculture	1.1083	0.1322	0.3119
	Mining	0.0792	1.2689	0.5947
	Manufacturing	0.2495	0.7931	1.8717

Direct effect (diagonal: 1.1083, 1.2689, 1.8717)
Indirect effect (all other cells)

Fig. 10.12 Direct and indirect effects

		Final Demand		
		Agriculture	Mining	Manufacturing
Industry	Agriculture	1.1083	0.1322	0.3119
	Mining	0.0792	1.2689	0.5947
	Manufacturing	0.2495	0.7931	1.8717
	Ouutput Multiplier	1.4370	2.1942	2.7779

Direct effect

Indirect effect

Fig. 10.13 Output multiplier calculation

Let us focus on a specific column of a Leontief matrix, as shown in the following. If we add up all the numbers of a column, it is the total direct and indirect impacts of one unit increase of a type of final demand on the economy. This is called ***output multiplier***. It describes the total amount of outputs stimulated by one monetary unit increase of a type of final demand. For instance, to respond towards one unit increase of agriculture final demand, the economic machine needs to produce 1.437 units of output. To respond to one unit increase in manufacturing final demand, the economic machine needs to produce 2.7779 units of output (Fig. 10.13).

- Employment Multiplier

  An ***employment multiplier*** describes the impacts of one unit of final demand on job creations. An increasing final demand will stimulate every industry to produce more. With expanding productions, all industries need to hire more workers. For the previous example, by studying the existing labor usage in production, one can know that every unit of agriculture output demands for 2 labors, every unit of mining output demands for 5 labors, and every unit of manufacturing output demands for 10 labors. The Leontief matrix informs 1.1083 units of direct impact from one unit of agriculture final demand on the agriculture industry. This requires hiring 2.2166 (1.1083*2) agriculture workers. The indirect impacts from one unit of agriculture final demand on the mining and the manufacturing industries are 0.0792 and 0.2495 unit of output. These require hiring 0.3960 agriculture worker (0.0792*5) and 2.4950 (0.2495*10) manufacturing workers. Totally, one unit increase of agriculture demand leads to adding 5.1076 jobs across the three industries. Similarly, the employment multiplier is 14.5399 for one unit increase of mining demand, and 22.3143 for one unit increase of manufacturing demand (Fig. 10.14).

- Income Multiplier

  Additional production also means more income for workers. Labor cost (i.e., income for workers) per unit of output information can be obtained by examining the existing payroll activities. ***Income multiplier*** describes the total workers' income across industries stimulated by one unit of specific type of final demand. Figure 10.15 illustrates the calculation. Given that labor cost per unit of output is 100, 200, and 300 for the agriculture, mining, and manufacturing industry, respectively, one can calculate the additional payments across the industries due to 1.1083 unit of increase of agriculture products, 0.0792 unit of increase of mining product,

	Final Demand						
	Agriculture	Mining	Manufacturing	Employment/ output	Agriculture	Mining	Manufacturing
Agriculture	1.1083	0.1322	0.3119	2	2.2166	0.2644	0.6238
Mining	0.0792	1.2689	0.5947	5	0.3960	6.3445	2.9735
Manufacturing	0.2495	0.7931	1.8717	10	2.4950	7.9310	18.7170
				Employment Multiplier=	5.1076	14.5399	22.3143

5*0.0792 = 0.3960
2*1.1083 = 2.2166
10*0.2495 = 2.4950

Fig. 10.14 Employment multiplier calculation

	Final Demand						
	Agriculture	Mining	Manufacturing	Labor cost/Output	Agriculture	Mining	Manufacturing
Agriculture	1.1083	0.1322	0.3119	100	111	13	31
Mining	0.0792	1.2689	0.5947	200	16	254	119
Manufacturing	0.2495	0.7931	1.8717	300	75	238	562
				Income Multiplier=	202	505	712

200*0.0792 =16
100*1.1083 =111
300*0.2495 =75

Fig. 10.15 Income multiplier calculation

and 0.2495 unit increase of manufacturing good. The income multiplier of one unit of agriculture final demand is 202. Similarly, the income multiplier for one unit of mining final demand is 505, and for one unit of manufacturing demand is 712.

References

Atkinson, R. D., Stewart, L. A., Andes, S. M., & Ezell, S. (2012). *Worse than the great depression: What experts are missing about American manufacturing decline* (pp. 327–339). Information Technology and Innovation Foundation.

Blakely, E. J., & Leigh, N. G. (2013). *Planning local economic development*. Sage.

Case, K. E., & Fair, R. C. (2007). Principles of microeconomics. Pearson Education.

DeSalvo, B., Limehouse, F., & Klimek, S. D. (2016). *Documenting the business register and related economic business data*. US Census Bureau Center for Economic Studies Paper No. CES-WP-16-17.

Gardocki, B. C., & Baj, J. (1985). *Methodology for estimating nondisclosure in County Business Patterns*. Center for Governmental Studies. Northern Illinois University.

Jorgenson, D. W. (2001). Information technology and the US economy. *American Economic Review, 91*(1), 1–32.

Kreahling, K. S., Smith, S. M., & Frumento, C. (1996). A procedure for estimating nondisclosed employment in the County Business Patterns data set. In *Staff Paper Series 299, Department of Agricultural Economics and Rural Sociology*. The Pennsylvania State University.

Leitner, H. (1990). Cities in pursuit of economic growth: The local state as entrepreneur. *Political Geography Quarterly, 9*(2), 146–170.

Li, J., & Whitaker, E. (2018). The impact of governmental highway investments on local economic outcome in the post-highway era. *Transportation Research Part A: Policy and Practice, 113*, 410–420.

Malizia, E., Feser, E. J., Renski, H., & Drucker, J. (2020). *Understanding local economic development*. Routledge.

Mathur, V. K. (1999). Human capital-based strategy for regional economic development. *Economic Development Quarterly, 13*(3), 203–216.

McCann, P., & Van Oort, F. (2019). Theories of agglomeration and regional economic growth: A historical review. In *Handbook of regional growth and development theories*. Edward Elgar Publishing.

Pierce, J. R., & Schott, P. K. (2016). The surprisingly swift decline of US manufacturing employment. *American Economic Review, 106*(7), 1632–1662.

Sechrist, R. P. (1986). *Procedures for estimating non-disclosed data from county business patterns. Research in contemporary and applied geography* (Vol. X, No. 1). Binghamton.

Stansel, D. (2005). Local decentralization and local economic growth: A cross-sectional examination of US metropolitan areas. *Journal of Urban Economics, 57*(1), 55–72.

Stijns, J. P. C. (2005). Natural resource abundance and economic growth revisited. *Resources Policy, 30*(2), 107–130.

Zhang, S. (2019). Rethinking US enterprise zones: The role of research design in program evaluation. *Local Economy, 34*(6), 545–571.

Zhang, S., & Guldmann, J. M. (2009). Estimating suppressed data in regional economic databases: A goal-programming approach. *European Journal of Operational Research, 192*(2), 521–537.

Zhang, S., & Guldmann, J. M. (2010). Accessibility, Diversity, Environmental Quality and the Dynamics of Intra-Urban Population and Employment Location. *Growth and Change, 41*(1), 85–114.

Zhang, S., & Guldmann, J. M. (2015). A regression-constrained optimization approach to estimating suppressed information using time-series data: Application to county business patterns 1999–2006. *International Regional Science Review, 38*(2), 119–150.

Chapter 11
Evaluation Research

Planners plan for various projects guided under goals, objectives, and policies, such as transportation improvement projects, affordable housing projects, and urban redevelopment projects. Using a transportation improvement project (e.g. reconfiguring a road from six lanes to four lanes, and adding bike lanes) as an example, a feasibility analysis needs to be conducted before implementing the project, and an evaluation study needs to be completed after the implementation to measure whether the project makes sense financially, socially, and environmentally. The process of conducting the feasibility analysis is part of the evaluation process, although evaluation also entails evaluating the effectiveness and impact of a project or a policy. In this sense, evaluation research includes ***ex-ante evaluation research*** and ***ex-post evaluation research***, simply put, evaluation before and after the project is implemented.

Ex-ante evaluation research measures the cost effectiveness of the project, estimating the costs, sources of funds, regulatory environment, social environment, natural environment, and other elements, to ensure the success of the project. Measuring the impact and success of the project is part of the ex-post evaluation research, meaning evaluating the project or policy after it is implemented. Success of the project depends on a myriad of factors, such as whether the project has been under the budget, or the project has realized the initial goal such as safety of the bikers and pedestrians and the positive economic impact that the project brings. Operational feasibility can be used as a measure of success for certain projects as well. If the project is predicted to yield a positive cash flow over the course of 20 years, the project may be also regarded as successful from the angle of financial feasibility. Twenty years is a long time, and unforeseeable events may happen; however, a cash flow projection, which incorporates certain level of volatility and sensitivity, will provide a general guidance on the future revenue and profit of the project.

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_11].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_11

The actual cash flows after the project is implemented, therefore, can be also used as a tool to measure success. Ex-post policy evaluation often requires using specific projects under the policy to measure the impact and success.

Overall, the evaluation research involves cost-benefit analysis of a project or a policy. Therefore, this chapter will cover various techniques and methods used in evaluation, before and after the project is implemented.

Fundamentals of Budgeting

Budgeting is the decision process to decide on the monetary value to be spent on activities and projects based on priority. The budgeting process should indicate what will be done, when will it be done, how long it takes, and how these activities and projects will be financed. Base budget of an operation or project usually includes ***line-item budget*** for resources purchased, ***performance budget*** for implementing the activity or project, ***program budget*** for services delivered, and other types of budget.

In public budgeting the budget cycle starts with budget preparation, budget adoption, agency execution, audit or evaluation, and then restart the cycle. By law, government agencies are required to prepare a budget and have the budget adopted after public hearings. Typically speaking, the ad valorem taxes, meaning taxes based on assessed values of properties, mostly real estate, are the largest share of revenues in local governments (see Fig. 11.1). Some other revenues include carryovers from previous *fiscal years* (FY), user fees, licenses and permits, etc. Fiscal years are typically different from calendar years and vary per different entities. The expenditure items usually include personnel expenses, operating expenses, capital project expenses, and others (see Fig. 11.2). Personnel expenses are oftentimes the largest expenses.

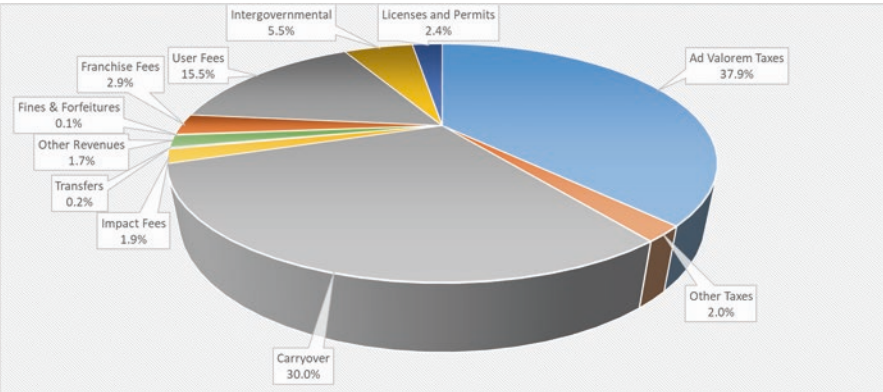


Fig. 11.1 Total Revenue Sources, City of Palm Beach Gardens, FL, FY 2020/2021

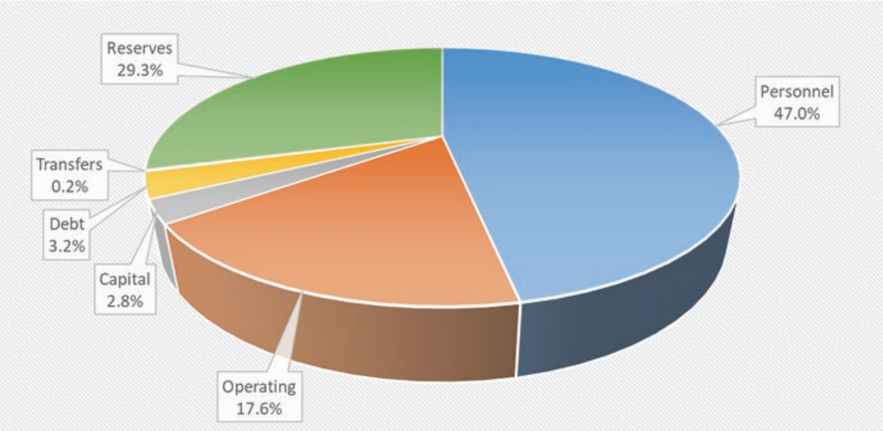


Fig. 11.2 Total Expenditures, City of Palm Beach Gardens, FL, FY 2020/2021

Ex-ante Evaluation Methods

Any project or course of action, regardless of disciplines, needs deliberate and careful planning, particularly so for projects that cost manpower or money, and needs to present significant benefits and outcomes once implemented. The first step of planning for a project is to develop the goals that the projects intend to achieve. Goals are different for different entities and stakeholders. For example, economic development goals for local communities often center around tax revenues and jobs. However, businesses which pay taxes and provide employment opportunities may often think about profit as their goals, especially if they are for profit entities. Even for nonprofit agencies being able to be solvent or at least break even is one of the operational goals, in addition to the social goals that these agencies often strive to achieve. Public agencies, such as the planning department and the municipal government, need to have public interests in mind, balancing the interests of private entities, residents, and other parties affected by a project.

Goals are broad while objectives are narrower, specific, and often have tangible outcome measures. For example, increasing job opportunities can be one goal for a potential project, while objectives will be outlining the specific number and types of jobs which should be created. Once a project has goals in mind, it is easier to conduct evaluation analysis before it is implemented.

The following sections will introduce various methods to evaluate whether a project is feasible. ***Feasibility*** is the likelihood of successful implementation of a project to achieve intended goals and outcomes, usually through a set of selected courses of action and under a geographic and political context. Essentially feasibility analysis is cost-benefit analysis. When the intended benefits are larger than the costs based on the project goals, the project is said to be feasible to be implemented. Benefits can be tangible, intangible, direct, or indirect. Costs can be tangible, intangible, direct, or indirect. Intangible benefits and costs are mostly associated with social benefits and costs.

Site Feasibility Analysis

Site feasibility analysis is mostly pertinent to planning projects that are site specific. Planning projects that are site specific include real estate development projects, transportation projects, historic preservation projects, environmental preservation projects, and other projects that are located in certain areas.

Site feasibility analysis is mostly conducted in real estate development, although different types of real estate development tend to have different site requirements. Generally speaking, site feasibility analysis should cover the following aspects:

- *Physical factors*
- *Legal factors*
- *Political factors*
- *Off-site factors*

Physical factors often focus on topography, soil and subsurface conditions, hydrology and drainage requirements, and stormwater runoff. All of the residential, commercial, and office development projects need to consider subsurface conditions, such as avoiding sink holes and environmentally sensitive areas. While topography either aids or hinders drainage, proper design and installation of drainage and stormwater runoff is imperative in any type of new developments.

When looking at site features, some of the opportunities to the project include views and vintage points, vegetation, existing water bodies, and other unique site features such as proximity to grocery stores, train stations, downtown, etc. Waterbodies and open space are two of the major physical features that are linked to residential satisfaction (Cao & Wang, 2016; Kaplan, 1985). However, large open spaces play less important roles than nearby trees, well-landscaped grounds, and places for walks. Additionally, residents value opportunities to grow plants as a sense of community (Kaplan, 1985). Overall, improving parks and open space, neighborhood safety, and neighborhood appearance is important to enhance quality of life and residential satisfaction (Cao & Wang, 2016). Physical site features and design can also be used to enhance safety and deter crime. CPTED (Crime Prevention through Environmental Design) leads initiatives incorporating safety features when designing a place.

Constraints of the site are often from environmental considerations, the shape of the parcel(s), and local regulations. Site specific constraints include brownfields, wetlands, waterways and drainage easement, flood plains, open space and conservation areas, and climate. Brownfields are sites that are either perceived to be contaminated or actually contaminated by pollutants. Development on brownfields often requires mitigation of the potential health impact of the sites prior to project implementation.

Off-site conditions, such as transportation networks (streets, public transit, sidewalks, traffic, etc.) and utilities (power line, gas line, water, and sewage), are also critical to ensure the success of implemented planning projects. Most contemporary developments use electricity, water, and sewage systems from an incorporated municipality; however, in more rural areas or unincorporated places well water or on-site septic tank systems may be needed to provide freshwater and handle wastewater. For more advanced environmental features, development may also consider reuse of graywater, waste water with minimal contaminants, such as from shower and washing machines. Electricity is usually provided by a company; however, a new development may use renewable energy sources, such as solar, wind, or water, to generate electricity. There are notable examples of "solar villages" or "solar towns" in the world where solar power is used to provide electricity for residential, office, or commercial uses nearby (Dharmadasa et al., 2015). Large-scale solar farms may be able to provide electricity sufficient for an entire town, not just the new developments. Windmills have been used in certain places; however, arguments about sky/air rights and noise mitigation have been challenging factors to implement such windmill projects. When considering sites for green buildings, water pollution reduction, human health protection, dust and air emission mitigation, sufficient green space consideration, and construction noise mitigation are the most important criteria in site planning and design (Huo et al., 2019). Therefore, environmental protection consideration, effective use of space, use of natural and existing resources, green parking and thermal environment, and efficient use of land resources are the five determining factors in site selection, planning, and design.

Data sources for site analysis depends on the type of analysis conducted. Maps, drawings and test, and land survey are all the tools that can be used in site analysis. If it is about topography, the U.S. Geological Survey (USGS) provides digital elevation models for professionals to extract information about the topography of the site. Topography analysis is mostly useful when the site is located in areas with uneven elevation, or in hilly areas. Most of the site analysis requires a base map, which is often from the appraiser's office of a local community. In the U.S., county appraisers usually maintain property parcel data with basic characteristics such as land use, size, geographic coordinates, address, appraised value, and property taxes levied. If the parcel has existing housing or building structures, the characteristics of the buildings, such as size, number of rooms, and other pertinent attributes, can be found from the appraiser parcel data as well. Most of the parcel data is stored as geographical databases where the user can easily retrieve the data and use them in GIS software, such as ArcGIS, to map or conduct spatial analysis.

Other site conditions, such as views, vegetation, and water features, can be from site reconnaissance or from the GIS databases maintained by local governments. Many of the U.S. county governments establish and maintain various GIS databases for features within the county, and these databases are easily available to the public to download. For example, many counties maintain land use, roads, water bodies, cultural facilities, medical facilities, and other GIS databases. Many state government or universities provide GIS data for the state. For example, Florida Geographic Data Library (FGDL) is a notable example of geodatabases which include a large amount of data ranging from physical features, economic and business data, to environmental and transportation planning data.

Conducting ***due diligence*** of the site-specific conditions also requires checking on local regulations to understand the land uses of the site, zoning requirements, building codes and design guidelines, and whether the site is located in a subdivision. Appropriate environmental assessments and approvals need to be processed if the environmental impact is significant enough that site is subject to any Environmental Impact Assessment (EIA).

Regulatory and Political Feasibility Analysis

Planning projects are heavily influenced by *political climate* of the local community, in addition to multiple levels and layers of regulatory requirements. ***Regulatory factors*** range from the federal and state levels, to local level regulations. Federal levels of regulations most relate to funded projects under different branches and the federal government, such as the U.S. Department of Housing and Urban Development (HUD), the U.S. Department of Agriculture (USDA), the Environmental Protection Agency (EPA), the U.S. Department of Transportation (DOT), or other branches that are closely related to urban and regional planning. However, sometimes projects are under the scrutiny of the U.S. Constitution and its amendments, such as the First Amendment (freedom of speech and assembly), the Fifth Amendment (the takings clause), and the Fourteenth Amendment (the due process clause).

Most projects are subject to local regulations, such as comprehensive plans, district plans, neighborhood plans, zoning, land use, subdivision regulations, environmental approvals associated with flood plains, conservation areas, coastal areas, Superfund/brownfield areas, building codes, and design guidelines, etc. Prior to implementation a project needs to demonstrate compliance with all applicable regulations. At times, request of rezoning or variance, or setting up special zoning districts (such as Overlays, Planned Unit Development, etc.) may be needed in order to implement the project.

Market Feasibility Analysis

Any type of planning projects usually requires a ***market feasibility analysis***, especially if the projects intend to serve potential customers and making profit. Market feasibility analysis centers around the supply and demand analysis. For example, target market identification and market capture ratio are all related to the demand side of market conditions. Knowing the characteristics of the demand is the key to determining the products to provide and where the project should be located. Market analysis is often used to (1) assess market feasibility for economic uses and type of development at specific sites; (2) define target industries or development projects to attract to a community; (3) determine market potential/recruitment targets for commercial districts; (4) improve understanding of customer markets for existing businesses; (5) inform zoning and land uses; and (6) secure funding from financial institutions.

There are four types of market feasibility studies: ***General studies***, ***site-specific studies***, ***highest-and-best-use studies***, and ***target market profile studies***. General studies of market feasibility usually focus on an entire industry, such as the office market outlook for a metropolitan area. General studies use aggregated data to describe the current demand and supply conditions, and then project the future market conditions. Site-specific studies focus on specific project site and explore market potentials of the project if located in that specific site. Many real estate developers use this approach if they intend to develop a project on a selected site. Highest-and-best use studies focuses on identifying the highest profit that a project may generate and the best uses to achieve such profit. Highest-and-best use is a fundamental principle in real estate development. Target market profile studies focuses on analyzing demand conditions of a product or a service. Target market profile may be site specific, industry specific, product specific, or service specific.

Market feasibility analysis covered under this chapter focuses on site-specific feasibility analysis, although many of the methodology can be used in other types of market feasibility analysis. There are two different approaches to market feasibility analysis. The first approach is to identify the best site to implement a project. The second approach is to focus on the market conditions once the site is identified. In the U.S., data used in market feasibility analysis are mostly from the demographic census of the U.S. Census Bureau, business receipts from local government or local chamber of commerce, tax revenues from local taxing authorities, and third-party data vendors. Some notable third-party data vendors for market analysis include ESRI Business Analyst, which has an online portal, and a desktop version. ESRI Business Analyst Online (BAO) is an online application that is very easy to use and includes large amounts of data to conduct a market analysis. The desktop version is more complicated to use but provides more functionality than the online version.

Identifying Suitable Sites

Identifying suitable sites uses multiple criteria to narrow down sites meeting these criteria, and then identifies the best suitable site by comparing specific site conditions. Some of the criteria may include the characteristics of potential consumers, such as household income, educational levels, and median housing values, etc. For example, when deciding on where to locate a Wholefoods Supermarket, it may be critical to find a target market which have middle to high income, is located along arterial roads, and on a site with large enough space to accommodate a grocery anchor store. The company has its own criteria about the market characteristics (Wholefoodsmarket.com., 2020):

- Typically, 200,000 people or more in a 20-minute drive time
- 25,000–50,000 Square Feet
- Large number of college-educated residents
- Abundant parking available for the store's exclusive use
- Stand alone preferred, would consider complementary
- Easy access from roadways, lighted intersection
- Excellent visibility, directly off of the street
- Must be located in a high traffic area (foot and/or vehicle)

The first step of finding a suitable site is, therefore, looking for areas meeting the size conditions, large, in vacant or redevelopable areas. Data for vacant parcels can be obtained from county appraisers and redevelopable areas can be identified through planning documents or field observations. Once the potential sites are determined, these sites should overlay with demographic conditions of a census track or census block groups, such as a high percentage of population with a college degree. This percentage is often calculated based on population older than 25. For example, the average percentage in a metropolitan area is about 30%. Maybe a criteria is to set as at least 50% of the population over the age of 25 has a college degree. This criterion can be also set based on the consumer demographics of high-performing stores. After narrowing down the areas further, other conditions, such as traffic and road network, can be added to the overlay. Once all the conditions are met, the site will be identified. If there are still a few suitable sites, further comparison is often needed to narrow down to one site with the most suitable conditions. Data about traffic and transportation networks are often from state governments, such as the Department of Transportation, although some third-party vendors, including ESRI BAO, has some traffic data that can be easily retrieved when necessary.

Site-Specific Market Potentials

Once the project site is identified, a detailed market analysis is often warranted, especially if the site is chosen based on only a few criteria, and without sufficient supporting facts. Conducting market analysis always needs to focus on supply and demand, and site-specific demand analysis stresses the characteristics and consumption behavior of the consumers in the target market areas. Supply-side analysis focuses on competitive advantage and analysis of competitors. Before conducting demand and supply side analyses, it is imperative to delineate market areas. There are a few approaches identifying the potential market areas and the following section explains these methods in detail. Some of the methods, such as Riley's law of gravitation, is overly simplified, but sometimes does provide a quick analysis of the market areas when considering competitors. Some other methods, such as GIS-based methods, are heavily data driven and may provide a better-defined market area.

Target Market Identification

Identifying the target market, particularly for those products and services that serve local areas, is the first step of market feasibility analysis. Some large corporations often have a market area larger than local areas, and these types of planning projects may not need to focus on market feasibility analysis; on the other hand, projects like these may benefit from analyzing the potential impact these projects have on local communities. There are a few methods in identifying local market areas, one is called Riley's law of gravitation, and another is using GIS as a tool to identify drive time, ring buffer, walk time, etc., depending on residents' travel behavior in local communities.

Riley's Law of Gravitation

Riley's law of gravitation is a formula derived based on gravitation theories, using demographic or social parameters to measure the gravity, i.e., the range of a market areas due to the attractive "gravity" of population, products, or services. The basic formula used to calculate market breakpoint is delineated below:

$$B_i = \frac{D_{ij}}{1+\sqrt{\frac{P_j}{P_i}}}$$

Where B_i is the market breakpoint of place i, in comparison of place j. D_{ij} is the distance between i and j. P_j is the population size, number of stores, or types of services in place j, depending on whether i and j are towns or shopping centers. P_i is the population size, number of stores, or types of services in place i.

For two towns, if town i has 25,000 residents, and town j has 30,000 residents, the distance is 35 miles. Then the market breakpoint from town i is:

$$B_i = \frac{35}{1+\sqrt{\frac{30,000}{25,000}}} = 16.7\,miles.$$

Then B_j is 35–16.7 miles = 18.3 miles.

Another example of using Riley's law of gravitation is to calculate the market breakpoints among competitive shopping centers, where P_i/P_j is the number of stores in shopping centers i and j, since presumably a higher number of stores indicate more attraction to consumers. However, this approach does not consider other characteristics of these stores, such as types of stores or stores catering to certain demographic segments.

Zip Code Tabulation to Capture Trade Area Data

Businesses often collect zip codes of the consumers; however, some consumers may not be willing to give out the zip code they are from. The simple approach of collecting zip codes of the consumers may be able to help researchers or practitioners find the zip codes with the largest number or the smallest number of consumers. This way a roughly estimated trade area can be created through the tabulation process.

Using GIS to Delineate Trade Areas

GIS has been widely used in analyzing various conditions for urban and regional planning. In market analysis, GIS has helped to map out typical travel distance of consumers, thus resulting in the trade area around a specific site. Drive and walk time (5 minutes, 10 minutes, and 25 minutes are usually used but rather arbitrary) are the mostly used criteria in the U.S., but specific criteria is subject to local consumer characteristics and travel behavior. If businesses have customer street addresses and/or zip codes, GIS can be used to map these addresses or tabulate the zip codes, thus creating a trade area for specific businesses, sites, or districts. Figure 11.3 delineated a drive-time distribution, 10 minutes, 20 minutes, and 30 minutes of Sawgrass Mills in Sunrise, FL.

30-minute drive time is used since this outlet shopping mall has a regional, even international, consumer bases. Due to the proximity of South Florida to the Caribbean and the Bahamas, international customers may be able to come to the mall and go home within 1 day.

Understanding Local Area Market Conditions

Certain market analysis is not based on specific project sites, but local area market conditions (University of Wisconsin-Madison, 2020). For example, analyzing the market conditions of a commercial district often starts with creating a business and building inventory in the district. Business mixes, meaning the type of businesses, should be analyzed. List of businesses can be obtained from local chamber of commerce, as well as from online search engines. Building inventory often relies on the property data from the local appraisers' office. Consumer demographics, such as age, income, education, race and ethnicity, educational attainment, etc. are often critical of understanding the demand conditions. Meanwhile, lifestyle characteristics such as consumption and expenditure patterns also play important roles in understanding local demand conditions. Demographic profile can be generated through analyzing the Census data directly, or using third-party data vendors such as ESRI Business Analyst. ESRI Business Analyst can generate demographic profiles and segmentation per drive time/walk time/rings, of census geographies, offering flexibility and infographics relevant to specific site or district (see Fig. 11.4 for

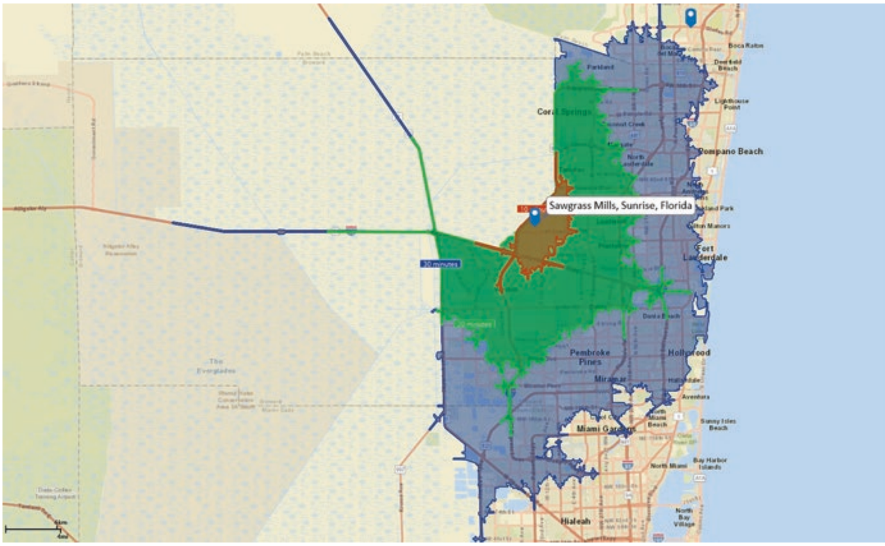


Fig. 11.3 An Example of Trade Area by Drive Time. (Created by the authors from ESRI BAO (Business Analyst Online))

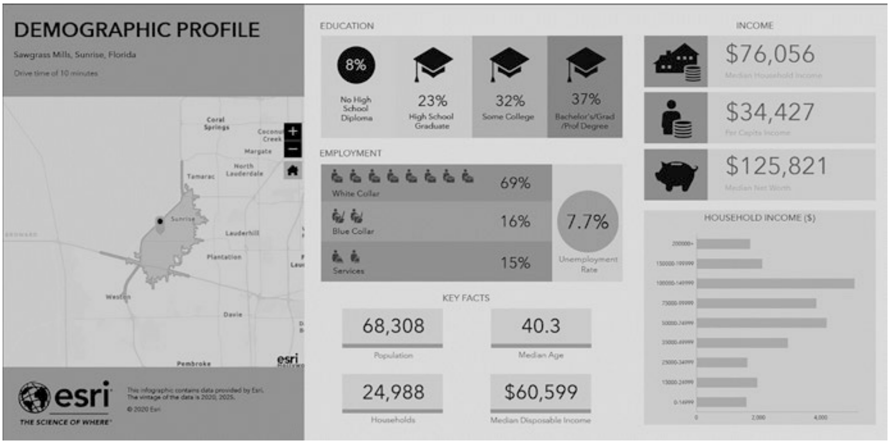


Fig. 11.4 Demographic Profile. (Created by the authors from ESRI BAO (Business Analyst Online))

infographic annotation of 10-minute drive time. The infographics can be obtained for all the selected drive time/walk time/ring distance).

Local economic conditions, on the other hand, create a larger environment for the business district. A surplus and leakage analysis as defined by ESRI Business Analyst can often indicate which industries are overly supplied, while which industries are inadequate in the area, and thus the area is leaking consumers to businesses outside the area. Figure 11.5 shows a surplus and leakage analysis of businesses within 10-minute drive-time of Sawgrass Mills Mall, located in Sunrise, Florida.

Using methods introduced earlier, trade areas of the commercial district can be calculated. Another critical component analyzing local market conditions is to conduct surveys, focus groups, or public meetings of the business owners and consumers. The purpose of those surveys is to identify the needs and opportunities of key stakeholders. If possible, business and workforce data, market and marketing data, and sales data should be obtained to understand the current conditions of businesses.

When proposing future plans for local areas, it is often important to conduct case studies in similar cities with similar socioeconomic conditions. Cases should include best-practice scenarios to learn of success stories and lessons.

After collecting these relevant market conditions, a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis will further help identify the market potential of different sectors, such as retail and service industries, restaurants, arts & entertainment, housing, office space, or lodging facilities.

A competitive advantage analysis, often based on Michael Porter's Diamond Model of National Advantage, may also provide a glimpse of the economic conditions of a local area (see Fig. 11.6 for the diamond model).

Financial Feasibility Analysis

Financial analysis is the key to project success, and planning projects are no exceptions. Planning projects are either projects related to making plans, planning analysis, or planned development projects. Funding sources are imperative to implement any type of projects and funding sources are often tied to the projected cash flows and profitability of the projects. For example, if a project needs to borrow money from investors or financial institutions potential financial performance of the project is the fundamental criteria for investors and financial institutions to loan funds to the project.

Time Value of Money

Due to the influence of inflation, market risks, and other various factors, money bears time value. Today's one dollar will worth more than one dollar in the future, while worth less than one dollar in the past. A discount rate, or interest rate, determined by risk-free rate or weighted average costs of capital, is used to discount or compound money. Costs of capital indicate borrowing costs to raise capital and one of the notable examples of borrowing costs is the interests paying to the lenders. The following formulas are used to convert money between present and future.

$$FV = PV\left(1+r\right)^{t}$$

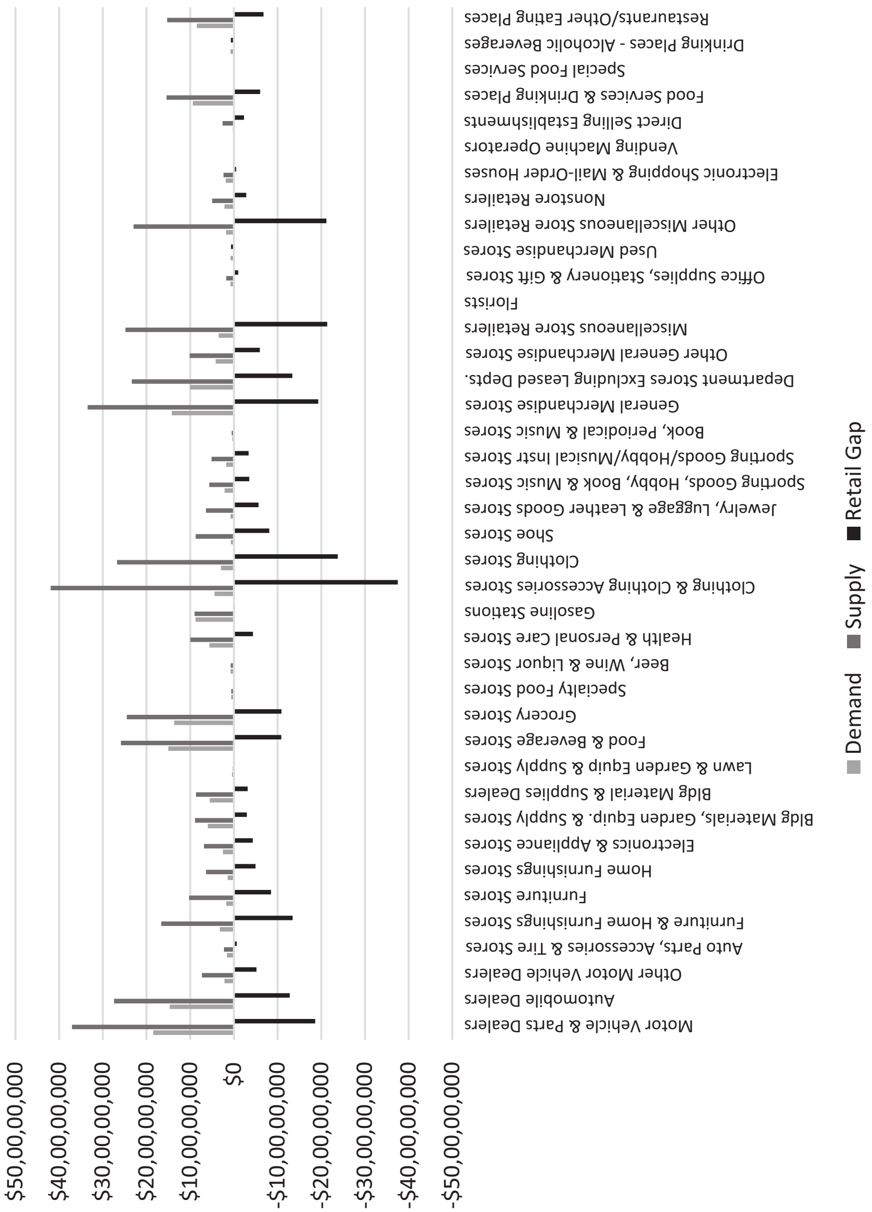


Fig. 11.5 Retail Gap Analysis. (Created by the authors from ESRI BAO (Business Analyst Online))

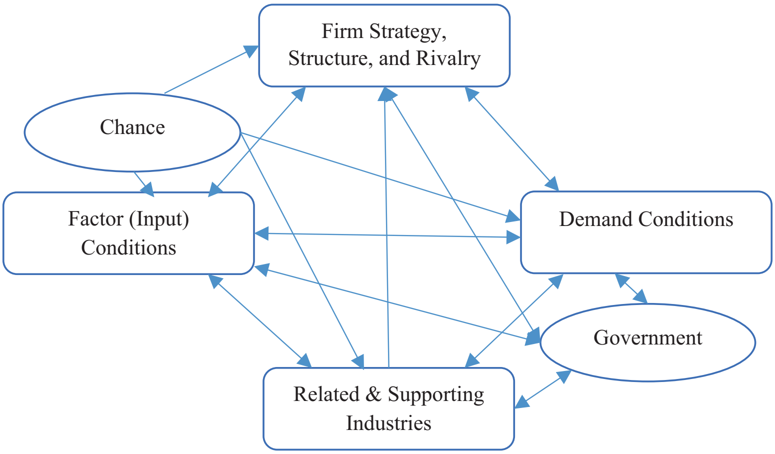


Fig. 11.6 Porter's Diamond Model of National Advantage. (Adapted from "The Diamond Model", Institute for Strategy & Competitiveness, Harvard Business School)

$$PV = \frac{FV}{(1+r)^t}$$

Where FV is future value, and PV is present value. r is discount rate, and t is the elapsed time t. The fundamental principle of time value of money is the basis of many financial calculations, such as annuity, mortgage, cash flow, etc. Therefore, it is imperative to understand how to convert future value into present value, and vice versa.

Sources of Funds Analysis

In private corporate world there are three fundamental methods to raising capital. The first method is ***debt financing***, which means borrowing from financial institutions, or from private and public entities and investors who invest in bonds. Bonds are a type of financial vehicle which are often issued by government entities to raise funds for public purposes. For example, many local municipalities issue revenue bonds to raise capital for public projects. Investors are promised a fixed return (dividend) from the bonds. Issuing bonds is associated with certain costs. When borrowing from financial institutions, there are various costs as well. These costs include interests that the borrower has to pay, the loan origination and closing costs, and other costs associated with raising debt capital.

A second approach to raising funds is through ***equity financing***. Equity means ownership and equity financing indicates selling ownership shares to raise funds. Therefore, the most common type of equity financing is financing through stocks. After a private company initiated Initial Public Offerings (IPO) the company becomes "public" and has the right issuing stocks to be sold to the investors. In turn the funds received by the company will be used in production or expansion of businesses. Investors gain or lose investment through increased or decreased share values of company stocks.

When combining debt and equity financing the approach is called ***hybrid financing***. Many businesses choose to use a combination of debt and equity financing to fund businesses. Regardless of the methods used in raising capital, there are costs associated with raising capital. In case of debt financing, the cost of capital is often associated with the interest. Therefore, to break even, the return rate of a project needs to be larger than the interest rate, which in this case is used to measure the average costs of capital. The break-even return rate is called Internal Rate of Return (IRR). When the cost of capital rate (e.g. interest rate of borrowing capital) is smaller than the IRR, projects tend to be profitable. When discounting future cash flows to today's value, Net Present Value (NPV), the discount rate is viewed as the average costs of capital. To maximize NPV, the discount rate (cost of capital) needs to be the smallest possible, but should be at least smaller than an IRR. For example, IRR can be calculated based on initial costs and projected net cash flows. If the IRR is calculated as 12% to break even, then the interest rate or the cost of capital should be smaller than 12%. Therefore, IRR can be used as a guiding threshold rate for cost of capital, when raising capital through debt, equity, or hybrid approaches.

For nonprofit and government agencies, funding often comes from tax revenues, various fees, loans with low interest rate, or grants. Many types of foundations and many government agencies provide grants to qualified small businesses, nonproft agencies, or individual households.

Uses of Funds Analysis

Uses of funds analysis relates to the budgeting process. In real estate and housing development, ***uses of funds*** are divided into hard costs and soft costs. Hard costs are associated with tangible assets, such as land acquisition and purchase of construction materials. Soft costs are often associated with personnel, consulting, or intangible costs; for example, architectural design, management, or payroll are soft costs.

Balance Sheet and Income Statement

Balance sheet, ***income statement***, and ***cash flow statement*** are the three fundamental financial statements indicating the financial status of an organization (Ross et al., 2013). Balance sheet measures how assets, liabilities, and shareholders' equity is balanced at a specific time (e.g. Table 11.1). Income statement demonstrates the sources of revenues, itemized expenses, and net income after all the expenses (e.g. Table 11.2). Cash flow statement extends the income statement by including multiple periods' revenues, expenses, net income, and critical financial ratios. Cash flow analysis can be retrospective or projected, depending on the purpose of the analysis. Retrospective cash flows are usually included in an annual report and can be used as

Table 11.1 Abridged balance sheet for Federal National Mortgage Association (Fannie Mae), a Publicly Traded Government Sponsored Enterprise

Fiscal year is January–December. All values USD Millions.	2020	2019	2018	2017	2016
Assets					
Cash & short term investments	115,624	61,407	49,423	60,260	62,177
Cash only	115,623	61,407	49,423	60,260	62,177
Short-term receivables	10,497	9258	9235	8991	8737
Investments – total	3,825,582	3,391,494	3,319,928	3,232,529	3,156,194
Net loans	3,659,144	3,327,389	3,241,694	3,173,537	3,076,854
Gross loans – net of unearned income	3,669,696	3,336,405	3,255,897	3,192,621	3,100,319
Loan loss allowances (reserves)	−10,552	−9016	−14,203	−19,084	−23,465
Other investments	166,438	64,105	78,234	58,992	79,340
Net property, plant & equipment	1261	2366	2584	3220	4489
Other assets (including intangibles)	19,838	26,884	23,960	23,179	22,841
Other assets	18,689	24,594	21,376	20,001	22,841
Total assets	3,986,057	3,503,319	3,418,318	3,345,529	3,287,968
Liabilities & shareholders' equity					
Total debt	3,935,736	3,467,386	3,391,920	3,330,054	3,262,316
ST Debt & current portion LT debt	12,173	26,662	24,896	33,756	35,579
Short term debt	12,173	26,662	24,896	33,756	35,579
Long-term debt	3,923,563	3,440,724	3,367,024	3,296,298	3,226,737
LT debt excl. capitalized leases	3,923,563	3,440,724	3,367,024	3,296,298	3,226,737
Deferred taxes	−12,947	−11,910	−13,188	−17,350	−33,530
Other liabilities (excl. deferred income)	24,754	21,325	20,158	19,161	19,581
Total liabilities	3,960,798	3,488,711	3,412,078	3,349,215	3,281,897
Total equity	25,259	14,608	6240	−3686	6071
Liabilities & shareholders' equity	3,986,057	3,503,319	3,418,318	3,345,529	3,287,968

Source: Fannie Mae

a reference to examine the financial wellbeing of an entity. Projected cash flow analysis is used to measure whether a project or an entity is feasible to operate in the near future. Cash flow analysis will be covered in a later section in this chapter.

Costs and Benefits

The essence of financial feasibility analysis is cost-benefit analysis. Cost and benefit can be direct costs/benefits, indirect costs/benefits, and tangible versus intangible costs/benefits. ***Direct costs/benefits*** are directly relating to the product or service, such as labor and materials, and product revenues; while ***indirect costs*** often associate with running or maintaining the businesses, for example, office supplies, utilities, computers, cell phones, etc. ***Indirect benefits*** relate to indirect impacts that the production generates, such as more jobs in other sectors. ***Intangible costs*** are costs that cannot be easily quantified. For example, many urban redevelopment projects are associated with unintended residential displacement caused by gentrification. It

Table 11.2 Abridged income statement for Federal National Mortgage Association (Fannie Mae), a Publicly Traded Government Sponsored Enterprise

Fiscal year is January–December. All values USD Millions.	2020	2019	2018	2017	2016
Sales/revenue	109,451	122,840	119,106	113,800	108,306
Total investment income	907	1770	952	1527	1291
Interest income	107,569	119,572	117,049	109,856	106,021
Trading account income	513	322	126	190	28
Trust income, commissions & fees	462	1176	979	2227	966
Commission & fee Income	462	1176	979	2227	966
Total expense	11,869	105,263	99,007	95,348	89,938
Total interest expense	82,703	98,610	96,098	89,123	84,726
Loan loss provision	678	−4011	−3309	−2041	−2155
Selling, general & admin. expenses	3068	3023	3059	2737	2555
Other operating expense	5450	5209	875	3433	2967
Operating income	97,582	17,577	20,099	18,452	18,368
Pretax income	14,879	17,577	20,099	18,447	18,333
Income tax – current – domestic	3803	2089	−114	−600	2014
Income tax – deferred – domestic	−729	1328	4254	16,584	4006
Net income	11,805	14,160	15,959	2463	12,313

Source: Fannie Mae

is difficult to monetize such costs and therefore it is intangible. ***Tangible costs*** are costs that can be easily identified and measured with monetary currencies.

Total costs are the total expenses for the project, and average costs are costs averagely divided among different projects. ***Fixed costs*** do not change along with the business output, but ***variable costs*** change. Therefore, fixed costs are mostly about real estate and machineries. However, fixed costs vary greatly per types of industries. Borrowing costs are related to the costs when using debt financing. Operating and maintenance costs are costs associated with day-to-day operation of the business. Opportunity costs are costs of resources diverted from other uses; for example, giving up on the potential profit from a commercial development project to develop affordable housing projects.

Another important concept in business finance is that when marginal benefit equals marginal cost the net benefit is maximized. Marginal benefit is the benefit generated by the production of an additional unit of output; while marginal cost is the cost incurred by the production of an additional unit of output. This principle can be used in public budgeting when projecting the size and user price of certain infrastructure, such as parking garages (Patton et al., 2013).

Cash Flow and Pro-forma Analysis

Cash flow analysis is the analysis to project the future cash flows, meaning revenues, expenses, and net cash flows (sometimes called net income), which is the difference between revenues (benefits) and expenses (costs), to measure the profitability of a business operation (see Table 11.3 and Table 11.4).

Table 11.3 Example of net cash flow

Year	0	1	2	3	4	5	6
Net cash flow/net income	-$1000	$200	$200	$200	$200	$200	$200

For the purpose of financial feasibility analysis, cash flow projections need to span cross multiple years, for example, 10–20 years. All the net cash flow in each year needs to be discounted back to present value, added together, and then compare with the initial investment. This discounted sum of net cash flows minus the initial investment is called Net Present Value (NPV) of a project. For example, the following table presents a simple scenario of projected cash flows of a project. The initial investment is $1000, and the net cash flow/net income for the future 6 years is projected. Table 11.5 provides a real example of cash flow analysis of a company.

The formula used to calculate NPV is:

$$NPV = CF_0 + \frac{CF_1}{(1+r)} + \frac{CF_2}{(1+r)^2} + ?+ \frac{CF_t}{(1+r)^t}$$

Where r is the discount rate, which is often measured as the cost of capital rate, sometimes called opportunity cost of capital. Opportunity cost of capital is the incremental return on investment, which needs to be lower than the breakeven rate of return. Breakeven rate of return is also called internal rate of return (IRR). Internal rate of return can be calculated simply based on the projected net cash flows. For example, IRR from the cash flow in Table 11.4 above is calculated as 5%, using the default IRR formula in Excel or a financial calculator. Therefore, 5% is the breakeven rate of return. In this case, the cost of capital, meaning the associated cost of raising the capital, should be lower than 5% in order for the project to be profitable. It makes sense since the smaller the discount rate used in the calculation of NPV, the larger the NPV value will be. It also means that the smaller the costs associated with raising business capital, the more profitable the project will be. Assume the discount rate is 4%. Then the NPV can be calculated as:

$$NPV = ?1000 + \frac{200}{(1+0.04)} + \frac{200}{(1+0.04)^2} + \frac{200}{(1+0.04)^3} + \frac{200}{(1+0.04)^4} + \frac{200}{(1+0.04)^5} + \frac{200}{(1+0.04)^6}$$

Therefore, NPV can be calculated as $48. When the cash flow gets more complicated, computerized software will help the calculation of NPV much easier.

In real estate project analysis, a cash flow statement is often called a pro forma ("as a matter of form" in Latin). Pro Forma first lays out the assumptions, for example, the costs of the project, return rate on investment

Table 11.4 Abridged cash flow for Federal National Mortgage Association (Fannie Mae), a Publicly Traded Government Sponsored Enterprise

Fiscal year is January–December. All values USD Millions.	2020	2019	2018	2017	2016
Operating activities					
Funds from operations	2542	−3124	7702	3745	2272
Funds from operations growth	181.37%	−140.56%	105.66%	64.83%	–
Changes in working capital	−1817	–	–	–	–
Receivables	−2749	–	–	–	–
Other assets/liabilities	932	–	–	–	–
Net operating cash flow	725	−3124	7702	3745	2272
Investing activities					
Sale of fixed assets &; businesses	5991	7425	9321	12,221	16,115
Purchase/Sale of investments	−13,740	1113	−11,735	16,391	14,758
Purchase of investments	−88,281	−20,990	−18,922	−4511	−6070
Sale/maturity of investments	74,541	22,103	7187	20,902	20,828
Increase in loans	766,699	261,808	172,155	189,593	233,935
Decrease in loans	1,139,889	582,437	433,638	468,435	574,206
Other uses	−339,043	−141,475	−108,294	−123,687	−140,147
Other sources	287	–	78	641	116
Net investing cash flow	−46,974	205,422	145,399	188,919	228,048
Financing activities					
Cash dividends paid – total	–	−5601	−9372	−12,015	−9624
Preferred dividends	–	−5601	−9372	−12,015	−9624
Change in capital stock	–	–	3687	–	–
Sale of common &; preferred stock	–	–	3687	–	–
Proceeds from stock options	–	–	3687	–	–
Issuance/reduction of debt, net	100,975	−185,193	−158,316	−182,572	−204,086
Change in long-term debt	100,975	−185,193	−158,316	−182,572	−204,086
Issuance of long-term debt	1,671,462	1,224,807	1,147,201	1,418,535	1,419,664
Reduction in long-term debt	−1,570,487	−1,410,000	−1,305,517	−1,601,107	−1,623,750
Other funds	−510	480	63	6	14
Other uses	−510	–	–	–	–
Other sources	–	480	63	6	14
Net financing cash flow	100,465	−190,314	−163,938	−194,581	−213,696
Net financing cash flow growth	152.79%	−16.09%	15.75%	8.94%	–
Net change in cash	54,216	11,984	−10,837	−1917	16,624
Free cash flow	725	–	7702	3745	2272

Source: Fannie Mae

Table 11.5 An example of goal achievement matrix

Criteria	Weight	Project 1	Project 2	Project 3	Project n
Criterion 1	0.10	4	5	3	5
Criterion 2	0.25	3	1	2	1
Criterion 3	0.50	4	2	4	2
Criterion 4	0.15	5	4	5	3
Weighted average		3.90	2.35	3.55	2.20

(capitalization rate), and other key parameters. Then a projected cash flow over the future years will delineate itemized revenues, expenses, and net cash flows. Debt service coverage ratio measures the ratio of operating income available to pay the debt, which is often the construction loans or mortgages in real estate development.

Ex-post Analysis

When implementing a policy or a project, constant monitoring is often required. After the implementation, it is beneficial to evaluate the effectiveness of the policy or project, and to measure whether the implementation has helped achieve intended goals. Policies and programs administered by the government agencies or receiving public or private grants often require evaluation or reporting of the policy or project outcomes. Many use success or failure to measure project outcomes; however, simply using success or failure overly simplifies the complex nature of many projects and policies. Seven common criteria are often used to measure project success: Technical performance, efficiency of project execution, managerial and organizational implications, personal growth, project termination, technical innovativeness, and manufacturability and business performance (Freeman & Beale, 1992). All these criteria, however, need to be consistent with the discounted cash flow (DCF) principle.

There are various methods evaluating the outcomes of a project or a policy. The following section will explain some of the main methods used in research or practice:

- Cost-oriented approaches
- Before-and-after comparisons
- With-and without comparisons
- Actual-versus-planned performance comparisons
- Experimental (controlled) methods
- Quasi-experimental methods
- Economic impact analysis
- Social impact analysis
- Environmental impact analysis

Cost-Oriented Approaches

The financial feasibility analysis section of this chapter introduces the time value of money, present value of projected future cash flows, and net present value (NPV, discounted net cash flows to present value subtract the initial project costs). After project implementation and operation real cash flows and real expenses will further help evaluate whether the costs of the project, when compared with other parameters (e.g. goals, net cash flows, or revenues), validate the continued operation of the projects. Cost-oriented analysis can also be used prior to the implementation of the project, with projected or estimated costs, revenues, or benefits, while evaluation using cost analysis after implementation uses actually incurred costs, revenues, or benefits. Four types of cost-based evaluation analysis will be explained:

- Goals achievement matrix (GAM)
- Cost effectiveness analysis
- Cost-benefit analysis
- Cost revenue analysis

Goals Achievement Matrix (GAM)

The method of goals achievement matrix is descriptive in nature, and positive goals achievement matrix indicates that all the goals have been achieved per required criteria in the intended timeline. The matrix usually includes rows and columns in a tabular format, including the criteria, scores, and weights if each criterion is measured differently regarding its importance. Table 11.5 is an example of a GAM table:

The weighted average is calculated by multiplying individual scores with the weight, and then summing the multiplication together under each project. For example, in Table 11.5 the weighted average for Project 1 is calculated as:

$$4 ? 0.10 + 3 ? 0.25 + 4 ? 0.50 + 5 ? 0.15 = 3.90$$

Scoring criteria should be listed in a separate table in a rubrics format to make them clear.

GAM provides a simple, yet comprehensive review of the project regarding the goals, since these goals are the ultimate guiding compass for the project or policy. Future operation and implementation should focus on why certain goals are not achieved, and why certain goals are achieved, and the focus should be diverted to either revising the goals or striving to achieve those that are not realized. GAM can be used before project implementation when one project needs to be selected among multiple projects, or can be used after projects are implemented to evaluate project success. GAM process is also very effective to engage participation of stakeholders to determine project priority (Cities Alliance, 2020).

Cost Effective Analysis (CEA)

Different from goals achievement matrices, cost effective analysis (CEA) usually focuses on specific objectives under the goals, and evaluate the project or policy's effectiveness relevant to the objectives. CEA is particularly relevant during circumstances where multiple alternative actions or policies achieved the same objective or the level of benefit. In that case, the alternative with the lowest cost yields the largest benefit. For example, if alternatives A, B, and C can all create 300 affordable housing units for low income people (those making less than 60% Area Median Income (AMI) in a local community, while alternative B has the lowest costs of achieving such objective. Under CEA, alternative B is the most cost effective approach. Therefore, CEA is a method which can be used both before and after a project or policy is implemented.

Cost-Benefit Analysis

The financial feasibility section covers basic concepts of NPV of net cash flows, which is the difference between the discounted future net cash flows (the differences between revenues and expenses) and the present value of the initial costs. Cost benefit analysis, in the financial sense, is the same approach, although actual data will be used for evaluation after project or policy implementation. For example, net cash flow for each year is the difference between the actual revenues and all the incurred expenses. Then these annual net cash flows for maybe 10 years are discounted to today's value and added up to present value of all net cash flows, for example, $4.2 million dollars. The initial investment cost is $3 million dollars. Under this circumstance, $4.2 million is called benefit, while $3 million is called cost. The NPV (net present value) for this project, therefore, is $1.2 million dollars. It is positive and therefore, the project is a success, although other financial ratio analysis may be needed to measure how successful the project is compared to similar projects, or the profit goals of the owners/developers of the project. Benefit to cost ratio is therefore $4.2/$3 = 1.4.

Discounted cash flow (DCF) analysis is the fundamental element of cost-benefit analysis, although costs and benefits may not be related to financial terms. For example, benefits may be socially related, for example, youth enrichment programs or environmental educational initiatives. For these programs or policies that do not have tangible financial benefits, it is usually difficult to use financial figures to justify the benefit of these programs. Under circumstances like these, cost benefit analysis may be descriptive or qualitative, and benefits may be measured using surveys or interviews to capture the societal benefits of these policies or programs.

Cost Revenue Analysis

Cost revenue analysis (CRA) is often used in measuring the effectiveness of local communities in providing services versus tax and other types of revenues that local communities receive from certain sectors or localities. Sectors or localities could be based on land uses, such as single-family, multiple family, agricultural, industrial, office, institutional, etc., or based on locations such as central business districts (CBD), suburbs, or certain neighborhoods or subdivisions. The ratio between revenues and costs can be used to measure which sectors or localities generate the most revenues, holding costs constant. Results from the cost revenue analysis, however, should not be used as the basis of providing services. These results are often used internally to help government agencies to improve service effectiveness by lowering costs, or can be used in other planning decisions about future land uses. Different from cost-benefit analysis, cost revenue analysis usually does not consider the benefits generated in these sectors or localities, and the analysis is often static, compared to the long-term nature of cash flow analysis associated with the cost-benefit approach.

Before-and-After Comparisons

Comparing the outcome measures before a plan or policy was implemented, or simply comparing the trend of a variable over multiple time periods, is a commonly used method to measure the effectiveness of the plan or policy. For example, a researcher is interested in how property values are affected by the large-scale redevelopment policies and projects within a special district, such as a few Community Redevelopment Areas (CRAs). One method is to collect the stratified sample of property values within these redevelopment areas before the special areas were established, and then a few years after the special areas were established, controlling for inflation. The average year of these CRAs being established was in 1995, with the earliest 1990 and the latest 1998. The first step for the researcher to do is to find the GIS boundary files for the CRAs. Then he or she uses the property appraisers' parcel taxation GIS shapefiles in 1994 and 2004 as the base files, using the geospatial processing tool in ArcMap to clip the parcel files based on the CRA boundary. Then the researcher only compares property values for selected single-family homes within these areas in 1994 and in 2004 since property value change in single-family homes often best indicates socioeconomic changes of local neighborhoods. The data is organized using the format presented in Table 11.5.

However, Table 11.6 is the standard data format for t-test analysis in SPSS or SAS. In Excel, the data should be organized as:

Then the researcher can either use Microsoft Excel or SPSS, even SAS, to calculate the t-test statistics. T-test is to test whether two groups' data have significant differences. If the two groups are the same subjects, for example, the same housing units, then a paired t-test is used. Please refer to Chap. 7 Statistical Analysis about detailed procedures to conduct a paired t-test.

Table 11.7 presents the t-statistics after the analysis for the housing value impact example. The planner assumes the housing value distribution for the 2 years have similar standard deviations, where it is assumed that the property value distribution in both years is dispersed or spread out at a similar magnitude. Therefore, in Excel, we use Data>Data Analysis, choose "t-test: Two-Sample Assuming Unequal Variance". Variance is squared standard deviation.

Table 11.6 Before-after comparison data format

Folio ID[a]	Year	Appraised property value (in 2004 constant value)
4,125,863,520,000,025	1994	$91,000
4,125,412,532,101,231	1994	$75,346
…	…	…
4,125,863,520,000,025	2004	$210,073
3,517,784,201,302,001	2004	$150,310
	…	…

[a]Folio ID is a unique identification of real properties assigned to each parcel of land by local government

Table 11.7 Before-after comparison data format

Folio ID*	1994	2004
4,125,863,520,000,025	$91,000	$210,073
4,125,412,532,101,231	$75,346	$150,310
…	…	…

Select the data range for the value variables for the 2 years as presented in Table 11.6. Put 0 as Hypothesized Mean Difference. If the variable names (1994 and 2004) are selected, check "label", if not, leave "label" checkbox unchecked. Keep other parameters and statistics default, and click "OK". Alpha of 0.05 means that the planner has 95% confidence that the property values in 2004 have changed, compared to the 1994 values. Then the t-test results are created. After adjusting the length of decimal places to two or three, Table 11.8 then presents the t-test results.

When looking at the P(T < =t) one-tail and two-tail results, we find 0.000, which is the rounded value to three decimal places, since the original value is very small. This indicates that the mean difference is very significant and falls into the rejection zone shown on the tails of the student's t distribution (similar to normal distribution) curve. The results then indicate that the property values in 2004 are significantly larger than property values in 1994. However, the results need to be interpreted carefully to rule out other factors leading to property value appreciation.

This process of using the t-test to compare differences assumes using a sample data to test the population data. If the planner just wants to compare the average property values before and after the establishment of the special district, simply comparing the average values in each CRA in 1994 and 2004 will allow us to see the change. However, in this case, adding data values, such as calculating the average property values in multiple years before and after the establishment of the

Table 11.8 T-test results when assuming unequal variances

	1994	2004
Mean	104,526.67	170,775.00
Variance	647,626,380.95	1,350,620,666.67
Observations	3564	4789
Hypothesized mean difference	0	
Df	8351	
t stat	−5.865	
P(T < =t) one-tail	0.000	
t critical one-tail	1.703	
P(T < =t) two-tail	0.000	
t critical two-tail	2.052	

district, may make the results more convincing since any longitudinal or dynamic trend is better than one-time period comparisons.

Box 11.1 Using Microsoft Excel for Creating Charts and Conducting Basic Statistical Analysis

Most planning offices or consulting firms do not have sophisticated statistical application packages since many of the advanced statistic models are often reserved for researchers who use large datasets in running various models. Regardless of different sub-disciplines, planning researchers are using more sophisticated models and machine learning techniques, such as Gradient Boosted Decision Tree (GBDT) models which creates state-of-the-art results for many complex and challenging data mining problems, to fit large amounts of data.

However, for a planning staff member, such methods are not highly necessary for day-to-day decision making. If the planner needs to use simple statistical methods, such as scatter-plot or box-plot to chart the data with smaller datasets, or using simple linear regression models to predict or fit data, Microsoft Excel will be sufficient. Although over the years, Excel has been more powerful and is able to accommodate large datasets, it is not the norm to use it to conduct advanced statistical analysis. Instead, Excel is a powerful tool in generating tables, charts, and sometimes forms. Forms is a special type of Excel workbook where each spreadsheet is like an application, where the users can input data into designated fields. Then reports and statistics will be created. Creating forms require certain level of mastery of visual basic programming language.

If the “Data Analysis” tab is not available in the toolbar section it needs to be activated first. To activate the statistics function, open Excel. Go to File>Options. Click Add-Ins from the menu bar on the left side. Once a pop-up window appears, choose “Manage: Excel Add-ins”, click “Go…” check Analysis Toolpack, and click OK. Now click the Data menu and “Data Analysis” should be activated. After clicking “Data Analysis”, common statistical procedures, such as regression, random number generation, histogram, correlation, descriptive statistics, t-test, F-test, etc. can be used to fit the data, when choosing the correct data range for independent or dependent variables.

With-and Without Comparisons

This approach intends to measure the outcome and its impact when the outcome and/or impacts are not easy to measure. In this case, we can choose two similar locations, one with the project implemented, and another does not have a similar project implemented. The outcome and impact of the project for the first location are measured, in comparison with the change of the second neighborhood without a similar project. This approach can also be used in the same location, comparing the status before the project is implemented to the outcome and impact after the project is implemented. For example, planners often use pilot projects as an “experiment” to explore whether such pilot projects have merits to become permanent projects. The outcome and impact of the pilot projects are often quantified to justify whether the pilot projects would become successful if becoming permanent.

Actual-Versus-Planned Performance Comparisons

Actual-versus-planned performance comparisons is an approach similar to the goal achievement matrix approach where planned performance estimates are compared with actual performance outcomes. The comparison, therefore, will help determine whether the actual performance meets or exceeds the initial goals.

Experimental Methods

As indicated in Chap. 4, experimental methods are widely used to measure the effectiveness of a stimulus or instrument, for example, a new medicine or vaccine, between control groups and experimental groups. Experimental design is widely used in natural and medicine science. Strictly speaking, in an experimental design, the intended outcome or effect is measured before the initiation of the experiment (pre-test). Then the experiment facilitator randomly assigns subjects to control and experimental groups, and the experimental groups are administered with the instrument, while the control group subjects use a placebo. The effects of the instrument or the placebo is therefore recorded and compared (post-test). Therefore, for each experimental subject, the status of whether receiving the instrument is treated as an independent variable, and the recorded effects or outcome is the dependent variable. Combining with the pre-test results, the experiment will be able to measure the effects or outcome of an instrument with scientific rigor.

Quasi-Experimental Methods

Quasi-experimental methods are similar to experimental design, but do not require randomly assigning subjects into control and experimental groups. Meanwhile, it is also possible that a control group may be missing. For example, one case evolving study only looks at the changing effect of using one case. Another example may be tracking the social impact of low-income housing in a neighborhood over a certain time period to explore how low-income housing may impact neighborhoods, by controlling for other mediating factors. Multiple case comparisons, for example, comparing waterfront redevelopment in multiple cities, is also an approach in quasi-experimental design.

Economic Impact Analysis

Economic impact analysis is the process of estimating the economic impact of a project. This type of analysis is critical in urban and regional planning when introducing a planning project. There are various methods conducting an economic impact analysis, but the fundamental economic impact of projects on localities lies in jobs and tax revenues. New jobs generated can be estimated using industrial standards, or use a modeling system based on input-output analyses of economic multipliers. In the U.S., the Bureau of Economic Analysis (BEA) created and maintains a Regional Input-Output Modeling System (RIMS II), which provides the multiplier estimate of different industries. Certain software applications, such as IMPLAN, also provides fiscal impact analysis of economic activities.

Social Impact Analysis

Social impact of planning and economic activities is often difficult to quantify. However, with the assistance of statistical analysis, such as simple outcome comparison and regression models, it is possible to measure how different activities relate to the change in socioeconomic statuses. For example, if a project is implemented in a neighborhood, certain measures, such as jobs created and tax revenues are quantifiable outcomes. If a researcher or planner intends to understand the health and psychological impact to the residents, surveys and interviews will help gather data to explore such impacts. Positive social impact is one of the most important criteria for planning project implementation, since protecting the public interests is the guiding principle of planning.

Environmental Impact Analysis

Environmental impact analysis sometimes is necessary, depending on the scale, the location, and the ownership arrangement of the project. Chapter 14 Environmental Analysis provides more detailed procedure of conducting the environmental impact assessment. Generally speaking, environmental impact assessment (EIA) is a systematic analysis of the potential environmental impacts of projects. The analysis also assesses the significance of such impacts and proposes measures to mitigate the negative impacts (Biamah et al., 2013). For projects not requiring submitting an EIA to environmental protection agencies, there is no clear guidelines about how the assessment can be conducted, although referring to a formal process may help with the procedure. When a formal environmental impact assessment is required, the following procedure needs to be followed (ELAW, 2020; IISD, 2020):

1. After a project or action is identified, conduct ***screening*** to see if the environmental and social impact of a proposed development project would be significant to warrant an EIA.
2. If an EIA is required, set up the ***scoping*** process to establish boundaries of the EIA, set up analysis basis for each stage, describe project alternatives, and consult the affected stakeholders.
3. Draft EIA by evaluating the ***environmental impacts*** of proposed projects and its alternatives, and then identify the mitigation strategies if the impact is significant. The draft should include feasibility studies, baseline studies, and impact studies.
4. Prepare plans to address ***mitigation measures*** and other project risks, such as technological failures and natural disasters.
5. Prepare the ***EIA final report***.
6. ***Review and approval***. If the project is rejected a revised or new EIA process should be initiated; if the project is approved, the report may be implemented or if applicable goes to a higher level administrative and judicial review.
7. ***Monitoring***. After the EIA report is approved, the project will be implemented and its environmental impact needs to be monitored during the project cycle.

Public participation is expected to be incorporated into the entire process. Public participation may consist of information meetings, public hearings, or opportunities to provide written comments.

Data Science, Technology, and Evaluation Research

In recent years, big data, which are data with huge volumes, digital, often collected continuously, and often designed to collect data for general purposes - not necessarily just for project or program evaluation, has increasingly become an important research tool in evaluation research (York & Bamberger, 2020). Along with the advancement in digital technology, such as sophisticated data collection using drones and applications, and data collection through social media or e-governance, big data allows program or project evaluators using larger datasets, machine learning, and artificial intelligence to better evaluate or simulate project or program outcomes.

Big Data, Internet of Things, and Information and Communication Technology are three major components of the new information technology (York & Bamberger, 2020). Big data that can be used in evaluation research include program administrative databases, household surveys, satellite images and remote sensing data, social media feeds, phone records, internet searches, electronic transactions, video and audio data, text, and integrated data platforms. Internet of things (IoT) can be sources of big data, and mostly encompasses information on the internet, connected and smart devices, wearable biometric monitoring devices, remote sensing, etc.

In addition to traditional statistical analysis, machine learning, the algorithms looking for patterns of big data, enables project or program evaluators to work on messy, and incomplete data. Machine learning does not need to follow the data assumptions intrinsic to many of the statistic modeling methods. Machine learning techniques are not within the scope of this book; however, using these techniques has become a trend in academic research in urban and regional planning. It is also used in zoning, design scenario modeling, site selection analysis, property management, parking analysis, transportation route optimization, program impact assessment, and many other planning sub-disciplines (Kalisky & Mani, 2019). For example, in a capstone project of Harvard Data Science, the class uses machine learning and GIS to estimate parking capacity for the city of Somerville, MA (Rentsch, 2019). Books related to smart cities and artificial intelligence, urban computing, computing and communication systems, and decision support systems should also provide theories and applications of using machine learning in urban planning and evaluation research.

As with any other data sources, there are a few challenges using big data, such as selection biases, poor quality data, construct validity, issues using and interpreting algorithms, research bias, and confusing correlation and causality.

- Selection biases mostly happen when the sample data are not reflective of the population characteristics. Data collected through social media, such as Facebook, Twitter, and Instagram, often suffers such limitations.
- The data collected may have poor quality as well, with missing information, inaccurate information, or false information.

- Sometimes, big data are collected for administrative, transactional, or tracking purposes, are not conditioned for research purposes, and therefore may suffer from construct validity issues.
- Many of the data vendors collecting and managing big data do not completely disclose proprietary information, making it harder for researchers to interpret research findings.
- Inherited in any research, research bias and confusing correlation and causality may be present when using big data in evaluation research.

Meanwhile, as with issues in traditional public engagement, digital data collection may perpetual inequality by excluding social groups who are not able to have easy access to smart devices and internet applications. Collecting and analyzing big data also requires tremendous amounts of resources where smaller organizations are not able to afford. In such circumstances elitism in evaluation research may be evident. Regardless of these challenges, big data and advanced analytical methods have helped opening a pathway to incorporating sophisticated quantitative computing into social sciences and evaluation research.

References

Biamah, E. K., Kiio, J., & Kogo, B. (2013). Chapter 18 – Environmental impact assessment in Kenya. *Developments in Earth Surface Processes, 16*, 237–264. https://doi.org/10.1016/B978-0-444-59559-1.00018-9

Cao, X., & Wang, D. (2016). Environmental correlates of residential satisfaction: An exploration of mismatched neighborhood characteristics in the Twin Cities. *Landscape and Urban Planning, 150*, 26–35. https://doi.org/10.1016/j.landurbplan.2016.02.007

CitiesAlliance.(2020).*Tool17:GoalAchievementMatrix(GAM)*.AccessedonNov.20from:http://city-development.org/tool-17-goal-achievement-matrix-gam/#1472738150329-85a5d0f0-9b68

Dharmadasa, I., Gunaratne, L., Weerasinghe, A. R., & Deheragoda, K. (2015). *Solar villages for sustainable development and reduction of poverty*. Sheffield Hallam University Research Archive (SHURA). http://shura.shu.ac.uk/9783

ELAW (Environmental Law Alliance Worldwide). (2020). Chapter 2 – Overview of the EIA process. In *Guidebook for evaluating mining project EIAs*. Accessed on Nov. 20 from: https://www.elaw.org/files/mining-eia-guidebook/Chapter2.pdf

Freeman, M., & Beale, P. (1992). Measuring project success. *Project Management Journal, 23*(1), 8–17.

Huo, X., Yu, A. T. W., Darko, A., & Wu, Z. (2019). Critical factors in site planning and design of green buildings: A case of China. *Journal of Cleaner Production, 222*(10), 685–694.

IISD (International Institute for Sustainable Development). (2020). *EIA: 7 steps*. Accessed on Nov. 20 from: https://www.iisd.org/learning/eia/eia-7-steps/

Kalisky, S. & Mani, A. (2019). *How GIS and machine learning work together*. ESRI User Conference, 2019. https://www.esri.com/content/dam/esrisites/en-us/about/events/media/UC-2019/technical-workshops/tw-6165-494.pdf

Kaplan, R. (1985). Nature at the doorstep: Residential satisfaction and the nearby environment. *Journal of Architectural and Planning Research, 2*(2), 115–127.

Patton, C. V., Sawicki, D. S., & Clark, J. J. (2013). *Basic methods of policy analysis and planning* (3rd ed.). Pearson.

Rentsch, A. (2019). *Machine learning for urban planning: Estimating parking capacity.* Harvard Data Science Capstone Project, Fall 2019. Towardsdatascience, December 15, 2019. https://towardsdatascience.com/machine-learning-for-urban-planning-estimating-parking-capacity-15aabd490cf8

Ross, S. A., Westerfield, R. W., & Jordan, B. D. (2013). *Fundamentals of corporate finance* (10th ed.). McGraw-Hill.

University of Wisconsin-Madison (UWM). (2020). *Downtown and Business District market analysis.* UWM extension market analysis toolbox. Retrieved on October 22, 2020 from: https://fyi.extension.wisc.edu/downtown-market-analysis/

Wholefoodsmarket.com. (2020). *Real estate*. Retrieved on October 21, 2020 from: https://www.wholefoodsmarket.com/company-info/real-estate

York, P., & Bamberger, M. (2020). *Measuring results and impact in the age of big data: The nexus of evaluation, analytics, and digital technology*. The Rockefeller Foundation.

Web Resources

Urban Land Institute.: https://uli.org/

Congress for New Urbanism.: https://www.cnu.org/

U.S. Department of Housing and Urban Development.: https://www.hud.gov/

Chapter 12
Urban Land Use Analysis

A ***land use*** describes an arrangement, modification, and/or activity that is carried out on land to provide a service to individuals and/or communities. Residential areas not only provide accommodation spaces for shelter, leisure, and child-rearing, but also serve as financial investments of households. Commercial areas serve households' demand for retail, financial, and personal services, and, at the same time, serve as employment centers for workers. A community may keep an open area untouched to protect the environment and to serve residents' needs of leisure and social interaction spaces. Land use links different specialized components of planning, such as housing, commercial, open space, and transportation. The orderly layout of a city requires a proper understanding of land uses from multiple perspectives to avoid clashing land uses that interfere with public health, economic efficiency, and environmental preservation. This chapter introduces concepts and methods for land use analyses, including land use classification systems, land use inventory and compatibility analyses, various measures for identifying land use patterns, and land use demand analysis.

Land Use Classification

The concept of land use in cities goes beyond physical activities and involves psychological, social, and economic dimensions. Land cover presents physical aspects of a land use. It focuses on physical materials as shown on a piece of land, i.e., what can be seen on a land. There are five major schemes for classifying land uses or land covers: the Anderson Land Classification Scheme, the Cadastral Land Use Classification, the traditional urban land use classification, the Land based Classification Standards, and the urban transects.

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_12].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_12

A land can be covered by natural or planted vegetation, bare rock, or man-made constructions. The ***Anderson Land Classification Scheme*** reflects these physical differences, i.e., land covers. It was developed by the U.S. Geological Survey (USGS) to provide a standardized scheme for interpreting remote sensor images and developing land use and land cover data across the nation and over time (Anderson et al., 1976). The original scheme as proposed by Anderson et al. includes four hierarchies of land use and land cover categories. Every hierarchy targets at a different type of remote sensor data. For instance, classification level 1 activities are designed for LANDSAT type of data, i.e., satellite images. The first level has nine land cover types, including urban or built-up land, agriculture land, rangeland, forest land, water, wetland, barren land, tundra, and perennial snow or ice. The second level includes 37 subcategories, some of which describe land use activities, for instance, residential, commercial, industrial, and transportation. The third and fourth categories required medium- and low-altitude images, which are often taken from aerial vehicles, and are less useful for the USGS.

With the Anderson Land Classification Scheme, the USGS has been producing national land coverage databases (NLCD) every 2–3 years since 2001. 2001–2016 data can be found on the Multi-Resolution Land Characteristics (MRLC) Consortium website (https://www.mrlc.gov/). Table 12.1 presents the current land cover categories used for the 2016 NLCD. Level 1 classes remain consistent with the original

Table 12.1 The 2016 NLCD land cover legends

Level I classes	Level II classes	
Water	11	Open water
	12	Perennial ice/snow
Developed	21	Low intensity residential
	22	High intensity residential
	23	Commercial/industrial/transportation
Barren	31	Bare rock/sand/clay
	32	Quarries/strip mines/gravel pits
	33	Transitional
Forested upland	41	Deciduous Forest
	42	Evergreen Forest
	43	Mixed Forest
Shrubland	51	Shrubland
Non-natural woody	61	Orchards/vineyards/other
Herbaceous upland natural/semi-natural vegetation	71	Grasslands/herbaceous
Herbaceous planted/cultivated	81	Pasture/Hay
	82	Row crops
	83	Small grains
	84	Fallow
	85	Urban/recreational grasses
Wetlands	91	Woody wetlands
	92	Emergent herbaceous wetlands

scheme. Level II classes have 21 different land cover types, which is much less than the original scheme.

In urbanized areas, administrators and assessors are more interested in economic attributes of the land, such as land use type and intensity, which directly determine the value and taxable status. The ***Cadastral Land Use Classification system*** addresses such interests. This system has been used by local property assessors when constructing a land recording system that facilitates property management and transfer activities (IAAO, 2016). Table 12.2 presents the land use classification used by the Property Valuation Administrator's (PVA) office of Jefferson County, KY, in 2019. Most local administrations adopt similar classification systems. General land use categories, including agriculture, industrial, commercial, residential, government, utilities, etc., not only address differences in how a land is used, but also build a structure based on which the local government can impose different property taxes. A local property classification system is dynamic, with changing land use activities over time to fit the needs of new policies and new planning visions.

Planning practices in the twentieth century have primarily focused on urban areas. There are a set of ***traditional land-use categories in planning*** that have been used. These land uses include residential, commercial, mixed use, institutional uses, open space, transportation, public and government lands, and environmental areas (Jeer & Bain, 1997). When preparing a land use map, the frequently used colors are yellow for residential, red for commercial, purple or hatched pattern for mixed-use, blue for institutional use, green for open space, black or gray for transportation, pink for public and government use, and green or blue for environmental areas.

Modern planners view land as multi-dimensional concept and introduce the ***Land Based Classification Standards*** (LBCS). The LBCS provides a more comprehensive description of a land. Such a land use database is flexible in serving different planning purposes. It interprets the same set of land uses from five different dimensions: activity, function, structure type, site development character, and ownership. Activity describes the land use based on what can be observed. Function reflects the economic attribute of a land use. Structure type explains structural or building characteristics. Site development character provides information about a land in relation to a development. Ownership refers to legal rights of a land. Every dimension consists of a hierarchy of categories. Table 12.3 lists general land use categories for each dimension. For every general land use category, there is a hierarchy of land uses nested within. Compared with other land use classification systems, this LBCS allows for more flexibility and enables planners or administrators to precisely target at a specific aspect of properties for a specialized planning or public policy analysis (APA, 2001).

More recently, with the rising attention on improving the integrity between an urban and the natural environment, the ***urban transect*** approach provides another land use classification scheme. It provides a land use classification strategy for planners who are motivated for planning for a harmonious relation between urban and rural areas, or, between urban and the natural environments. Different from the previous land use classifications, which focus on describing individual land parcel,

Table 12.2 Jefferson County, KY, PVA land use code

Property class	Description	Property class	Description
100	Agricultural vacant land	500	Residential vacant land
101	Agricultural with dwelling	510	Residential 1 family dwelling
199	Agricultural with outbuildings	520	Residential 2 family dwelling duplex
296	Telecom companies	530	Residential 3 family dwelling triplex
300	Industrial vacant land	540	Mixed use residential and comm
305	Manufacturing	542	Sec. 42 tax incentive
320	Industrial warehouse	546	Mobile home without land
330	Industrial office	548	Mobile home with land
400	Com vacant land	549	Residential patio/condos
401	Com tri-Plex apartment	554	Residential condo 1 thru 50 units
405	Apartments	555	Residential condo land and amenities
410	Commercial motels/hotels	556	Residential condo 51 or more units
412	Commercial hospitals/nursing homes	561	Residential co-ops
415	Commercial Mobile home parks	580	Moratorium
420	Commercial retail	598	Boat slip/condo
430	Commercial restaurant	599	Outbuildings
442	Commercial sect. 42 spec assess	601	Exempt Federal Gov't
443	Commercial medical clinic/ offices	602	Exempt State Gov't
445	Commercial office	604	Exempt Local Gov't
450	Commercial condos	606	Exempt religious
451	Commercial condo land and amenities	608	Exempt education
455	Commercial auto services	609	Exempt other
460	Commercial parking facilities	620	Exempt metro government
465	Commercial entertainment	646	Mobile home Pk lot unoccupied
470	Commercial non-exempt schools	650	Exempt right of way
480	Commercial warehouse	655	Exempt open space
496	Commercial cell tower	656	Res condo master lot
497	Commercial moratorium	657	Com condo master lot
499	Commercial paving, fencing, yard item	660	Exempt parking
		690	Exempt telecommunications
		699	GIS account/match
		700	REMF voids (Don't use)
		820	Utility industrial
		830	Utility commercial

Table 12.3 General LBCS land use categories by dimension

Activity	Function	Structure	Site	Ownership
1000 Residential activities 2000 Shopping, business, or trade activities 3000 Industrial, manufacturing, and waste-related activities 4000 Social, institutional, or infrastructure-related activities 5000 Travel or movement activities 6000 Mass assembly of people 7000 Leisure activities 8000 Natural resources-related activities 9000 No human activity or unclassifiable activity	1000 Residence or accommodation functions 2000 General sales or services 3000 Manufacturing and wholesale trade 4000 Transportation, communication, information, and utilities 5000 Arts, entertainment, and recreation 6000 Education, public admin., health care, and other inst. 7000 Construction-related businesses 8000 Mining and extraction establishments 9000 Agriculture, forestry, fishing and hunting	1000 Residential buildings 2000 Commercial buildings and other specialized structures 3000 Public assembly structures 4000 Institutional or community facilities 5000 Transportation-related facilities 6000 Utility and other nonbuilding structures 7000 Specialized military structures 8000 Sheds, farm buildings, or agricultural facilities 9000 No structure	1000 Site in natural state 2000 Developing site 3000 Developed site -- crops, grazing, forestry, etc. 4000 Developed site -- no buildings and no structures 5000 Developed site -- nonbuilding structures 6000 Developed site -- with buildings 7000 Developed site -- with parks 8000 Not applicable to this dimension 9000 Unclassifiable site development character	1000 No constraints--private ownership 2000 Some constraints--easements or other use restrictions 3000 Limited restrictions--leased and other tenancy restrictions 4000 Public restrictions--local, state, and federal ownership 5000 Other public use restrictions--regional, special districts, etc 6000 Nonprofit ownership restrictions 7000 Joint ownership character--public entities 8000 Joint ownership character--public, private, nonprofit, etc. 9000 Not applicable to this dimension

urban transects emphasize on characteristics of neighborhoods. Figure 12.1 illustrates the basic idea of urban transects. It describes a regional system of urban elements gradually transitioning into rural elements from the urban center to the edge. There are seven transect zones: special district, urban core zone, urban center zone, general urban zone, sub-urban zone, rural zone, and natural zone (Pinnell, 2009). Different zones have different architecture, landscaping, building type and style, and street layout standards (Talen, 2002).

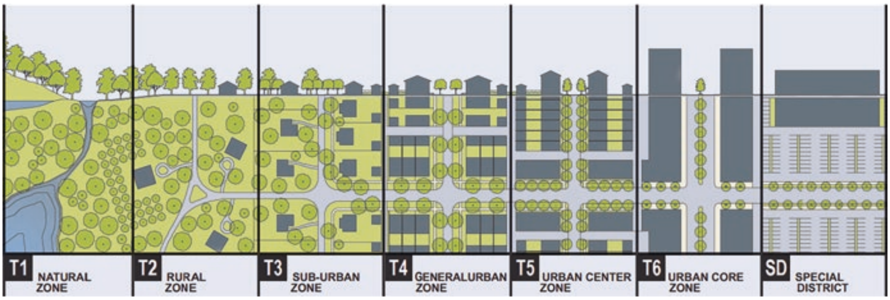


Fig. 12.1 Urban Transect Zones. (Source: Pinnell (2009). This Figure was originally created by Duany Plater-Zyberk & Company and became free of licensing in 2005)

Land Use Inventory Analysis

Many cities maintain active land use inventory systems, for instance, a property value assessor's land record database. It is a parcel-based information system. It provides information about existing land uses, land use regulation, property ownership, property value, and past transactions for all land parcels. A land use inventory analysis is different from a land use inventory system. A land use inventory analysis has analytical purposes of understanding the current level of land supply and explaining how land has been used in the city.

For preparing a ***land use inventory analysis***, one needs to first determine what type of land use classification will be used. The four land use classification systems introduced in the previous section serve different purposes. For an economic planner, the Cadastral style of land use classification system can serve as an appropriate scheme because it focuses on functional and economic differences of land uses. However, an environmental planner may use the Anderson Land Classification Scheme because it captures more difference among natural attributes.

The next step is to collect current land use data. One may need to require such data from multiple sources, including the planning agency, the property assessor's office, and other related entities. For areas without existing land use data, field surveys become necessary.

The third step is to re-classify land use activities into a desired land use classification scheme. For instance, the data from a local property assessor's office has 62 different land uses if it uses the classification scheme as presented in Table 12.2. One needs to regroup the data into more meaningful categories, such as agriculture, manufacturing, residential, and commercial etc.

With land use data ready, one can present the data in a land use inventory table, which summarizes usage by land use categories. Table 12.4 provides an example of Jefferson County, KY. With this information, a planner can identify (1) what types of land uses exist in the county? (2) what land uses dominate? (3) whether there is a supply of land to accommodate future developments? A land use inventory

Table 12.4 2018 Land use inventory of Jefferson County, KY

	Area (Acre)	% of Total
Commercial	15,021	6.04%
Farmland	30,353	12.20%
Industry	18,281	7.35%
Multi-family	8007	3.22%
Parks and open space	26,423	10.62%
Public and semi-public	17,701	7.12%
Right-of-way	31,444	12.64%
Single family	82,614	33.21%
Vacant	18,895	7.60%
	248,738	

analysis often comes with a map of existing land uses, which provides a thematic mapping analysis about the locations and distributions of land uses.

Land Use Incompatibility

Land use incompatibility explores adverse effects of different types of land uses imposing on each other. For instance, introducing high density development into an environmentally sensitive area could impair the quality of the natural environment. Placing a large employment center near a residential area may bring in more traffic and noise and interfere with residents' daily life. A neighborhood being placed too close to a highway may face the problems of excessive noise and emission, and this has a harmful effect on people's health.

With proper planning, adjacent land uses can be compatible with each other. All urban land uses serve varying needs of people and are unavoidable urban elements. Large factories could bring various issues related to traffic, noise, and pollution. However, they provide employment opportunities as well. One land use such as open space, parks can serve as the buffer for other land uses to minimize adverse impacts. A land use buffer is a common strategy to deal with conflicting land uses. It is to put a physical barrier between conflicting land uses. A tree belts or an additional setback can serve as a buffer for noise, pollution, odor, or unpleasant visuals (Akbari, 2002; Mok et al., 2006).

Many land uses can be supportive for each other. This provides the foundation for mixed-use developments in the U.S. To promote a neighborhood with functionally integrated land-uses, the real estate industry has extensively explored the relations among land uses (Brett, 2019). Table 12.5 analyzes potential relations among land uses. Residential and small-sized retail land uses complement each other, with residential activities providing a retail demand and retail activities improving the

Table 12.5 Market grid for mixed-use development (MXD)

Primary use components	Health Care	Marina	Entertainment: Sports	Entertainment: Theaters	Entertainment: Bars & Restaurant	Retail: Comparison	Retail: Specialty Stores	Retail: Convenience	Hotel	Offices	Residential
Residential	□	●	x	x	□	□	□	●	x	●	
Offices	□	–	□	–	●	□	□	□	●		●
Hotel	□	□	●	□	●	□	●	□		●	x
Retail: convenience	–	●	□	□	□	□	`x		□	□	●
Retail: specialty stores	–	●	□	●	●	●		x	●	□	□
Retail: comparison	–	□	□	●	●		●	□	□	□	□
Entertainment: bars & Restaurant	–	●	●	●		●	●	□	●	●	□
Entertainment: theaters	–	–	□		●	●	●	□	□	–	x
Entertainment: sports	□	□		□	●	□	□	□	●	□	x
Marina	–		□	–	●	□	●	●	□	–	●
Health care		–	□	–	–	□	–	–	□	□	□
Level of market synergy in MXD											

Source: Brett (2003). Reproduced by permission from Urban Land Institute and the author
● Strong, □ Weak or uncertain, – Neutral absence of synergy, x Potential market conflict

quality of life of residents and making the neighborhood more attractive. Co-location of different retail and entertainment activities creates an economic advantage of a shared investment and an increased customer attraction. This provides mutual benefits to all commercial activities.

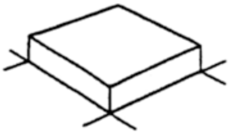
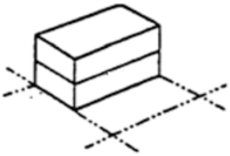
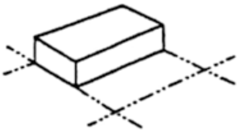
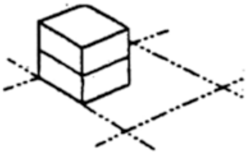
Land Use Indicators

Galster et al. (2001) propose eight dimensions for studying a sprawled built environment (density, continuity, concentration, clustering, centrality, nuclearity, mixed use, and proximity) and discuss how to measure them in a GIS environment. Ewing and Cervero (2010) propose a set of D variables to measures the nature of a city in relation to travel, including density, diversity, design, destination accessibility, and distance to transit. The design dimension captures many elements that Galseter et al. propose, such as continuity, concentration, and clustering. This section follows the Ewing-Cervero system and presents operationalized measures of different dimensions of a built environment.

Density

Density measures the level of activities for every unit of land area. It can be generalized measurements such as the size of a population or employment per acre. In the planning area, there are many measures specifically focusing on physical development. ***Gross density*** measures the total number of building units per unit of area, which includes supportive spaces such as roadways, parks, and public and infrastructure spaces. There are another two measures focusing on site specific development intensity: lot coverage and floor area ratio. ***Lot coverage*** is the ratio of the size of a building structure's footprint to its ground area. It measures the proportion of a land that is covered by a building structure, which includes non-livable areas such as garage, swimming pool, sheds, etc. ***Floor area ratio*** (FAR) is the ratio of the gross floor area of a building to its ground area. Gross floor area is the total area of a building measured from between middle of outside walls. Table 12.6 presents four examples to further illustrate the concepts of lot coverage and floor area ratio. A FAR explores the level of activities a site can accommodate, while a lot coverage focuses on the level of modification of a development on a site. For example, both case 1 and 2 can accommodate the same level of activities. However, case 2 has a lower lot coverage percentage and therefore a lower level of environmental impact.

Table 12.6 Examples of lot coverage and floor area ratio (FAR)

	Building lot coverage	FAR
Case 1 	1	1
Case 2 	0.5	1
Case 3 	0.5	0.5
Case 4 	0.25	0.5

Diversity

Land use diversity accounts for the number of different land uses within an area. Number of land uses, which is available from a land use inventory analysis, serves as one simple measure of diversity. A neighborhood with more numbers of different land uses is more diverse. The present of non-residential land uses is another indicator for diversity. In U.S. cities, most land is dominated by residential activities. A higher percentage of non-residential land uses is expected to have a higher level of land use diversity.

The Shannon index and the Simpson index are another two commonly used diversity indicators. When applied for a land use analysis, the ***Shannon index*** (H) of a community can be defined as:

$$H = -\sum_{I}^{i=1} \frac{a_i}{A} * \ln\left(\frac{a_i}{A}\right)$$

and the ***Simpson Index*** (S) is:

$$S = 1 - \sum_{I}^{i=1} \left(\frac{a_i}{A} \right)^2$$

where A is the total area of the community, a_i is the area of type-*i* land use, and *I* is the number of land use categories in the area. The calculations should not include vacant land. For both indices, a higher index value indicates a higher level of land use diversity. The Shannon diversity index has been extensively used in the existing literature to measure land use mix (Frank et al., 2006; Brown et al., 2009).Both the Simpson index and the Shannon index measure the evenness, i.e., whether there is the same amount of land uses in a community or a city, with a more even distribution having a higher index value. These indices do not depict local variations of land use diversity, i.e., whether a census tract has more diverse land uses than another.

Design

Two types of neighborhoods dominant U.S. cities: traditional and suburban. They differ in the design of street networks, development scale, and walkability. Popular street network measures include ***length of street per square mile***, ***number of intersections per square mile***, and the ***line-to-node ratio***, which is the number of street segments divided by the number of intersections in an area. A traditional neighborhood uses a grid street network. As a comparison, cul-de-sac style of streets are more common in suburban areas. A lower level of street density and line-to-node ratio is expected to be associated with suburban land use development. As for development scale, common measure are average building setbacks, average block size, and average street width. At last, walkability is another design attribute. ***Walkability indicators*** include numbers of pedestrian crossings, street trees, and availability of sidewalks and so on.

To measure the degree of sprawl, the ***Sprawl index*** of a region (*SI*) can be calculated as:

$$SI = \left(\frac{S\% - D\%}{100} + 1 \right) * 50$$

where *D%* is the percentage of the total population in high-density census tracts and *S%* is the percentage of total population in low-density census tracts. In a study of U.S. metropolitan areas, Laidley (2016) defines high-density tracts as tracts with 3500+ persons per square mile, low-density tracts as tracts with 200–3500 persons, and rural tracts as tracts with less than 200 persons. The value of this sprawl index varies between 0 and 100, with a higher value indicating a higher level of urban sprawl.

Destination Accessibility

Accessibility refers to the convenience of a location to be reached. Many factors influence the level of accessibility of a location, including demographic attributes, mobility, roadways, and the distribution of land use activities (Handy et al., 2002; Levine et al., 2019). From the land use aspect, accessibility can be assessed by the distance to the closest desired destination. For instance, the retail accessibility of a residential neighborhood can be measured as the distance to the closest retail store, with a shorter distance indicating a higher retail accessibility. Accessibility can also be measured as number of activities within a reachable area. For instance, the job accessibility of a location can be measured as number of jobs within 10 minutes driving distance.

Land Use Suitability Analysis

Land use suitability analysis is to identify the most suitable sites for a specific land use. This chapter focuses on the economic dimension of land and introduces the land use suitability as a tool to achieve economic functionality and efficiency. Suitability analysis can be a tool for environmental analysis. We will re-introduce suitability analysis in Chap. 12 for dealing with environmental considerations.

Land use suitability analysis is necessary in cities because of the heterogeneous nature of urban land and land use, as well as the dynamic relation among them. Every parcel in a city is different in locational convenience, economic importance, and legal and environmental constraints. On the contrary, a specific land use demands for a particular set of site attributes. For instance, for a commercial land use, the accessibility to its potential customers is the most important factor. For an industrial land use, it is critical to have a large site available for laying out its assembling lines. A land use suitability analysis brings together factors related to land, land use, and other related socio-economic factors.

The process of a land use suitability analysis includes (1) deciding on criteria; (2) collecting data; (3) constructing suitability indices; and (4) deciding which land is more suitable. When constructing a set of criteria, or decision factors, one should consider both site attributes and land use characteristics, such that one can assess the capacity of land to accommodate a land use activity.

Site attributes can be analyzed from three scales, regional, neighborhood, and the site itself. A site synthesizes multiple attributes from different dimensions and scales into one bundle of goods to be offered. At the regional scale, the overall urban structure determines the functionality of every land in a city. There is a functional separation of urban spaces in a region. For many cities, the central business area (CBD) is the most important job market. Suburban areas are dominated by single family residential uses but may have modern manufacturing and commercial activities integrated. These regional scale factors constitute the first set of factors for a suitability analysis. A neighborhood contains not only the site, but also socio-economic relations that interact with the subject site. When focusing on neighborhood attributes, one should identify dominate land uses and study the relations between the subject land use and existing land uses. This is the key step for avoiding incompatible land uses. The property site itself is the finest geographical scale of analysis. Physical attributes of the land, the improvements, the current conditions, and legal ownership all affect suitability level of this land to accommodate a land use activity.

Different land uses have different expectations on land and therefore, different suitability criteria apply. Table 12.7 provides an illustration. Modern industrial land use activities have a unique production cycle, which highly depends on the availability of land to layout factories, an abundance of utilities to support the production, and a convenient access to air freight, railway, and highway to ship in material inputs and ship out products. In many cities, there are taxation policies specifically designated to support or regulate industrial activities. Therefore, potential taxation cost is one of the factors for analyzing land use suitability for industrial uses. For commercial activities, there are different sets of criteria for a land use suitability

Table 12.7 Suitability analysis criteria for different land uses

		Residential	Commercial	Industrial
Regional factors	Proximity to air freight			X
	Proximity to railway			X
	Proximity to interstate highway		X	X
	Direction of growth	X	X	
	Accessibility to jobs	X		
	Accessibility to people		X	X
	Air and water quality and compliance cost			X
Neighborhood factor	Adjacent land uses	X	X	X
	Neighborhood prestige	X		
	Land cost			X
	Taxation cost			X
	Natural amenity	X		
	Crime	X	X	X
	Accessibility to shopping centers	X		
	School quality	X		
Site factors	Landscaping		X	
	Parking		X	X
	Accessibility by truck or train			X
	Parcel size	X	X	X
	Frontage and visibility		X	
	Topography		X	X
	Utilities	X		X
	Zoning		X	X

analysis. For regional factors, whether a commercial land use is in line with the growth direction (with a higher level of growth potential) of a region affects the profitability of this commercial activity. It is also important to have a location close to potential customers, i.e., the population, while other types of accessibility are less important. For residential land uses, many neighborhood factors are important, such as neighboring land uses, neighborhood prestige, school quality, the availability of natural amenities, the convenient accessibility to shopping centers, and neighborhood safety.

A suitability analysis involves a ***multi-criteria decision making***. The first step is to analyze the context information and explore the purpose of the suitability analysis and nature of the subject land use. Then one can proceed with selecting criteria that are related to the subject land use activity. The third step is to assign a performance score for every criterion. Table 12.8 illustrates an example of a land use suitability analysis for introducing electrical vehicle (EV) charging stations. With criteria selected based on an analysis of the operation of EV charging stations, one needs to assign a performance score. Every candidate location is evaluated for its performance for every criterion. A performance score can take a value between 1 and 10, with 1 indicating the least suitable and 10 indicating the most suitable for all criteria. The fourth step is to assign a weight to every criterion. When adding up all the weights, it should equal one. When assigning a weight to a criterion, one should consider the importance of the subject criterion in suitability analysis and the reliability of the data of this specific criterion.

The last step is to calculate the overall ***suitability score*** for every candidate location based on the weights and the performance scores, as

$$SS_l = \sum_{I}^{i=1} w_i * a_{il}$$

where SS_l is the suitability score of location l, w_i is the weight for the ith criterion, a_{il} is the performance score for the ith criterion at location l, and I is the number of criteria. Based on this suitability score, one can make the decision about which locations are better choice. In the case of the EV charging station land use suitability analysis (Table 12.8), location L18 has the highest suitability score (5.3), followed by location L4 with a suitability score of 5.2.

Future Land Development Projection

Chapters 9 and 10 have introduced methods for projecting future population and employment levels. With increasing demographic and economic activities, there will be an increasing demand for land for building homes, stores, offices, and factories. The section addresses the questions of how much land will be needed to accommodate the future growth and how to translate population and employment projections into land use expansion. The methods presented in this section are

Table 12.8 A land use suitability analysis for electric vehicle charging stations

Criteria	Weight	Candidate locations																	
		L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18
Total population level	0.07	1	1	1	2	2	1	1	1	1	1	1	1	1	2	1	6	1	1
Total employment level	0.08	5	5	5	5	5	5	5	4	4	5	5	5	4	5	4	5	5	5
# of nearby apartments and condos	0.09	1	1	1	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1
EV ownership	0.07	9	10	10	9	9	9	10	9	10	10	10	10	10	9	9	9	9	9
Distance to the closest existing level 2 station	0.07	1	2	1	1	2	1	2	2	2	2	2	1	2	1	3	2	3	3
Distance to the closest existing fast charging station	0.12	3	3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	2	2
Traffic	0.11	3	4	6	5	3	3	4	5	5	4	5	5	4	2	3	4	5	5
# of votes for an EV Charging Station	0.07	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Distance to the closest highway exit	0.1	10	10	10	10	9	10	10	10	9	10	9	9	10	10	10	9	8	8
Total gross leasable areas	0.11	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	10
Existing number of parking	0.11	5	5	5	5	5	5	3	1	4	4	5	5	5	2	1	1	4	5
The overall score		4.3	4.6	4.7	**5.2**	4.4	4.3	4.3	4.1	4.4	4.5	4.6	4.5	4.4	4.2	3.8	4.3	4.2	**5.3**

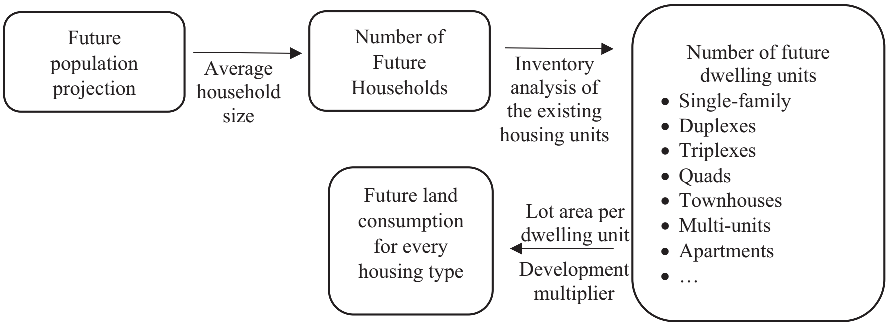


Fig. 12.2 Projecting Future Residential Land Consumption

developed based on the experience of the city of Norman in Oklahoma (City of Norman, 2004).

Figure 12.2 presents a method for projecting future residential land uses based on population projection. First, with the future population projected and with average household size information available, one can project the number of households in the future. It is reasonable to assume that one household demands for one housing unit. Therefore, the total number of housing units in the future can be projected.

Second, the composition of a city's housing market tends to be stable over time. An inventory analysis of the existing housing units can reveal the housing unit distribution by dwelling type. Applying this distribution, one can project numbers of future housing units by type.

At last, a residential neighborhood consists of land used for residential parcels and land used for circulation and supportive services. To project the amount of land used for residential lots, one can search the zoning (which regulates development size and intensity for different type of land uses) for lot area per dwelling unit information. If such information is not available in the local land use regulation, one can study the existing parcel information, which is available with a county's property auditor's office, and calculate the lot area per dwelling unit information for different type of housing units. This lot area per dwelling unit information enables the conversion of the number of future dwelling units to the area of residential parcels/lots.

To project the amount of land for supportive services, one needs to identify the land development multiplier for every type of housing units. This multiplier describes the proportion of land that is used for non-dwelling purposes, such as streets, club houses, and parking. The city of Norman, Oklahoma, complied a table of residential development parameters (Table 12.9), which include land development multipliers for different dwelling units, based on historical building information and data collected from Planning and Community Development Department. With this land development multiplier information available, one can project the amount of land for non-dwelling uses. The total future residential use can be projected accordingly.

Figure 12.3 presents the method for projecting future economic land uses based on the projection of employment by industry.

Table 12.9 Residential development parameters – City of Norman, Oklahoma

	Average lot area per dwelling	Land development multiplier
Country residential	435,600	5%
Very low density residential	87,120	10%
Low density residential	10,500	25%
Duplex	4600	25%
Triplex/quadruplex	5445	20%
Townhouse	5445	15%
High density residential	2420	10%

Source: City of Norman (2004)

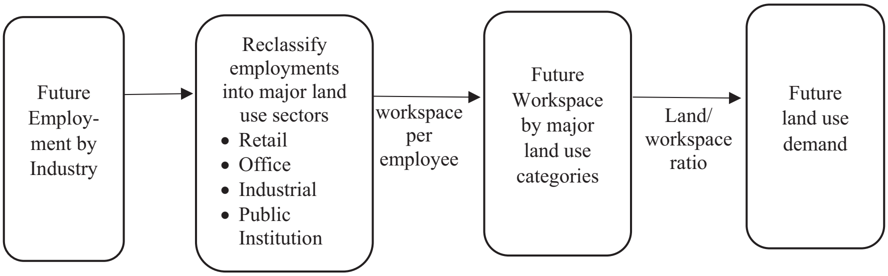


Fig. 12.3 Projecting Future Economic Land Consumption

Table 12.10 Non-farm employment by land use type

	Retail	Office	Industrial	Public/Institution
Mining	0.00%	25.84%	0.00%	0.00%
Construction	0.00%	18.84%	18.84%	0.00%
Manufacturing	0.00%	1.05%	98.95%	0.00%
Transport, Communication & Public Utilities	5.53%	25.93%	68.54%	0.00%
Wholesale trade	0.00%	14.67%	85.33%	0.00%
Retail trade	92.91%	0.91%	6.18%	0.00%
Finance, Insurance & Real Estate	21.25%	78.75%	0.00%	0.00%
Services	34.99%	49.80%	7.02%	8.18%
Government employment	0.00%	3.80%	0.00%	96.20%

Source: City of Norman (2004)

The major challenge is to convert economic activities into another classification scheme that is consistent with land use activities. For instance, for a manufacturing industry, its land uses may include three major economic land uses: industrial, office, and retail (e.g., showrooms). Public institution is the fourth popular land use with job opportunities. Therefore, the first step of the analysis is to convert industry-based employment to land-use based employment. To do so, one must study every industry and identify their workers' working space type. Table 12.10 provides an example. Based on this information, industry-based employments can be converted to land-use based employments.

With employment data converted into land-use based, one can proceed with an analysis of the demand of workspaces based on workspace per employee information. Many local and federal agencies study space usage per employee for different types of structures. For instance, the Federal Energy Information Administration (FEIA) regularly conducts building energy consumption surveys. The survey provides useful information for calculating workspace per employee for different types of buildings.[1] In the real-estate industry, space-per-employee is also well studied. Fanning (2014) recommends 235 square feet per employee for office jobs.

With the demand of different types of workspace information available, one can continue to link workspace with land consumption based on land/workspace ratios, which can be derived from floor area ratios (FAR) as regulated by zoning. Total land demand can be projected accordingly.

Analytical Programs and Models

A land use decision is influenced by many factors, for instance, the market, the lifestyle, people's preference of amenity, and the local policy. It also can influence many other decisions directly or indirectly. The introduction of a new infrastructure can stimulate new growth in the area. Adding a business plaza can make a neighborhood residentially more attractive. Land use analytical programs and models attempt to uncover such interactions and can facilitate a decision-making process about future facility placements and infrastructure additions and affect future urban structure.

The Lowry-type land use model is one of the early efforts (Goldner, 1971). It is based on the economic-base theory and a gravity model to simulate the distribution of residential and employment activities across a region. With the basic employment decision made and the plant built, residential locations will be simulated based on the gravity model that considers distance as a deterrent factor. These residential neighborhoods generate demand for services and service employment locations will be similarly simulated using a gravity model. These new service employments attract more residential neighborhoods to grow, and the growth of new residential neighborhoods stimulates additional service activities. These iterative dynamics continue until an equilibrium is reached.

Empirical analyses represent another stream of land use modeling analysis. It uses cross-sectional data and econometric methods to explore the determinants of land use changes. For instance, Nelson and Hellerstein (1997) used geophysical and socioeconomic variables to explain different types of land use activities.

[1] https://www.eia.gov/consumption/commercial/data/2012/bc/cfm/pba2.php.

Optimization models focus on the best decision according to one or more criteria. They minimize one or more objectives subject to a selected set of constraints. For example, one can study a land use pattern that minimizes the total development cost, total commuting time, total cost for public facilities, and/or environmental impacts. Constraints can include available land and income, costs of development, and pollution generated based on the current development intensity, energy consumptions, urban runoff, and so on.

More recently, planning researchers and practitioners are interested in addressing what-if questions. What are the impacts if a development is approved? What will be the growth pattern if a project is implemented? A scenario planning offers a tool to planners to explore possible ways that the future may unfold and adopt development strategies accordingly. A scenario planning starts with a comprehensive analysis of past and current activities, as well as socio-economic, political, and environmental factors that can influence future growth. Then it produces a set of possible scenarios, i.e., possible futures, and the associated policies, development strategies, and initiatives for each scenario. It also studies the likelihood of each scenario to occur. Such an analysis provides useful inputs to advice decision-makers to adopt appropriate tactics for a more sustainable growth in the future.

References

Akbari, H. (2002). Shade trees reduce building energy use and CO2 emissions from power plants. *Environmental Pollution, 116*, S119–S126.

APA (American Planning Association). (2001). *Land Based Classification Standards*. https://planning-org-uploaded-media.s3.amazonaws.com/document/LBCS.pdf.

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A land use and land cover classification system for use with remote sensor data* (Vol. 964). US Government Printing Office.

Brett, D. L. (2003). *Real estate market analysis: Trends, methods, and information sources*. Urban Land Institute.

Brett, D. L. (2019). *Real estate market analysis: Trends, methods, and information sources* (3rd Ed.). Urban LandInstitute.

Brown, B. B., Yamada, I., Smith, K. R., Zick, C. D., Kowaleski-Jones, L., & Fan, J. X. (2009). Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity. *Health & Place, 15*(4), 1130–1141.

City of Norman, Oklahoma. (2004). *Norman 2025 land demand analysis*. Local government report.

Ewing, R., & Cervero, R. (2010). Travel and the build environment: A meta-analysis. *Journal of the American Planning Association, 76*(3), 265–294.

Fanning, S. F. (2014). *Market analysis for real estate: Concepts and applications in valuation and highest and best use* (2nd ed.). Appraisal Institute.

Frank, L. D., Sallis, J. F., Conway, T. L., Chapman, J. E., Saelens, B. E., & Bachman, W. (2006). Many pathways from land use to health: Associations between neighborhood walkability and active transportation, body mass index, and air quality. *Journal of the American Planning Association, 72*(1), 75–87.

Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., & Freihage, J. (2001). Wrestling sprawl to the ground: Defining and measuring an elusive concept. *Housing Policy Debate, 12*(4), 681–717.

Goldner, W. (1971). The Lowry model heritage. *Journal of the American Institute of Planners, 37*(2), 100–110.

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine, 23*(2), 64–73.

IAAO (International Association of Assessing Officers). (2016). *Standard on manual cadastral maps and parcel identifiers*. https://www.iaao.org/media/standards/Manual_Cadastral_Maps_2016.pdf.

Jeer, S., & Bain, B. (1997). *Traditional color coding for land uses*. American Planning Association.

Laidley, T. (2016). Measuring sprawl: A new index, recent trends, and future research. *Urban Affairs Review, 52*(1), 66–97.

Levine, J., Grengs, J., & Merlin, L. A. (2019). *From mobility to accessibility: Transforming urban transportation and land-use planning*. Cornell University Press.

Mok, J. H., Landphair, H. C., & Naderi, J. R. (2006). Landscape improvement impacts on roadside safety in Texas. *Landscape and Urban Planning, 78*(3), 263–274.

Nelson, G. C., & Hellerstein, D. (1997). Do roads cause deforestation? Using satellite images in econometric analysis of land use. *American Journal of Agricultural Economics, 79*(1), 80–88.

Pinnell, P. (2009). *Smart code Version 9.2 – Center for urban rural interface*. The Town Paper Publisher.

Talen, E. (2002). Help for urban planning: The transect strategy. *Journal of Urban Design, 7*(3), 293–312.

Chapter 13
Transportation Planning Analysis

Transportation is a key component of the society since the beginning of human civilization. Through the nineteenth century, walking, using animals and animal-related vehicles, and steam-trains were the major transportation tools. The invention of electricity, automobiles, air crafts, and modern trains greatly transformed human life. In conjunction with roads paved with concrete and asphalt, highway systems became a reality. Transportation is one of the urban and regional planning fields that witnessed the most technological advancement over time. Artificial intelligence and remote sensing has continued to be incorporated into transportation engineering, making automobiles, airplanes, ships, trains, and other transportation vehicles safer, faster, and more fuel efficient.

However, the key issues of transportation remains. First of all, how to transport human and products most efficiently with the lowest cost is the fundamental question that transportation planning is interested in. ***Transportation planning*** is the planning process involving multiple disciplines, such as urban and regional planning and civil engineering, and multiple government agencies at the national, state, and local levels to coordinate the improvement of the existing transportation system based on demographic characteristics, travel patterns, and the needs of communities (ITE, 2010). Transportation planning practice focuses on coordinating land use, housing, and transportation needs of local communities through interaction between planning, design, and operation of transportation services, balancing consumption of renewable and non-renewable energy and promoting the use of alternative mode of transportation (walk, bike, or public transit) (ITE, 2016). The fundamental goal of transportation planning is improving mobility, accessibility, and safety, which will be explained in more details in later sections. Secondly, increasing transportation safety is especially important in implementing transportation policies. Lastly, maintaining or improving the current transportation infrastructure is one of the critical functions of transportation planning agencies. As with other fields in planning,

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_13].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_13

transportation planning is very multi-disciplinary, and is closely related to civil engineering, land use, and makers of transportation vehicles. Land use planning highly relates to the efficiency of the transportation systems. Sprawling urban areas and mismatched housing supply and employment centers tend to increase commute time, congestion, and traffic accidents.

Resolving issues in transportation mostly lies in the hands of local agencies, such as the Metropolitan Planning Organizations (MPOs). Along with states and tribes, MPO is required to coordinate transportation improvement programs through Capital Improvement Plans (CIPs) and long range plans. Various management programs are also associated with intermodal facilities (facilities as connectors of various modes of transportation), pavement, bridges, public transportation, transportation safety, and traffic congestion. With the increasing recognition of the collaborative efforts between land use, housing, and transportation agencies, it is imperative to establish an efficient and effective transportation network when considering the coordination among these three important functions in cities. Transportation analysis heavily relies on Big Data, mathematical and statistical calculation and manipulations, machine learning, and various sophisticated software (e.g. TransCAD, ArcGIS, etc.) and hardware systems (GPS, remote sensor, smart card systems, satellite, etc.). Therefore, advanced modeling in transportation analysis requires more advanced knowledge in transportation data analysis and statistics, and mastering the methods of comprehensive analysis requires time in course-lengths and possibly multiple courses. This chapter can only offer a glimpse of the complex procedure of data and analysis in transportation planning. Readers interested in learning further may benefit from taking courses in transportation planning and civil engineering, and refer to various data, toolkits, and manuals published by the U.S. Census Bureau (Census Transportation Planning Products), Transportation Research Boards, Department of Transportation, American Public Transportation Association, and the Institute of Transportation Engineers. Freight and aviation logistics related to transportation planning requires a separate set of data and methods and therefore is not specifically covered in this chapter.

This chapter will explore the fundamental focus of transportation, starting with a few important concepts such as mobility, accessibility, and level of services. The chapter then introduces basic methods in Transportation Demand Management (TDM) modeling system, followed by travel behavior and safety analysis.

Key Concepts in Transportation Planning

Before getting into modeling methodology and optimization analysis of transportation networks, it is critical to understand some of the key concepts in transportation planning, and how land use and road networks work together to move people and goods effectively and efficiently.

Transportation and Mobility

Mobility and ***accessibility*** are the two fundamental functions of a transportation network. Transportation moves people and goods from one location to another through different modes of transportation such as automobile vehicles, bicycles, rail, train, airplanes, and ferries/boat, etc., while mobility is more than simply moving people and goods, but the ability to move them freely in an affordable and safe manner.

Mobility is associated with the following aspects:

- Accessible and inclusive. Transportation systems need to be easily accessible to all, regardless of socioeconomic statuses, individual mobility levels, and locations. The system must be easy to navigate with minimal obstacles.
- Affordable and economically viable. The transportation system must be affordable to all, regardless of modes of transportation.
- Efficient and smart. Transportation systems should be efficient, sometimes incorporating advanced technologies, so that it takes shorter time with minimal barriers to travel.
- Safe. Transportation systems should be safe with minimal traffic accidents.
- Sustainable and environmentally friendly. Contemporary transportation systems should incorporate the principles of sustainability in design and implementation. Use of renewable energy and clean energy, and mass public transit are the keys to transportation sustainability.

Functional Classification of Streets and Roads

Roads and streets are often classified based on the capacity of mobility and accessibility levels. Classification also provides design standards of the roadways system. Only a limited percentage of the total street miles and vehicle miles of travel (VMT) is allowed in different types of roads in urban and rural areas.

U.S. Federal Highway Administration defines the classification criteria and procedures through the following guiding principles (FHWA, 2020). Tables 13.1 and 13.2 list examples of the classification system:

Interstate. Interstates are the highest classification of arterials and were designed and constructed with mobility and long-distance travel in mind.

Other Freeways & Expressways. Roadways in this functional classification category look very similar to Interstates. …for the purpose of functional classification the roads in this classification have directional travel lanes, usually separated by some type of physical barrier, and their access and egress points are limited to on- and off-ramp locations or a very limited number of at-grade intersections.

Table 13.1 VMT and mileage guidelines by functional classifications – arterials

	Arterials			
	Interstate	Other freeways & expressway	Other principal arterial	Minor arterial
Typical characteristics				
Lane width	12 feet	11–12 feet	11–12 feet	10–12 feet
Inside shoulder width	4–12 feet	0–6 feet	0 feet	0 feet
Outside shoulder width	10–12 feet	8–12 feet	8–12 feet	4–8 feet
AADT[1] (Rural)	12,000 – 34,000	4000 – 18,500[2]	2000 – 8500[2]	1500 – 6000
AADT[1] (Urban)	35,000 – 129,000	13,000 – 55,000[2]	7000 – 27,000[2]	3000 – 14,000
Divided/ Undivided	Divided	Undivided/Divided	Undivided/Divided	Undivided
Access	Fully controlled	Partially/fully controlled	Partially/ Uncontrolled	Uncontrolled

[1]*AADT* Annual Average Daily Traffic
Source: U.S. Department of Transportation Federal Highway Administration

Table 13.2 VMT and mileage guidelines by functional classifications – collectors and locals

	Collectors		
	Major collector[2]	Minor collector[2]	Local
Typical characteristics			
Lane width	10–12 feet	10–11 feet	8–10 feet
Inside shoulder width	0 feet	0 feet	0 feet
Outside shoulder width	1–6 feet	1–4 feet	0–2 feet
AADT[1] (Rural)	300–2600	150–1110	15–400
AADT[1] (Urban)	1100 – 6300[2]		80–700
Divided/Undivided	Undivided	Undivided	Undivided
Access	Uncontrolled	Uncontrolled	Uncontrolled

[1]*AADT* Annual Average Daily Traffic
Source: U.S. Department of Transportation Federal Highway Administration

Other Principal Arterials. These roadways serve major centers of metropolitan areas, provide a high degree of mobility and can also provide mobility through rural areas.

Minor Arterials. Minor arterials provide service for trips of moderate length, serve geographic areas that are smaller than their higher arterial counterparts and offer connectivity to the higher arterial system.

Major and Minor Collectors. Collectors serve a critical role in the roadway network by gathering traffic from local roads and funneling them to the arterial network.

Local Roads. Locally classified roads account for the largest percentage of all roadways in terms of mileage. They are not intended for use in long distance travel, except at the origin or destination end of the trip, due to their provision of direct access to abutting land.

Level of Service

Level of service of roads and streets are determined by the ability of the transportation network to accommodate traffic flow, free flow or delayed flow. Traffic congestion/interruption/restrictions, freedom to travel with desired speed, driver comfort and convenience, excessive land access points, operating costs, poor signal timing, and other factors determine the level of services (Mathew & Rao, 2007). The Highway Capacity Manual published by Transportation Research Board (TRB) explains the level of services as summarized in Table 13.3 (TRB, 2010):

Transit levels of services are also available regarding rail and transit vehicles. See Highway Capacity Manual for the full description and definition of these levels of services. Standards for transit services require not only traffic speed and maneuverability, but also capacity, peak loading, operating conditions, and mobility factors.

There are various ways to improve levels of services, for example, allowing left turn lanes, allowing right turn in red, allowing adequate space for traffic weaving, expanding transit service, expanding road capacity, and improving Transportation Demand Management (TDM) systems. Access control is often used to improve mobility and accessibility of roads. Access control strategies include spacing of access points, prohibited access (such as one-way street), prohibited left turns, etc.

Transportation Demand Management (TDM) Modeling

The major purpose of ***transportation demand management*** (TDM) is to improve mobility and accessibility of a transportation system. There are a myriad of strategies used in promoting transportation demand management, therefore guiding trip generation, distribution, mode choice, and/or route selection over time and across space (Ferguson, 1990). For example, express lanes on highways, high occupancy

Table 13.3 Level of service criteria for signalized and unsignalized intersections

Level of service	Average control delay (seconds/vehicle)		General description
	Signalized intersections	Unsignalized intersections	
A	≤ 10	0–10	Free flow
B	>10–20	>10–15	Stable flow
C	>20–35	>15–25	Stable flow
D	>35–55	>25–35	Approaching unstable flow
E	>55–80	>35–50	Unstable flow
F	>80	>50	Forced flow

Source: Highway Capacity Manual 2010 (Transportation Research Board, 2010) vehicle lanes on highways, staged working schedule, remote working arrangement, and incentives for commuters to use alternative transportation modes, such as walking, biking, and public transit. TDM may influence individuals' transportation decisions and promote sustainability in transportation. Data for transportation planning, monitoring, control, and information are mostly from the state and federal agencies, such as the Department of Transportation, and local agencies, such as Metropolitan Planning Organization (MPO) and transit authorities. Data collection methods rely on satellite images, remote sensing, automation, surveillance systems, field data, and sometimes surveys, focus groups, public forums, or interviews. With advancement in technology, much of the transportation data can be collected through automated systems, such as smart cards (transit cards, highway passes, etc.), GPS vehicle locations, cellular data, mobility tracking apps, parking tracking apps, and other methods to collect data. These data help estimate and predict the state of the transportation systems, travel demand, travel behavior, and help aid in transportation demand management.

Trip Generation, Distribution, Mode Choice, and Route Selection

Various factors determine the mode of transportation for individuals and households. Purpose of the trip, distance, route traveled, shortest time spent on trips, safety, income, and availability of alternative modes are common factors determining travel demand such as trip generation and mode of choice (Salter, 1976). Road network data and TAZ (Transportation Analysis Zone) demographic data and employment data are the two largest segments of data required in order to carry out a travel demand modeling analysis.

After collecting all relevant transportation analysis data, travel demand analysis usually involves a four-step process.

1. **Trip generation**. Trip generation determines the number of trips to be made by the travelers based on transportation analysis zones (TAZ). Traffic Analysis Zone (TAZ) is a geographic unit commonly used in the CTPP (Census Transportation Planning Products) to estimate travel demand models for transportation planning purposes, even though many regional travel demand models use different TAZs and the CTPP TAZs are different from standard census geographies. Various factors, such as number and size of households, vehicle ownership rate, land use, and development density, determine the trips generated (Federal Highway Administration, 2018). Trips are classified as external to external trips, external to internal trips, and internal to internal trips, in reference to the TAZs.

   The Trip Generation Manual of the Institute of Transportation Engineers (ITE) and the Quick Response Urban Travel Estimation Techniques and Transferable Parameters ("The Quick Response") are the two guiding sources for key parameters in trip generation. For example, the Quick Response Urban

Table 13.4 Average vehicle trips per unit

Residential land uses	Vehicle Trips per Unit
Single Family Residential	9.1 to 10.2
Duplexes and Townhouses	7.0
Apartments	6.0
Mobile Homes	5.5
Retirement Homes	3.5
Condominiums	5.9
Planned Unit Developments (PUDs)	7.9

Source: Quick Response Urban Travel Estimation Techniques and Transferable Parameters, 1987

Travel Estimation Techniques and Transferable Parameters found that each type of residential land uses generates different numbers of trips (Table 13.4).

Based on these average trips, local communities are able to estimate the total trips generated in a Traffic Analysis Zone (TAZ) or any other geographic units.

The "The Quick Response" manual also provides average trips based on different functional land uses, such as retail, office, manufacturing, restaurants, banks, parks, and educational facilities, to name a few. The Trip Generation Manual (ITE, 2017) provides comprehensive trip generation data for multiple modes of transportation in various places and it is a golden standard when analyzing trip generations.

In combination with data provided in the CTPP, planners are able to estimate trip distribution, mode choice, and route selection of the generated trips. All these will help planning the best transportation networks to move people and freight efficiently and effectively.

2. **Trip distribution**. This step determines the destination of the trips generated in each TAZ. Each TAZ is weighted based on an attractiveness factor for the attractions it has and its distance from other TAZs. Trip matrixes delineating the relationship between trip generations and attractions and gravity models help model trip distribution.
3. **Mode Choice**. The step determines the mode used by travels from origins to destinations. Modes include automobiles, transit, carpooling, walking, cycling, ferry, and others. Existing or forecasted transit and carpooling capacities are considered in this step. Logit and nested logit models are the typical models used in estimating mode choice.
4. **Route Assignment**. Based on the assumption that travelers will choose the route with the least travel time (shortest path) or least costs (minimum generalized cost), this step assigns the routes of the travelers from origins to destinations.

After trip assignment analysis, the models need to be validated and calibrated before being used to development a transportation plan. Travel demand analysis is usually performed every 10 years to project for the typical life span (for example, 20 years) of the transportation facility. Travel demand models are also location specific, may be countywide, multi-county, sub-regional, regional, statewise, or across different states.

Travel demand analysis is a traditional technique requiring sophisticated models and large amounts of data, and individual behavior and activity scheduling factors are not readily incorporated into the models. This translates to using various transportation planning software packages, such as TransCAD, to estimate the models. Nevertheless, travel demand modeling provides baseline estimation of travel behavior, and therefore plays significant roles in transportation planning analysis. Other models, such as transportation demand management, congestion studies, detour planning, public transit design and expansion, are examples of using transportation data in design and planning analysis.

Parking Analysis

Parking demand analysis is an important element in travel demand management. Parking needs vary based on different land uses, the demographic characteristics of the residents, available alternative modes of transportation, availability of parking spaces, cost of parking, accessibility, peak and periodic use factors, and employee, visitor, and patron parkers (ULI, 1993). For example, older residents may need to have more disability spaces compared to other types of residents. Commercial establishments may require more parking than industrial land uses. Parking Generation Manual, consummate with the Trip Generation Manual, provides standards and data about parking generations. Based on peak hour parking demand, Table 13.5 delineates required parking spaces for different types of land uses. Box 13.1 uses machine learning and GIS to estimate the parking needs of a city.

Table 13.5 Selected examples of parking generator

Parking Generator	Average Rate	Range
Affordable Housing	0.99/DU	0.32–1.66/DU
Hotel	0.74/ROOMS	0.43–1.47/ROOMS
Health/Fitness Club	1 per 211 SF	1 per 95–500 SF
General Office (general urban/ suburban)	1 per 418 SF	1 per 180–2000 SF
General Office (Dense multi-use Urban)	1 per 613 SF	1 per 430–1030 SF)

Source: Parking Generation Manual (5th Ed.) (Institute of Transportation Engineers, 2010)

Box 13.1 Using Machine Learning to Estimate Parking Capacity

Traditionally transportation analysis has incorporated big data and sophisticated computing software (e.g. TransCAD, Census Transportation Planning Package, ArcGIS Network Analyst, and Remix, etc.) in estimating travel demand and making transportation plans. With the advancement in computing technologies and artificial intelligence, machine learning has become a popular methodology in urban planning sub-disciplines, including transportation planning. For example, a Harvard Data Science Capstone Project uses machine learning to estimate parking capacity for the City of Somerville, MA (Rentsch et al., 2019). The following summarizes the process of conducting the study.

The City of Somerville was planning for a light rail line and needs to find out whether the current on-street and off-street parking will meet the demand from the new transit line without affecting the needs of the residents. The City was also curious where to place subway stations. The project intends to provide the city with an accurate estimate of parking capacity probabilities at the parcel level, and address the uncertainties associated with these probabilities. The project started with sampling a subset of residential parcels and labeled those with driveways and garages (both labeled as having driveways) by using the satellite images provided by the City. This sample data will be used as a comparison basis for the projected data for all the parcels.

To predict whether a residential unit has a driveway, the project uses the building and parcel footprint data, and street parking permits as the data sources for a learning model. Features such as distance from building to the lot line of the parcels, distance to the neighbors, etc. are extracted and calculated to feed into the learning model. The project team then used a variety of classifiers, such as logistic regression (with and without interactions), random forest, AdaBoost, XGBoot, and feed forward neural network. Please note that all these techniques are beyond the scope of this book. Please refer to literature in machine learning and the applications in urban planning for further reference. The random forest approach was then proven more superior to capture nonlinearities in the relationships between feature classes and the labels (binary variable of whether a parcel has a driveway or not). The project team then used a Platt scaling technique to calibrate the model results to reflect the true probability of each parcel having a driveway. The project team then estimated the aggregated driveways and mapped the block-level predictions. The results will help the City of Somerville to have an estimated parking capacity and identify the best locations for the subway stations.

Fundamentals of Traffic Data Analysis

Traffic data analysis involves around analyzing the key parameters of traffic to aid in transportation planning. There is a large amount of traffic indicators and this book will only cover the important ones deemed by the authors. For a full scale of these indicators please refer to the "Traffic Data Computation Method Pocket Guide" ("Guide") published by Federal Highway Administration (Federal Highway Administration, 2018). Traffic Analysis Toolbox, also formulated by the Federal Highway Administration, provides comprehensive tools in traffic analysis (Box 13.2). The following parameters are mostly based on the Guide and the Traffic Analysis Toolbox.

Annual Average Daily Traffic (AADT) and Vehicle Miles Traveled (VMT)

Annual average daily traffic is the mean total volume of traffic across all days for a year at a given location along a roadway. The average is calculated based on the Federal Highway Administration standards and AADT is one of the most frequently measured traffic statistics in local settings. It measures how busy roadways are and is one of the most important parameters used in transportation planning and funding allocations. When multiplying the length of the road segment, daily vehicle miles traveled (DVMT) can be computed. When multiplying AADT with the section length (in miles) and number of days in a year, Vehicle Miles Travelled (VMT) is calculated. Therefore, VMT is

$$Vehicle\,Miles\,Travelled = AADT * Section\,Length\left(miles\right) * number\,of\,days\,in\,a\,year$$

K-Factor

K-factor is the proportion of AADT occurring in the peak hour. For example, K-30 is the 30th highest hourly volume of the year as a percentage of the AADT.

$$K - factor = \frac{K^{th}\,highest\,volume}{AADT}$$

K-factor can be used in capacity planning, functional classification of roads, and sometimes analysis of traffic operations such as lane closures. In order to calculate K-factor, AADT and 1 year hourly volume data are needed.

Capacity and LOS

Capacity is the "maximum sustainable hourly flow rate at which vehicles reasonably can be expected to traverse a point or a lane on a roadway during a given time period" (Federal Highway Administration, 2018, p.24). Capacity is vehicles per hour per lane. Since capacity measures the maximum vehicle flows when the traffic demand meets the capacity, critical density and speed is achieved. The *Highway Capacity Manual* by Federal Highway Administration provides default capacity values on basic freeways and multilane highway segments.

Critical Speed, Maximum Flow Rate, and Critical Density

Critical speed is the traffic speed when the roadway meets its capacity. Critical speed is measured in Miles per Hour (mph).

$$Critical\ Speed\left(mph\right)=\left(\frac{Maximum\ Flow\ Rate\left(\frac{Veh}{H}\right)}{Critical\ Density\left(\frac{Veh}{Mile}\right)}\right)$$

For uninterrupted traffic flow, the default value for critical density is 45 passenger car/mile/lane. If the maximum flow rate is 2350, then the critical speed will be 50 mph. Flow rate is the traffic volume on a facility (vehicle\per hour). The task for the traffic manager, is then to make sure the traffic speed at or above the critical speed to avoid breakdown in traffic flow.

Number of Signalized Intersections

A count of at-grade intersections, where intersections at the same elevation or grade, with full-scale or partial signal devices, is an important parameter calculating capacity and estimate delay for a roadway segment. The distance between signals also play important roles in designing the signal cycle timing.

Box 13.2 Federal Highway Administration: Traffic Analysis Toolbox

The **Traffic Analysis Toolbox** Programs was formulated by the Federal Highway Administration and is a series of software and methodologies used in traffic and transportation analyses. The analysis is highly technical and therefore interested readers should refer to the published documentation of the toolbox through Federal Highway Administration's website: https://ops.fhwa.dot.gov/trafficanalysistools/. Fundamentally there are five methodologies in the Toolbox (FHWA, 2012a):

- **Sketch Planning**. Sketch planning analysis tools are often used to analyze the impact of regional growth on transportation infrastructure and therefore is often suitable in regional transportation planning analysis. Sketch planning tools are mostly based on simple queuing techniques or volume-to-capacity relationships from the Highway Capacity Manual (HCM).
- **Travel Demand Modeling**. Travel demand models are mathematic models in the prediction of travel demand, when considering destination choices, mode choices, time-of-day travel choices, and route choices.
- **HCM/Analytical Models**. These procedures are able to analyze facility performance (capacity, density, speed, delay, and queuing) as a function of demand, geometry, and traffic controls.
- **Simulation Models**. The macroscopic models simulate traffic flows as an aggregated quantity, based on the relationships among the flow, speed, and density of the traffic. Microscopic models are based on individual vehicles. Mesoscopic simulation models combine the properties of both microscopic and macroscopic simulation models.
- **Traffic Signal Optimization Tools**. These tools are primarily designed to optimize signal phasing and time plans for signalized intersections, arterial streets, or signal networks.

Regarding Measures of Effectiveness (MOEs), the following measurements are often used (FHWA, 2007):

- Travel Time, which is the mean travel time of the traced vehicle travel times for a road segment.
- Speed, which is the mean speed for a road segment, estimated by dividing the segment length by the mean travel time.
- Delay, is the difference between the measured travel time and the free-flow travel time for a road segment.
- Queue. The best way to measure a queue is to count the second by second arrivals at an upstream point and the second by second departures at a downstream point. The difference between the cumulative arrivals and the cumulative departures at time "t" is the queue at time "t".

- Stops, can be measured using the travel time measure if along a road segment, using the intersection delay method if at an intersection.
- Density. The average density over time at a given point in space is computed by dividing the measured flow rate per hour by the measured arithmetic mean speed at the point.
- Travel-Time Variance, is computed as the sum of the squared differences between the observation and the mean value.

All of these measurements have their limitations; however, they do provide the best possible assessment of the effectiveness of the transportation infrastructure.

Transit Data Analysis

Transit systems are public transportation systems encompassing busses, trains, lightrail, ferries, trolleys, and other forms of systems transporting passengers. The purpose of transit data analysis focuses on analyzing the state of the system and its usage patterns by passengers. With the advancement of technology, transit data can be often collected through Automated Data Collection Systems (ADCS) such as automated fare collection (AFC) and automated vehicle location (AVL) (Koutsopoulos et al., 2019). Ridership, customer classification, and congestion pattern analysis are key procedures estimating the usage of public transit systems. Transit scheduling, monitoring, and management mostly focuses on demand management to better balance demand and supply of the system.

Ridership and Boarding Estimation

Ridership estimation relies on data in automated fare collection (AFC) systems and revenues when AFC data is missing. Therefore, *direct count*, *revenue-based estimation,* and *sampling methods* are the three fundamental methods estimating ridership data. Direct count data can be obtained from AFC systems or fareboxes, and sometimes transit operators such as bus drivers if there is no AFC in place. However, count data registered manually by transit operators tend to create systematic errors in estimation (Furth, 2000). For revenue based estimation, if fares are differentiated by distance or if transit passes are for unlimited uses, it may be difficult to match revenues with passenger trips, in which case additional data collection (e.g. surveys) may be needed. The revenue estimation method include *farebox revenue* and *transit pass sales*. Based on revenues, ridership and boarding can be estimated. Additionally, transit agencies often conduct surveys to a sample of riders to estimate ridership.

Random sampling, a *concentrate sample*, or the *100 percent ride check* are the most commonly used sampling methods in collecting transit ridership and boarding data.

Passenger Miles

Passenger miles measure trip length of passengers. When boarding data is known through automated fare collectors (AFC) such as fareboxes estimating passenger trip length is straightforward. However, when boarding count data is not known, surveys and sampling methods can be used to estimate boarding, and then passenger miles. Route stratification and cluster or concentrated sampling are the major techniques to estimate boarding and passenger mile. However, route stratification can be only applied when boardings are known by routes, which may conflict with boarding data collected by fareboxes. Therefore, route-level ridership and boarding estimated from cluster sampling and the 100% ride check are most accurate when being used to estimate passenger miles.

Load Profiles

Load profiles analyze passenger activity, such as *boardings* and *alightings*, and passenger load at each stop along a single route, in a single direction (Furth, 2000). Boardings and alightings are the processes that passengers get aboard and then situated in a transit system, or get off the transit system. Load data can be departing or arriving load, however, departing load is most commonly used in transit analysis. Data sources of load profile often come from manual ride checks or automated fare collectors. Systematic transit planning often requires aggregated load profile data instead of single route, or single-day data. However, single-route summaries can often provide more detailed state of the transit system. *Peak load* data are also critical in planning for peak-hour capacity of the transit system.

Schedule Adherence Analysis

Scheduling is critical to the success of public transit systems. Scheduling relies on ridership data, passenger miles, boarding and aligning time, and operator factors. ***Schedule adherence analysis*** explores how the schedule of public transit is followed in the reality. Delays and unexpected accidents will often deviate the official schedule of a route. Therefore, schedule adherence analysis provides quality analysis of the public transit system and help improve and monitor future scheduling and trip time.

Other Types of Transit Analysis

Route economic performance analysis, geographic analysis, historical analysis, and passenger behavior and perception analysis are some of the other types of analysis used commonly in transit planning. Route economic performance analysis measures the cost and benefit of a set route. Geographic analysis uses GIS software and data to measure how ridership, boardings, alightings, passenger miles, schedules, performances, historical analysis, and other data vary geographically. Historical analysis measures the same sets of transit analyses in historical contexts, revealing longitudinal performance of the transit system.

Cycling, Walking, and Other Alternative Mode Analysis

Data collection methods are similar across different modes of transportation. Demographic data, ridership and usage data, traffic volumes, and statistics apply in cycling, walking, and other alternative mode analysis. When data are not readily available, primary data methods, such as 100 percent ride check or sampling, should be used to collect data. Data analysis, therefore, will aid in planning for a multimodal transportation system.

Travel Behavior and Safety Analysis

Travel behavior analysis relies on individuals' travel data to project patterns and aggregated characteristics. In the U.S., travel data can be obtained from the CTPP (Census Transportation Planning Products) and primary data collection (surveys and interviews). Demographic data are usually from the American Community Survey. Various travel behavior links to different exposure risks to traffic accidents.

Traffic crashes are the leading cause of death for U.S. adolescents (CDC, 2010) and remain one of the top actual causes of death in the United States, following tobacco use, poor diet, physical inactivity, alcohol consumption, microbial agents, and toxic agents (Botchwey et al., 2014; Mokdad et al., 2004, 2005). Various factors have been found to be associated with traffic crashes, including the built environment, transportation design, travel behavior, vehicle safety, and demographic characteristics, which were classified as "human infrastructure" versus "engineering infrastructure" (Evans, 1990). Consistent with previous findings, Evans (1990) found that human behavioral factors, especially social norms, plays the most important role in determining traffic safety.

Modeling traffic accidents and their moderating factors often relies on sophisticated models, such as Zero Inflated Negative Binomial Regression Model. Traffic crash data are non-negative, sporadic, and discrete, and therefore stochastic modeling is more suitable than the traditional deterministic models (Sharma & Landge, 2013). In deterministic models, a set of parameters or initial conditions always yield the same output, while in stochastic modeling, the same set of parameters and initial conditions might generate different sets of output. Traffic crash data are point data, and therefore should use models fitting point data. Poisson regression models are the typical models fitting point data; however, these models are not able to handle the over dispersion issues related to traffic crashes, since many study areas may have zero crashes, while some other areas have large amounts of crashes. In this case, a zero inflated negative regression model will overcome the over-dispersion problem, and therefore, is used to measure traffic safety, when controlling for a myriad of independent variables, such as demographic variables, built environment variables, and transportation regulation variables. Zero inflated regression models are two regime models. First, a probability model determines whether a count number is zero or positive number, and this model is called inflated model. Then the positive part of the distribution is described by a suitable stochastic distribution, called base model. The negative binomial regression model can be written in the following form (NCSS, 2020):

$$P_r\left(Y = y_i \mid \mu_{i,}\alpha\right) = \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(y_i + 1\right)\Gamma\left(\alpha^{-1}\right)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}$$

where

$$\mu_i = t_i\mu$$

$$\alpha = \frac{1}{v}$$

The parameter μ is the mean incidence rate of y per unit of exposure. It is the risk of a new occurrence of the event during a specific exposure period, t. Exposure may be time, space, distance, area, volume, or population size. t_i is the exposure period for a particular observation. When no exposure is given, it is assumed to be one.

The regression coefficients are estimated using the method of maximum likelihood and most statistical software, such as R, STATA, and SAS, can easily help achieve such estimation.

However, one limitation of a zero inflated negative binomial regression model is the lack of its ability to truly model the moderating factors leading to variances in crashes. A modified model, named zero-inflated negative binomial structural equation model, explores the inter-linkages of the independent variables, hoping to probe the moderating factors leading to traffic crashes. This model has been rarely used in the field of urban and regional planning, but has been used in business and economics (Jeong & Shin, 2017). Figure 13.1 is a structural equation model indicating the determining factors of traffic crashes, when being incorporated into a zero-inflated negative binomial regression model.

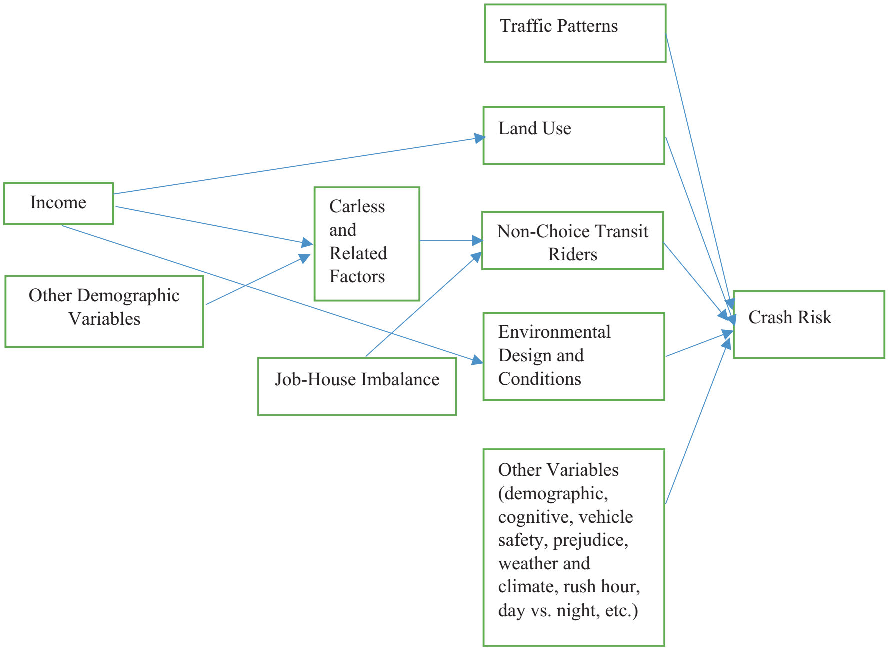


Fig. 13.1 Zero inflated negative binomial structural equation model

Capital Improvement Planning

Capital Improvement Plan (CIP), or capital improvement program, is a short-range transportation plan, usually 4–6 years, to identify capital projects in transportation and their associated budget and funding sources. CIP often also prioritizes projects based on horizons of implementation timeline. CIP is a working document and should be updated annually to reflect the implementation activities. A CIP usually should include:

- Estimated overall cost of each project
- Estimated operational/maintenance cost of each project
- Estimated timeline for the construction and completion of the project
- Project justifications
- Revenues from each project (if any)
- Funding sources and a financing plan
- Ranking of each project based on priority
- Performance indicators and project development milestones

A CIP is usually developed through the following four steps (Opengov.com, 2019):

Step 1. *Organizing the Capital Improvement Plan*, including plans for a CIP and designating a lead department or personnel. This process also needs to specify criteria to prioritize projects, develop a CIP budget calendar, and complete public participation process.

Step 2. *Identify projects and funding options*. Projects are ranked based on urgency, funding needs, and other criteria.

Step 3. *Prepare and recommend a capital plan and budget*. The selected projects, plans, timelines, ranking, and financing options are summarized in a CIP and presented for approval by the elected officials. The CIP should also include performance indicators and project development milestones for subsequent reporting purposes.

Step 4. *Adoption of the capital budget*. The CIP's first year is the capital budget year.

CIP provides a blueprint for spending on transportation projects in a short range and plays important roles in transportation planning. Transportation planning agencies are also responsible for compiling long-range transportation plans (LRTP), which usually spans 20–25 years and applies programmatic transportation goals to project prioritization. Local Metropolitan Planning Agencies (MPOs) are the leading entity in local long-range transportation planning, with the assistance from collaborating agencies and local communities. Both CIP and LRTP are the foundation for the development of Transportation Improvement Programs (TIP) which will be administered and implemented at appropriate regional levels where the plans apply to. All these documents are critical steps to achieving an efficient and effective transportation planning system.

Long-Range Transportation Planning

Long-range transportation plans (LRTP) are planning documents addressing transportation goals, objectives, and strategies with a minimum of 20-year horizon, prioritizing financing and projects based on the transportation goals and objectives. These plans are usually updated every 5 years looking into the future. LRTP is the foundation for Transportation Improvement Programs (TIP). The planning process is conducted at the national, state, and regional levels. State departments of transportation conduct state-level plans, while the Metropolitan Planning Organization (MPO) create regional-level plans.

Federal Highway Administration (FHWA) and Federal Transit Administration (FTA) has specific requirements regarding LRTP, for example, stakeholder coordination and input, fiscal constraint, and certain technical topics, administrative topics, and new requirements periodically (FHWA & FTA, 2018). Based on *Federal Strategies for Implementing Requirements for LRTP Updates for the Florida MPOs*, regarding stakeholder coordination and input, MPOs are required to explicitly describe the procedures, strategies, and outcomes of stakeholder involvement in the planning process in the public participation plan (PPP) for all the MPO products and processes. When the planned area includes Indian Tribes and federal land, consultation with the tribal government and federal land management agencies is required as well. MPOs are also required to make consultation of various state and local agencies in land use management, natural resources, environmental protection, conservation, and historical preservation. MPOs are required to periodically review the effectiveness of the procedures and strategies in the PPP.

Regarding fiscal constraint, projects in long-range transportation plans are required to present with enough detail to develop cost estimates and how the projects will be implemented. Updated plan every 5 years needs to assess and revise projects in the previous LRTP. The LRTP also should include safety components, freight needs assessment, environmental mitigation/consultation, congestion management process, and Americans with Disabilities (ADA) Transition Plans or ADA Program Access Plans.

Long-range transportation plans need to be consistent with State Transportation Improvement Plans (STIP) and Transportation Improvement Plans (TIP), which are updated annually, and need to be adopted by the MPO board before they become active. FHWA has detailed specific requirements about what should be included in the adoption document.

In 2018, FHW established four new requirements, new planning factors, transportation performance management, multimodal feasibility, and transit asset management. Two new planning factors need to be considered in the next LRTP, (1) improving the resiliency and reliability of the transportation system and reducing or mitigating stormwater impacts of surface transportation; and (2) enhancing travel and tourism. Assessment of performance measures and targets, long-range and short-range strategies/actions for the development of integrated multimodal transportation systems need to be incorporated into future LRTPs. If there are multiple asset classes offered in the metropolitan planning area, the MPO should set targets for each asset class.

Other emerging and proactive issues, such as mobility on demand (MOD), which primarily comprises of Automatic Driving Systems (ADS) and Connected Vehicles (CV), consultation with agencies in tourism and natural disaster mitigation, summary of public involvement strategies, impact analysis/data validation, state-level department of transportation (DOT) revenue forecast, sustainability and livability in context, and scenario planning, may be addressed in future LRTPs per discretion of individual state LRTPs or local LRTPs.

Currently at the state level there are a few major types of LRTP plans (FHWA, 2012b):

- Policy-based. These plans mostly focus on policy directions instead of other elements, such as investment scenarios, performance measures, and specific projects.
- Performance-based. These plans use quantifiable metrics, targets, or timeframes to guide planning decisions.
- Needs-based. These plans focus on demographic or travel demand projections to analyze transportation needs, considering available or alternative revenues.
- Project-based. These plans focus on assessing whether project investments will meet the plan's transportation policies or goals.
- Fiscally realistic/constrained. These plans "set directions for the transportation system through analysis of projected capital and operating costs".

- Vision-based. These visions are often set through public input on a preferred vision.
- Corridor-based. These plans focus on specific transportation corridors.

The Future of Mobility

Technological advancement has tremendous impact on transportation. Modern transportation modes, such as automobiles, aircrafts, and speed trains, have greatly improved the mobility, accessibly, and safety of transportation. However, resource constraints and the pressure from population growth have prompted the transportation industry to look for better alternatives such as renewable energy, and faster and safer mass transit to accommodate the needs of individuals and households. Three new urban models are the future trend of mobility (Hannon et al., 2019):

- *Clean and shared*. This aspect stresses using cleaner fuel, such as electric vehicles (EVs), and shared mobility where individual ownership of automobiles is discouraged.
- *Private autonomy*. In areas where reliance on automobiles is significant, self-driving and electric vehicles, in conjunction with car sharing, ride hailing, and high-speed public transit services, might provide greater assistance in achieving private autonomy. Meanwhile, drones and unmanned Aerial Vehicles (UAVs) will provide significant assistance in freight delivery industries.
- *Seamless mobility*. In this system, "mobility is door to door and on demand…Mobility is delivered through a combination of self-driving, shared vehicles, with high-quality public transit as the backbone."

Seamless mobility is also manifested as a system of Flexible on Demand Transport Services (FDTS), which requires the involvement, commitment, and support of all the stakeholders, local communities, and the policy makers to be fully implemented (Barreto et al., 2019).

New transportation systems will require new sets of data and methods to analyze the key elements of transportation demand system, trip generation, trip distribution, mode choices, and route selections. By far, transportation is the most evolving system in the urban and regional environment, followed by technologies used in housing and e-governance. Challenges remain for the future of mobility, and regulation challenges and cybersecurity risks will add layers of uncertainty about what the future mobility entails.

References

Barreto, L., Amaral, A., & Baltazar, S. (2019). Flexible on demand transport services (FDTS): The future of mobility systems. In *2019 IEEE international conference on engineering, technology and innovation (ICE/ITMC)* (pp. 1–9). Valbonne Sophia-Antipolis. https://doi.org/10.1109/ICE.2019.8792649

Botchwey, N. D., Falkenstein, R., Levin, J., Fisher, T., & Trowbridge, M. (2014). The built environment and actual causes of death: Promoting an ecological approach to planning and public health. *Journal of Planning Literature, 30*(3), 261–281.
CDC (Centers for Disease Control and Prevention). (2010). *Teen Drivers: Fact Sheet. Injury Prevention & Control: Motor Vehicle Safety*. Accessed June 6, 2021. http://www.cdc.gov/MotorVehicleSafety/Teen_Drivers/teendrivers_factsheet.html
Evans, L. (1990). An attempt to categorize the main determinants of traffic safety. *Health Education Research, 5*(2), 111–124. https://doi.org/10.1093/her/5.2.111
FHWA (Federal Highway Administration). (2007). *Traffic analysis toolbox volume VI: Definition, interpretation, and calculation of traffic analysis tools measures of effectiveness*. Publication No. FHWA-HOP-08-054. U.S. Department of Transportation.
FHWA (Federal Highway Administration). (2012a). *Traffic analysis toolbox volume XII: Work zone traffic analysis – Applications and decision framework*. U.S. Department of Transportation.
FHWA (Federal Highway Administration). (2012b). Trends in Statewide Long-Range Transportation Plans.
FHWA (Federal Highway Administration). (2018). *Traffic data computation method pocket guide*. U.S. Department of Transportation.
FHWA (Federal Highway Administration) & FTA (Federal Transit Administration). (2018). Federal strategies for implementing requirements for LRTP updates for the Florida MPOs.
Federal Highway Administration (FHWA). n.d. Highway Functional Classifications. Retrieved from: https://www.fhwa.dot.gov/planning%20/processes/statewide/related/highway_functional_classifications/section03.cfm#Toc336872980
Furth, P. (2000). Data analysis for bus planning and monitoring. In *Synthesis of transit practice 34*. Transportation Research Board, National Research Council.
Hannon, E., Knupfer, S., Stern, S., Sumers, B., & Nijssen, J. T. (2019). *An integrated perspective on the future of mobility*. McKinsey Center for Future Mobility. McKinsey&Company.
ITE (The Institute of Transportation Engineers). (2010). *Designing walkable urban thoroughfares: a context sensitive approach.*
ITE (The Institute of Transportation Engineers). (2016). *Transportation planning handbook* (4th ed.). John Wiley & Sons, Inc..
ITE (The Institute of Transportation Engineers). (2017). *The trip generation manual* (10th ed.).
Jeong, I., & Shin, S. J. (2017). High-performance work practices and organizational creativity during organizational change: A collective learning perspective. *Journal of Management, 45*(2), 1–17. https://doi.org/10.1177/0149206316685156
Koutsopoulos, H. N., Ma, Z., Nouralehi, P., & Zhu, Y. (2019). Chapter 10 – Transit data analytics for planning, monitoring, control, and information. In *Mobility patterns, big data, and transport analytics: Tools and applications for modeling*. Elsevier.
Mathew, T. V., & Rao, K. V. K. (2007). *Introduction to transportation engineering*. NPTEL. A Project Funded by MHRD, Govt. of India.
Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2004). Actual causes of death in the United States, 2000. *Journal of the American Medical Association, 291*(10), 1238–1245.
Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2005).'Correction: Actual Causes of Death in the United States, 2000. *Journal of the American Medical Association 293*, 293–294.
NCSS. (2020). Chapter 326: Negative binomial regression. NCSS statistical software. Retrieved on October 20 from: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative_Binomial_Regression.pdf.
Opengov.com. (2019). *Capital planning in the digital era*.
Rentsch, A., Feldman, J., & Ramaswamy, L. (2019). Machine learning for urban planning: Estimating parking capacity. Harvard Data Science Capstone Project, Fall 2019. Towarddatascience.com, Dec. 15, 2019. https://towardsdatascience.com/machine-learning-for-urban-planning-estimating-parking-capacity-15aabd490cf8
Salter, R. J. (1976). *Highway traffic analysis and design*. Palgrave. https://doi.org/10.1007/978-1-349-06952-1

Sharma, A. K., & Landge, V. S. (2013). Zero inflated negative binomial for modeling heavy vehicle crash rate on Indian rural highway. *International Journal of Advances in Engineering & Technology, 5*(2), 292–301.

U.S. Department of Transportation Federal Highway Administration (FHWA). (2020). *Highway functional classification concepts, criteria and procedures*. Accessed on 20 Oct 2020 from: https://www.fhwa.dot.gov/planning/processes/statewide/related/highway_functional_classifications/section03.cfm.

Urban Land Institute (ULI). (1993). *The dimensions of parking*. National Parking Association.

Web Resources

Transportation Research Board: http://www.trb.org/AboutTRB/AboutTRB.aspx
Federal Highway Administration: https://safety.fhwa.dot.gov/
APA Transportation Planning Division: https://transportation.planning.org/
Urban Land Institute: https://uli.org/
Institute of Transportation Engineers: https://www.ite.org/
Federal Transit Administration: https://www.transit.dot.gov/
Transit Planning Resources Library: https://www.transit.dot.gov/regulations-and-guidance/transportation-planning/planning-resource-library
U.S. Department of Transportation: https://www.transportation.gov/
Transportation Planning Resources Library: https://www.transportation.gov/leadershipacademy/resource-library

Chapter 14
Environmental Analysis

Two types of worldviews, expansionist and ecological, influence planning practice (Rees, 1995). Traditional planning, which focuses on the physical layout and the economic efficiency, follows the expansionist worldview. These planners view the environment as the provider of material and energy resources, life-support and waste assimilation services, and aesthetic amenities to support endless growth and to maintain the quality of life for people. They consider environmental scarcity as an economic issue and rely on technology to solve environmental constraints on the economic expansion. With the growing environmental consciousness (Krause, 1993), the ecological worldview began to influence the planning process and urban policy formations. Following this view, planners view the environment as an independent and complex system and urban growth depends on the ecosphere. The introduction of the prism sustainability model into the planning profession integrates the two worldviews into one theory to guide planning practice (Campbell, 1996). The sustainability model acknowledges the importance of both the environment and the human society, which is interpreted as economy and equity, and presents a holistic development paradigm which internalizes the conflicting but complementary relations among economy, equity, and environment. Guided by this model, environmental analysis becomes necessary for all types of planning activities. Urban growth should consider the impacts on the environment and should be paced out compatible with the physical and biological constraints. This chapter introduces basic environmental elements, revisits suitability analysis for analyzing environmental sensitivity, and addresses various other environmental analysis techniques.

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_14].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_14

Environmental Elements

The environment includes four subsystems: the lithosphere, the hydrosphere, the atmosphere, and the biosphere (Kemp, 2004). Corresponding to each of the subsystems, the chapter introduces topography and soil, watershed, climate, microclimate, air, and biodiversity.

Topography describes the surface of a land. The U.S. Geological Survey (USGS) regularly publishes topographic maps, which present information of elevation, hydrography, and places. These maps have been extensively used by local agencies for planning and land use management purposes. Figure 14.1 presents an example. The lines on the map are contour lines. The locations on the same contour line have equal elevation. For example, the elevation of location A is 520 feet. The elevations of all the locations on the same contour line are 520 feet. Contour lines also show the features of a terrain. Areas with denser contour lines are steeper and on the contrary, areas with contour lines farther apart are flat. In Fig. 14.1, location B has a dense set of contour lines. This means that the elevation quickly changes here. As a comparison, the area around location C has a sparse set of contour lines, suggesting a flat area which is more suitable for urban development.

Slope evaluates the change in the elevation over a given horizontal space, with a value between 0 degree and 90 degree. It can be calculated based on elevation data. Considering slope in development can reduce construction costs and potential risks. An area with a steeper slope tends to be more unstable and has a higher probability of natural hazard occurrences such as flooding and landslide. There is no consensus

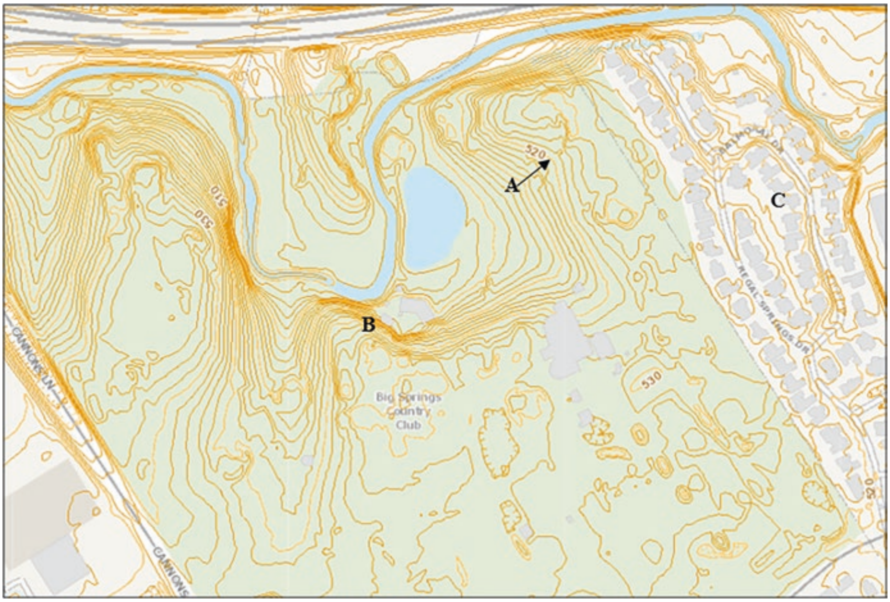


Fig. 14.1 An example of topographic map. (Data Source. The LOJIC (2020))

about what the maximum slope for a development is. Marsh (2010) recommends the following maximum slopes for various land uses: 20–25% for residential development, 3–4% for industrial uses, 15–17% for streets, and 3% for parking. He also recommends 1–2% as the optimum slope for most land uses to avoid the drainage issue and environmental risks.

Soil contains water, air, mineral particles, and organic matters. When study soil, one needs to study the whole soil profile, which includes the organic horizon, the surface horizon, the sub-surface horizon, and the substratum. Soils can be different in texture, with varying composition of sand (0.05–2 mm particle size), silt (0.002–0.05 mm), and clay (<0.002 mm). There are physical differences among soils, as presented in the following.

- The ***infiltration rate***, sometime called the intake rate, measures how fast water moves into the soil surface. Soil texture, coverage, and the speed of water supply affects the infiltration rate. When the infiltration speed is slower than the water accumulation speed, water will pool on the surface and move as runoff. Urban activities modify the soil surface and reduce the infiltration rate. As a result, urban runoff becomes an urgent issue for planners to deal with.
- ***Erodibility*** measures the proportion of soil that will erode under standard conditions. It is also called the *k*-factors. Rainfall attributes, such as droplet size and maximum intensity, directly affect the possibility of a soil erosion. Different soils have different propensity to erode. Silty soils tend to be more erosive and clay soils less erosive. Topographic attributes contribute another set of reasons for erosion. Soil erosion increases with slope steepness and length. The presence of vegetation on the soil surface may reduce erosion, as plants can act as barriers to reduce the amount of runoff and increase the soil infiltration rate. At last, the presence of appropriate erosion control techniques is expected to reduce erodibility.
- ***Bearing capacity*** is the ability of soil to support the weight on top of it. Soils with finer grain tend to have lower level of bearing capacities than coarse granular soils. Bearing capacity of the soil affects the stability of a foundation and the safety of a building (Bowles, 1996).
- ***Shrink-swell potential*** is the tendency of soil to shrink when dry and expand when wet. Expansive soils, soils with high shrink-swell potentials, cause billions of dollars of damage in properties in the U.S. every year, as they can damage building foundations (Thomas et al., 2000).

There are chemical differences between soils, for example, corrosion potential and fertility. Soils with high levels of moisture, acidity, and salt may be corrosive. Chemical reactions can occur between such soils and construction materials (e.g., concrete, metals), and this may damage building foundations and underground water and sewer pipelines. Fertility refers to soils' ability to support plant growth by providing necessary nutrients.

From the land management perspective, the major concerns related to soil include (1) the declining quality of soils as rooting environments, (2) the erosion and loss of topsoil by wind and water, (3) a loss of vegetation cover, (4) acidification, soil fertility decline, and plant nutrient depletion, and (5) the salinization in the soil, especially in irrigated systems.

The U.S. Natural Resources Conservation Service (NRCS) regularly produces the ***Soil Survey Geographic Database*** (SSURGO), which provides information about soil types and the distribution across the nation. This database provides basic physical information of soils, such as slope gradient, bedrock depth, water table depth, flooding frequency, available water storage, drainage condition, and potential erosion hazard. More importantly, this database also provides ratings about soil suitability for different types of developments, including dwelling with or without basement, septic tank absorption fields, sewage lagoons, local roads and streets, and paths and trails. Planners can use this information to assess whether the local soil can support a specific type of development.

Climate refers to the weather condition of a region over a long time. It is measured by temperature, humidity, and air movements. Climate affects individual's comfort levels, with 70–80 degree temperature and 20–50% relative humidity bringing the highest level of human comfort. Climate directly determines heating and cooling degree days and affects households' energy use. Climate presents as design constraints for regional infrastructure systems, such as water supply and stormwater treatment that are influenced by rainfall levels. Extreme climate events (flooding, drought, hurricane, extreme temperature or wind) make cities and city dwellers vulnerable.

Microclimate is a concept of interest for cities. Urban activities modify the surface of a land, interfere with water and air movements, and, consequently, may create a local atmosphere that differs from the surrounding areas. The ***urban heat island effect***, that is the temperature of urban areas tend to be higher than surrounding areas, offers an example of microclimate. Planners need to understand microclimate and utilize this unique set of knowledge to promote sustainability. For instance, Erell, Peralmutter, and Williamson (2012) propose climatically responsive urban design to better use natural energy in buildings. Bourbia and Boucheriba (2010) explore the possibility of using the street canyon effect to influence the local microclimate.

A ***watershed***, sometimes called a drainage basin, is an area with all water, including streams, precipitation, ground runoff, flowing to a common outlet. Figure 14.2 presents major watersheds in the U.S. These watershed boundaries are not consistent with any existing political boundaries, including the country's boundary. For instance, the Great Lakes watershed boundary extends into Canada. Urban growth modifies the surface of a watershed, redirects hydrological patterns, and introduces pollutants into the system. Watershed planning becomes increasingly important to protect watersheds and to restore watersheds' ability for providing ecosystem services (Brabec et al., 2002; Ryan & Klug, 2005).

Air conditions directly affect people's health and the environmental quality. The U.S. Environmental Protection Agency (EPA) regularly monitors six common air pollutants, ground level ozone, particulate matter, carbon monoxide, lead, sulfur dioxide, and nitrogen dioxide, as required by the Clean Air Act. Table 14.1 summarizes information of the six pollutants (2020a). The EPA, state, and local

Fig. 14.2 Watersheds in the U.S. (Source: The USGS (2020))

agencies have established a network of pollutant monitoring stations. Data collected from these stations are published in the EPA AirData database which can be accessed through https://www.epa.gov/outdoor-air-quality-data. The AirData database allows for displaying and downloading hourly, daily, and annual pollutant information from each monitor. This database also provides annual air quality index reports for U.S. counties and metropolitan areas.

Biodiversity, or biological diversity, refers to the number of different kinds of life in an ecosystem. Biodiversity is the indicator for the health of an ecosystem. A more mature and productive ecosystem is expected to be able to accommodate more species, including both animals and plants. Direct benefits from a high biodiversity include food security and recreational benefits. The diversity of an ecosystem also enables a fast recovery from natural disasters. The main cause of the loss of biodiversity is related to human activities that degrade the environment. With an increasing level of pollution in water, air, and soil, the living environment of animals and plants are being destroyed which leads to a declining number of species in an ecosystem. The loss of species further accelerates the environmental degradations of air, water, and soil. Biodiversity can serve as an indicator for the status of a specific environmental element. For instance, the EPA uses the condition and the types of species living in the water as the measure for the health of an aquatic system.

Many cities address all or part of these environmental elements in their comprehensive plan. For instance, the environmental element of the Seattle 2035 Comprehensive Plan addresses land, water, and climate (City of Seattle, 2020). Various federal, state, and local agencies closely monitor the conditions of major

Table 14.1 Major air pollutants, causes, and impacts

Pollutants	Causes	Impacts
Ground level ozone	Cars, power plants, factories emit pollutants, which have chemical reaction in the presence of sunlight and create ozone.	Aggravate asthma Cause permanent lung damage Affect sensitive vegetation and ecosystems
Particulate matter (PM)	Some particulate matters are directly emitted into the air from urban activities (e.g. construction, fires, driving). Some are formed in the air due to chemical reactions of pollutants.	Affect lungs and hearts Reduce visibility
Carbon monoxide (CO)	Most CO is from vehicles and machines that burn fossil fuels.	Reduce the oxygen level in blood Cause dizziness, confusion, unconsciousness, and death
Lead	Lead may come from Ore and metals processing and piston-engine aircraft operating on leaded aviation fuel, waste incinerators, utilities, and lead-acid battery manufacturers	Affect the oxygen carrying capacity of the blood, the nervous system, kidney function, immune system, reproductive and developmental systems, and the cardiovascular system Contribute to behavioral problems, learning deficits and lowered IQ for young children and infants.
Sulfur dioxide (SO_2)	The major source is the burning of fossil fuels by power plants and other industrial facilities.	Harm the human respiratory system Lead to the formation of other sulfur oxides in the air which contributes to particulate matter (PM) pollution Harm trees and plants by damaging foliage and decreasing growth Cause visibility and regional haze
Nitrogen dioxide (NO_2)	NO_2 forms from emissions from cars, trucks and buses, power plants, and off-road equipment.	Irritate airways in the human respiratory system React with other chemicals in the air to form both particulate matter and ozone form acid rain Make the air hazy Contributes to nutrient pollution in coastal waters

environment elements. The U.S. and state EPA agencies operates a network of outdoor monitors to track major pollutant levels around the clock and throughout the year. The U.S. Department of Agriculture maintains the Soil Survey Geographic (SSURGO) databases. This database provides useful information for communities to understand the characteristics of local soil.

Measures of the Environment

The ***universal soil loss equation*** provides an empirical technology to estimate soil loss by rainfall impact and surface runoff (Wischmeier & Smith, 1965). The amount of soil loss in tons per acre and per year can be calculated as:

$$\mathrm{A} = \mathrm{R}^{*}k^{*}\mathrm{LS}^{*}\mathrm{C}^{*}\mathrm{P}$$

where

R is the rainfall factor, which can be estimated as a function of an average annual precipitation. The EPA (2020b) offers a rainfall erosivity factor calculator service. The system asks for inputs of a location and a specific time and returns the rainfall factor accordingly.
k is the soil erodibility factor, which is available in the SSURGO soil database.
LS is the slope length and degree factor. It is calculated based on the slope length and the slope gradient. Renard et al. (1997) provide a detailed description for calculating LS factors. They also provide several LS tables where a user can directly look for a LS factor for a combination of a slope length and a slope gradient.
C is the cropping and management factor. It reflects the impact of cropping and management activities on erosion rates. One approach to find the C-factor is to explore the existing literature and assign uniform C-factor values (Panagos et al., 2015). Some state governments may provide advisory information on selecting an appropriate C-factor value. For instance, Michigan State Administrative Rules R 280.9 provides suggested runoff coefficients for different types of land.
P is the conservation practice factor. Supportive conservation practices are expected to reduce runoff volume. The P-factor reflects this. It is the ratio of soil loss with a specific conservation practice to the soil loss without any conservation practice.

The ***air quality index*** (AQI) is an indicator for the pollution level in the air, with an index value between 0 and 50 indicating a low level of pollutants and therefore a good air quality, a value between 51 and 100 suggesting a moderate air quality, a value between 101 and 150 indicating an unhealthy air quality for sensitive groups, a value between 151 and 200 indicating an unhealthy air quality, a value higher than 201 and 300 indicating a very unhealthy air quality, and a value 301 and above suggesting the existence of hazardous air.

To find the AQI, first calculate the index for every pollutant (Ozone, PM2.5, PM10, CO, SO_2, and No_2), according to the following formula:

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}}\left(C_p - BP_{Lo}\right) + I_{Lo}$$

where I_p is the index for pollutant p, C_p is the rounded concentration of pollutant p, BP_{Hi} is the breakpoint that is greater than or equal to C_p, BP_{Lo} is the brekpoint that

Table 14.2 Breakpoint values

Breakpoints of an observed value (***BP**$_{Hi}$*, ***BP**$_{Lo}$*)							AQI (I_{Lo}, I_{Hi})	Description
O3(ppm)8-hour	O3(ppm)1-hour	PM10 (μg/m3)24-hour	PM2.5 (μg/m3)24-hour	CO(ppm)8-hour	SO2(ppb)1-hour	NO2(ppb)1-hour		
0.000–0.059	–	0–54	0.0–15.4	0.0–4.4	0–35	0–53	0–50	Good
0.060–0.075	–	55–154	15.5–40.4	4.5–9.4	36–75	54–100	51–100	Moderate
0.076–0.095	0.125–0.164	155–254	40.5–65.4	9.5–12.4	76–185	101–360	101–150	Unhealthy for Sensitive Groups
0.096–0.115	0.185–0.204	255–354	65.5–150.4	12.5–15.4	186–304	361–649	151–200	Unhealthy
0.116–0.374	0.205–0.404	355–424	150.5–250.4	15.5–30.4	305–604	650–1249	201–300	Very unhealthy
–	0.405–0.504	425–504	250.5–350.4	30.5–40.4	605–804	1250–1649	301–400	Hazardous
–	0.505–0.604	505–604	350.5–500.4	40.5–50.4	805–1004	1650–2049	401–500	Hazardous

Source: The EPA (2020c)

is less than or equal to C_p, I_{Hi} is the AQI value correponding to BP_{Hi}, and I_{Lo} is the AQI value corresponding to BP_{Lo}. Table 14.2 provides the breakpoint values for an index calculation. With the indices calculated for all the pollutants, the AQI is defined as the highest index value among the available ones. The pollutant with the highest index value, i.e. causing the most serious concern about the air quality, is the critial pollutant.

The Environmental Inventory Analysis

An ***environmental inventory analysis*** can help planners to understand the existing physical opportunities and limitations imposed by the environment. For instance, the U.S. Department of Agriculture operates the National Resource Inventory program to produce information on the status and trends of major natural resources, including land, soil, and water. According to Berke and Stevens (2016), an environmental inventory analysis is to assemble information of major environmental elements, which have been explored at the beginning of this chapter. The purposes of such an analysis are to help planners to understand (1) the current stock and productivity of the existing natural resources, (2) the ecological function that a natural resource offers the community, (3) the potential public health threat, and (4) ethical, aesthetic, and spiritual values of natural resources. Many state and local governments have internalized environmental inventory analysis and environmental impact analysis in their regulatory systems. For instance, the State of New Jersey requires an environmental assessment for a development permit application. Box 14.1 lists the major environmental elements as listed in the assessment form of New Jersey. The existing condition section is a survey of the existing environmental inventory.

An ***environmental inventory analysis*** can help planners to understand the existing physical opportunities and limitations imposed by the environment. For instance, the U.S. Department of Agriculture operates the National Resource Inventory program to produce information on the status and trends of major natural resources, including land, soil, and water. According to Berke and Stevens (2016), an environment inventory analysis is to assemble information of major environmental elements, which have been explored at the beginning of this chapter. The purposes of such an analysis are to help planners to understand (1) the current stock and productivity of the existing natural resources, (2) the ecological function that a natural resource offers the community, (3) the potential public health threat, and (4) ethical, aesthetic, and spiritual values of natural resources. Many state and local governments have internalized environmental inventory analysis and environmental impact analysis in their regulatory systems. For instance, the State of New Jersey requires an environmental assessment for a development permit application. Box 14.1 lists the major environmental elements as listed in the assessment form of New Jersey. The existing condition section is a survey of the existing environmental inventory.

Box 14.1 Environmental Elements in the Environmental Assessment Checklist for Development Permit Applications in New Jersey

Existing Condition

- Geology
- Soils
- Surface waters.
- Subsurface water
- Topography and existing development features
- Wetlands and State open waters
- Floodplains
- Vegetation
- Wildlife
- Endangered or threatened species
- Air Quality
- Noise
- Cultural, historical and archaeological resources
- Unique features and scenic resources
- Land use

Construction Impacts

- Soil erosion and sedimentation resulting from surface runoff
- Flooding and flood plain disruption
- Degradation of surface water quality

- Ground water pollution
- Reduction of ground water capabilities
- Sewage disposal
- Solid waste disposal
- Vegetation destruction and disruption of vegetative communities
- Disruption of wildlife habitats particularly of protected species
- Destruction or degradation of scenic and historic features
- Air quality degradation
- Noise levels
- Energy utilization
- Wetland impacts

Information Source: https://www.nj.gov/dep/pcer/introcklist.htm.

The Environmentally Sensitive Area Analysis

The suitability analysis has been introduced in Chap. 12. When applying the suitability analysis in an environmental analysis, the primary focus is on identifying ***environmentally sensitive areas*** (ESAs). ESAs are areas with special environmental attributes worthy of protection. Table 14.3 lists several examples of ESAs and analyzes the potential benefits of protecting them. First, most ESAs have their economic functionality and can enhance the quality of life for people. The literature has shown that the accessibility to environmental amenities, such as beaches, creeks, parks, increases property values (Doss & Taff, 1996; Crompton, 2001). The protection of prime agriculture lands, i.e., lands with the highest quality of agriculture yields, preserves the agriculture productivity and enhances food security. Second, designating wildlife habitats, which could be a lake, a beach, a forest, or a

Table 14.3 Examples of environmentally sensitive areas and the benefits of preservation

	Economic/Recreational importance	Bio-diversity	Hazard mitigation	Pollution mitigation
Surface water	X	X		X
Beaches	X	X		
Dunes		X	X	
Wetlands	X	X	X	X
Wildlife/Vegetation Habitats	X	X		
Forests	X	X		X
River corridors	X	X	X	
Floodplains			X	
Prime agricultural lands	X			

specifically designated wildlife habitat, as ESAs is one of the strategies to preserve biodiversity. Biodiversity serves as the primary indicator for the health of our ecosystem. Third, some ESAs are critical for hazard mitigations. For instance, river corridors provide physical spaces for rivers to express their energy and wander around. It is important to preserve these areas and implement various hazard mitigation strategies, such as bank stabilization practices and vegetated buffers. At last, there is a pollution mitigation benefit of ESAs. Surface water sources are susceptible to pollution because of people's convenient accessibility to lakes, streams, and creeks. Designating these areas as ESAs and implementing stronger preservations policies can reduce pollution. Trees in forests can reduce particulate pollution by intercepting particles in the air and reduce hazardous nitrogen oxides, sulfur dioxide, ozone, and ammonia pollutions by absorbing these gaseous pollutants.

It is not difficult to identify the geographic coverage for the ESAs listed in Table 14.3. The common exercise is to define specific areas (e.g. the wetland, a wildlife habitat, or a forest) as ESAs or to designate a buffered area around a subject area (e.g. river corridor) as an ESA. Many times, planners need to monitor environmental degradation and identify ESAs based on multiple environmental elements, such as soil, slope, water, and air. The suitability analysis can be used to identify the most suitable areas to be preserved. It will be a four-step process. Figure 14.3 provides an example of Symeonakis et al. (2016). Their study identifies ESAs in Lesvos, Greece.

The first step is to decide on criteria. As shown in Fig. 14.3, the selected criteria cover different environmental aspects, water, soil, vegetation, and climate. Each aspect of the environment has different intrinsic characteristics for determining which locations are suitable for ESA designations. Social-economic attributes are included as well. This is related to the purpose of designating ESAs in urbanized areas, which is to protect them from human activities. Next, Symeonakis et al. (2016) collect data and calculate suitability indices. Given the complexity set of criteria in the analysis, they first calculated five sub-indices, based on which they calculate the final ESA indices. These final ESA indices allows for a further decision about which land should be preserved as ESAs.

An ESA analysis has a comprehensive geographic coverage (i.e., the whole region or city) and involves data from many sources of different environmental attributes. Such an analysis relies on raster calculation tools in a Geographical Information System (GIS). Raster calculation tools are for analyzing image-based data. Figure 14.4 illustrates the use of raster calculation tools for a hypothetical ESA analysis, which involves three factors. With a raster data, a real-world phenomenon is presented with a series of continuous cells with equal size. Slope, soil, and vegetation attributes are stored in three different raster datasets. The measures of these three attributed are converted to the same scale, with 1 indicating the least level of environmental concerns and 3 the highest level of concerns. With the weights determined (0.3 for slope, 0.5 for soil, and 0.2 for vegetation), one can calculate the ESA index with a weighted overlay tool in GIS, with the ESA Index = 0.3*Slope + 0.5*Soil +0.2*Vegetation. Based on the results, ESAs can be designated.

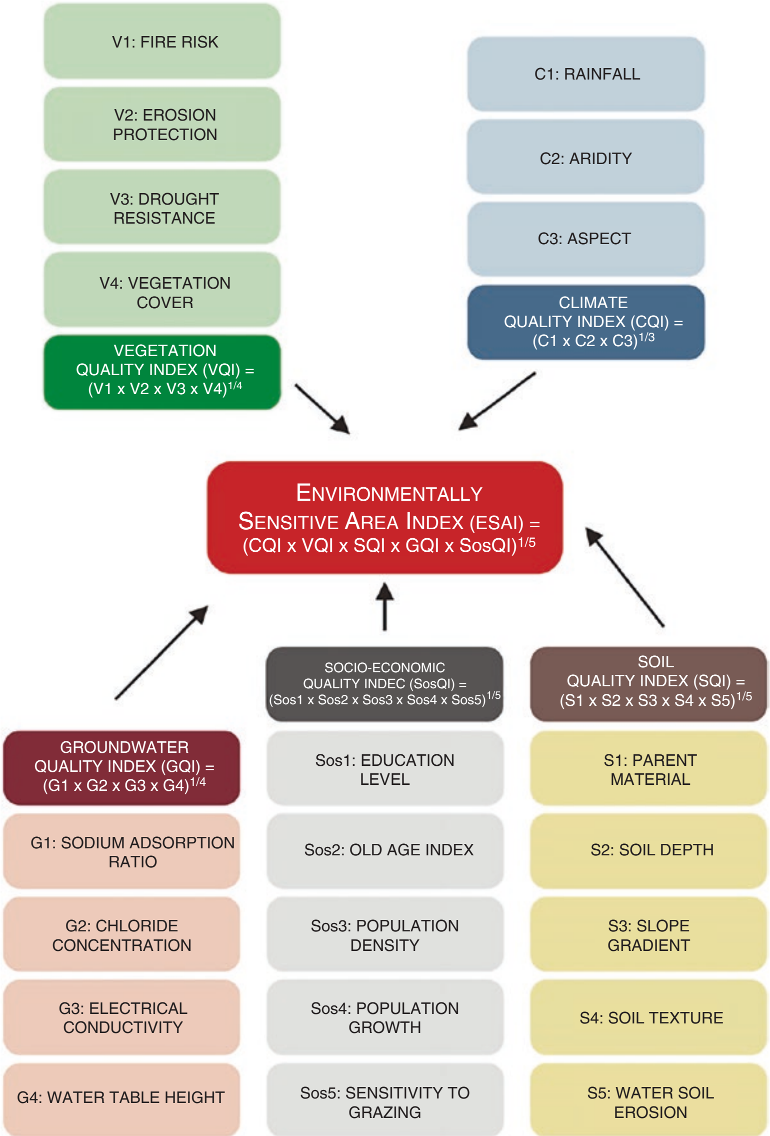


Fig. 14.3 An example of constructing ESA index. (Source: Symeonakis et al. (2016). Reproduced by permission from Wiley)

There are two reasons for conducting an ESA analysis using raster analysis tools. First, such an analysis uses data of multiple environmental elements for varying sources, with inconsistent environmental boundaries. Raster data are cell based and can handle the problem of varying boundaries. Second, many environmental data are produced as raster data, for instance, the elevation data and the vegetation data.

Slope

1	2	2	1
2	3	2	2
2	3	3	1
1	3	2	2

Soils

2	3	3	1
3	3	1	1
2	3	1	1
3	1	2	3

Vegetation

1	2	2	3
2	3	3	1
3	2	1	1
2	3	3	2

Weighted overlay
slope=0.3, soil=0.5, vegetation=0.2

1.5	2.5	2.5	1.4
2.5	3.0	1.7	1.3
2.2	2.8	1.6	1.0
2.2	2.0	2.2	2.5

Fig. 14.4 ESA analysis in GIS with raster data

The Environmental Carrying Capacity

A ***carrying capacity*** is the growth limit of an environment and the maximum level of development without weakening the productivity and the functional integrity of the environment. The purposes of conducting a carrying capacity analysis are to promote environmentally-conscious urban growth and to guide land use developments based on environmental constraints. The general approach to assess an environment's carrying capacity is to compare the demand for an environmental resource with the supply of the environment resource. Many factors drive the demand for environmental resources. Our economy is based on an intensive use of resources and relies on technology and innovation that improve economic productivity but, at the same time, increase the demand for natural resources. Further, the growing population and increasing wealth and quality of life impose pressures on the environment with increasing demand for water, energy, and materials. Factors affecting the supply side of a carrying capacity analysis include the availability of and the accessibility to environmental resources.

The Florida Keys Carrying Capacity Study (Clarke, 2002; URS Corporation, 2002) presents an example of such study. The study balances between human preferences and the need to protect the ecological system. The end product of the study is a GIS-based interactive scenario generator that accounts for various relationships between population growth and the natural environment and evaluate the impacts of each scenario on natural resources, human infrastructure, and the social environment in relation to the tolerance-limits of the natural environment. Maps are generated to show areas with human impacts exceeding natural capacities.

The Ecological Footprint

The ***ecological footprint*** is a concept related to the environmental carrying capacity. Both address the relation between the environment and an urban growth. The environmental carrying capacity focuses on analyzing the growth as constrained by the environment. The ecological footprint studies environment-growth relation from the perspective of the environment, that is the amount of environmental resources (e.g. land, air, soil) required to support the growth and to absorb its waste. The EPA has developed a spreadsheet to guide an ecological footprint analysis, as shown in Table 14.4. This spreadsheet documents the potential impacts of an activity on different elements of the environment.

Galli et al. (2007) propose an accounting method to calculate the ecological footprint of a region. To do the calculation, one needs to collect annual resource

Table 14.4 Spreadsheet for environmental footprint analysis

Core element	Green remediation metric		Unit of measure
Materials & waste	M&W-1	Refined materials used on site	tons
	M&W-2	Percent of refined materials from recycled or waste material	percent
	M&W-3	Unrefined materials used on site	tons
	M&W-4	Percent of unrefined materials from recycled or waste material	percent
	M&W-5	Onsite hazardous waste generated	tons
	M&W-6	Onsite non-hazardous waste generated	tons
	M&W-7	Percent of total potential onsite waste that is recycled or reused	percent
Water		Onsite water use (by source)	
	W-1	Source, use, fate combination #1	millions of gallons
	W-2	Source, use, fate combination #2	millions of gallons
	W-3	Source, use, fate combination #3	millions of gallons
	W-4	Source, use, fate combination #4	millions of gallons
Energy	E-1	Total energy use	MMBtu
	E-2	Total energy voluntarily derived from renewable resources	
	E-2A	Onsite generation or use and biodiesel use	MMBtu
		E-2B- Voluntary purchase of renewable electricity	MWh
	E-2C	Voluntary purchase of RECs	MWh
Air	A-1	Onsite NOx, SOx, and PM10 emissions	lbs
	A-2	Onsite HAP emissions	lbs
	A-3	Total NOx, SOx, and PM10 emissions	lbs
	A-4	Total HAP emissions	lbs
	A-5	Total GHG emissions	tons CO_2e
Land & Ecosystems	Qualitative description		

Source: https://clu-in.org/greenremediation/SEFA/

consumption and waste generation data for both the nation and the study region. Table 14.4 can provide a guidance for the data collection process. The ecological footprint of the subject region (EF) can be calculated as:

$$EF = \sum_i \frac{T_i}{Yw_i} * EQF_i$$

where T_i is the total annual consumption of the i^{th} product, i.e., the final demand from local households for product i, Y_{w_i} is the annual national yield of the i^{th} product, and EQF_i is the equivalence factor the i^{th} product. To use the method, one can collect consumer expenditure data from the Census Bureau to understand the amount of total consumption of different type of goods. Annual national yields data of the U.S. are available from multiple sources, such as the Economic Census and the national input-output accounts. Every type of product is related to one or a set of land types. The equivalence factor is calculated for every land type, as the ratio of the maximum potential ecological productivity of a specific land type over the average productivity of all productive lands. A higher than one equivalence ratio indicates that the production of the product involves the use of higher-than-average-productivity lands, and therefore, increases the ecological footprint. For instance, in the Ecological Footprint study of San Francisco, California, the equivalent factors used are 2.51 for cropland, 1.26 for forest, 0.46 for grazing land, 0.37 for marine and inland water, and 2.51 for built-up land (Global Footprint Network, 2011). Box 14.2 presents the method used for calculating San Francisco ecological footprint. The method is slightly different from the formula presented above. It replaces the total annual consumption of a product with the economic impact of this annual consumption, i.e., the number of goods and services produced from the economic system to satisfy the consumption from pure consumers, i.e., local households.

Box 14.2 Ecological Footprint Calculations

The Global Footprint Network and the San Francisco Planning and Urban Research Association teamed up in 2011 and studied the ecological footprint of the San Francisco Metropolitan Statistical Area. The following summarizes the overall process for calculating the ecological footprint.

1. Study the final demand of a region. The study can be based on multiple data sources of consumer expenditures. Per the request of the Bureau of Labor Statistics, the Census Bureau regularly conducts the consumer expenditure survey. Household expenditure by category information is available from this database for many metropolitan statistical areas in the U.S. The Bureau of Economic Analysis of the U.S. Department of Commerce releases annual data on personal consumption expenditures for U.S. states. These data reflect household consumptions, i.e., the amount of final demand from individuals of a region.

2. Use the national input-output matrix to calculate the Leontief inverse. The Bureau of Economic Analysis publishes national input-output accounts data. Such data for metropolitan areas can be purchased from the IMPLAN. As introduced in Chap. 10, the Leontief inverse matrix can be derived from a matrix that contains input-output relations. This Leontief inverse matrix directly links final demand to economic productions.
3. Study the impacts of the region's final demand based on the final demand and the Leontief inverse matrix. The results present the total outputs from every industry due to the final demand in this region.
4. Study the environmental impacts from the economic productions. Calculate the footprints of each production on different types of land, including cropland, grazing land, fishing ground, forest land, built up land, and carbon uptake.

The Environmental Impact Assessment

An ***environmental impact assessment*** is to evaluate the potential impacts, which can be beneficial or adverse, of a proposed development on the environment. The National Environmental Policy Act requires assessments of environmental impacts for any federal activities. Many states followed up with the national legislative effort and adopted state environmental policy acts. Sixteen states require environmental impact assessments for state actions and projects. These states include California, Connecticut, Georgia, Hawaii, Indiana, Maryland, Massachusetts, Minnesota, Montana, New Jersey, New York, North Carolina, South Dakota, Virginia, Washington, and Wisconsin. California, Minnesota, and New York further require environmental impact assessments for some or all private activities from individuals and businesses.

In practice, an environmental impact assessment can be conducted through several steps. The first step is to create a checklist of potential impacts. In states with strong environmental regulations, it will be the responsibility of the public agency or the planning agency to develop a comprehensive checklist of potential environmental impacts from projects. The list should cover major environmental components and collect project information that can be used to analyze environmental impacts. For instance, California's environment impact checklist covers nineteen categories, including aesthetics, agriculture and forestry resources, air quality, biological resources, cultural resources, geology/soils, greenhouse gas emissions, hazards and hazardous materials, hydrology and water quality, land use and planning, mineral resources, noise, population and housing, public services, recreation, transportation and traffic, tribal cultural resources, utilities and service systems, and mandatory findings of significance. Figure 14.5 presents the checklist questions for the biological resources category.

IV. BIOLOGICAL RESOURCES: Would the project:	Potentially Significant Impact	Less Than Significant with Mitigation Incorporated	Less Than Significant Impact	No Impact
a) Have a substantial adverse effect, either directly or through habitat modifications, on any species identified as a candidate, sensitive, or special status species in local or regional plans, policies, or regulations, or by the California Department of Fish and Game or U.S. Fish and Wildlife Service?	☐	☐	☐	☐
b) Have a substantial adverse effect on any riparian habitat or other sensitive natural community identified in local or regional plans, policies, regulations or by the California Department of Fish and Game or US Fish and Wildlife Service?	☐	☐	☐	☐
c) Have a substantial adverse effect on federally protected wetlands as defined by Section 404 of the Clean Water Act (including, but not limited to, marsh, vernal pool, coastal, etc.) through direct removal, filling, hydrological interruption, or other means?	☐	☐	☐	☐
d) Interfere substantially with the movement of any native resident or migratory fish or wildlife species or with established native resident or migratory wildlife corridors, or impede the use of native wildlife nursery sites?	☐	☐	☐	☐
e) Conflict with any local policies or ordinances protecting biological resources, such as a tree preservation policy or ordinance?	☐	☐	☐	☐
f) Conflict with the provisions of an adopted Habitat Conservation Plan, Natural Community Conservation Plan, or other approved local, regional, or state habitat conservation plan?	☐	☐	☐	☐

Fig. 14.5 An example of environmental impact checklist. (Source: This figure is part of the CEQA Appendix G: Environmental Checklist Form)

With the information collected from the environmental impact checklist, one can develop an impact matrix. Figure 14.6 presents an example. Such matrix lists tabular information of the impacts of a project on the environment. In Fig. 14.6, column

	Envisaged impact factors	PROJECT ACTIVITIES: Placement of wind turbines	Foundation of columns	The use of building materials	Substation construction	Transmission line construction	Construction of internal roads	Operation of construction equipment	Waste material Treatment	Project exploitation	Sum of IF values by types and bio. comp..	Average values
PHYSICAL COMPONENTS	Water	0	0	0	0	0	0	0	0	0	0	0.00
	Microclimate	0	0	0	0	0	0	0	0	0	0	0.00
	Land	1	2	1	1	1	1	2	2	1	12	1.33
	Erosion	0	0	0	0	0	0	0	0	0	0	0.00
	Air	0	0	0	1	0	2	3	1	0	7	0.77
	Noise	1	1	1	2	1	2	3	0	2	14	1.55
BIOLOGICAL COMPONENTS	Diversity of flora	0	1	0	0	1	0	1	1	0	4	0.44
	Diversity of fauna	2	1	1	1	1	0	2	2	2	12	1.33
	Ornithofauna	2	1	1	1	1	0	2	2	2	12	1.33
	Chiropteran fauna	2	1	1	1	1	0	2	2	2	12	1.33
	Barriers/corridors	2	1	1	1	1	0	1	1	2	10	1.11
SOCIO-CULTURAL COMPONENTS	Landscape	2	2	1	2	1	1	1	3	2	15	1.66
	Land use	1	2	1	1	1	1	1	2	1	11	1.22
	Economy	0	0	0	0	0	0	0	0	0	0	0.00
	Cultural heritage	0	2	0	0	0	0	0	0	0	2	0.22
	Accidents	2	0	0	1	0	0	0	2	2	7	0.77
Cumulative values of IF according to environmental factors		15	14	8	12	9	8	18	18	16		
Average		0.93	0.87	0.50	0.75	0.56	0.50	1.12	1.12	1.00	IF = 0.82	

Fig. 14.6 The environmental impact matrix. (Source: Josimović et al. (2014). Reproduced by permission from the authors)

factors describe all the phases/components of the project and raw factors represent environmental elements. Every cell of the matrix records the impact of a specific aspect of the project on an environmental element. The checklist information needs to be converted into score information. For instance, in Fig. 14.5, for every environmental factor, 1 indicates no impact, 1 low impact, 2 tolerable impact, 3 medium high impact,4 high impact, and 5 very high impact.

The environment impact matrix enables a planner or a developer to assess such questions as (1) what is the overall impact of a project on the environment? (2) what environmental element is the most affected? and (3) what phase or component of the project has the most environmental impact? In the example of Fig. 14.5, the average environmental impact factor (IF) is 0.82, indicating an overall low impact from the project on the environment. Every row corresponds to the impacts on the specific environmental elements. Comparing among rows, one can conclude that the project has the most impact on landscape (with cumulative IF value 15 and average IF value 1.66), followed by noise (with cumulative IF 14 and average IF value 1.55). Every column presents the impacts from a specific component of the project. Comparing between columns, one can conclude that two phases, the operation of construction equipment and the waste material treatment, generate the highest level of environmental impacts (with cumulative IF value 18 and average IF value 1.12).

References

Berke, P. R., & Stevens, M. R. (2016). Land use planning for climate adaptation: Theory and practice. *Journal of Planning Education and Research, 36*(3), 283–289.

Bourbia, F., & Boucheriba, F. (2010). Impact of street design on urban microclimate for semi arid climate (Constantine). *Renewable Energy, 35*(2), 343–347.

Bowles, L. E. (1996). *Foundation analysis and design*. McGraw-hill.

Brabec, E., Schulte, S., & Richards, P. L. (2002). Impervious surfaces and water quality: A review of current literature and its implications for watershed planning. *Journal of Planning Literature, 16*(4), 499–514.

Campbell, S. (1996). Green cities, growing cities, just cities?: Urban planning and the contradictions of sustainable development. *Journal of the American Planning Association, 62*(3), 296–312.

City of Seattle. (2020). Citywide Planning. http://www.seattle.gov/Documents/Departments/OPCD/OngoingInitiatives/SeattlesComprehensivePlan/CouncilAdopted2020_CitywidePlanning.pdf.

Clarke, A. L. (2002). Assessing the carrying capacity of the Florida keys. *Population and Environment, 23*(4), 405–418.

Crompton, J. L. (2001). The impact of parks on property values: A review of the empirical evidence. *Journal of Leisure Research, 33*(1), 1–31.

Doss, C. R., & Taff, S. J. (1996). The influence of wetland type and wetland proximity on residential property values. *Journal of Agricultural and Resource Economics*, 120–129.

Erell, E., Pearlmutter, D., & Williamson, T. (2012). *Urban microclimate: Designing the spaces between buildings*. Routledge.

Galli, A., Kitzes, J., Wermer, P., Wackernagel, M., Niccolucci, V., & Tiezzi, E. (2007). *An exploration of the mathematics behind the ecological footprint* (pp. 249–256). Wit Press.

Global Footprint Network. (2011). Ecological Footprint Analysis San Francisco-Oakland-Fremont, CA. https://www.spur.org/sites/default/files/2013-10/Ecological_Footprint_Analysis.pdf

Josimović, B., Petrić, J., & Milijić, S. (2014). The use of the leopold matrix in carrying out the EIA for wind farms in Serbia. *Energy and Environment Research, 4*(1), 43–54.

Kemp, D. D. (2004). *Exploring environmental issues: An integrated approach*. Routledge.

Krause, D. (1993). Environmental consciousness: An empirical study. *Environment and Behavior, 25*(1), 126–142.

Marsh, W. M. (2010). *Landscape planning: Environmental applications* (5th ed.). Wiley.

Panagos, P., Borrelli, P., Meusburger, K., Alewell, C., Lugato, E., & Montanarella, L. (2015). Estimating the soil erosion cover-management factor at the European scale. *Land Use Policy, 48*, 38–50.

Rees, W. E. (1995). Achieving sustainability: Reform or transformation? *Journal of Planning Literature, 9*(4), 343–361.

Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., & Yoder, D. C. (1997). *Predicting soil erosion by water: A guide to conservation planning with the revised universal soil loss equation (RUSLE)*. U.S. Department of Agriculture, Agriculture Handbook No. 703.

Ryan, C. M., & Klug, J. S. (2005). Collaborative watershed planning in Washington state: Implementing the watershed planning act. *Journal of Environmental Planning and Management, 48*(4), 491–506.

Symeonakis, E., Karathanasis, N., Koukoulas, S., & Panagopoulos, G. (2016). Monitoring sensitivity to land degradation and desertification with the environmentally sensitive area index: The case of lesvos island. *Land Degradation & Development, 27*(6), 1562–1573.

The EPA. (2020a). *Criteria air pollutants*. Retrieved on Nov. 21, 2020 from: https://www.epa.gov/criteria-air-pollutants.

The EPA. (2020b). *Rainfall erosivity factor calculator for small construction sites*. Retrieved on Nov. 21, 2020 from: https://lew.epa.gov/.

The EPA. (2020c). *Technical assistance document for the reporting of daily air quality –the Air Quality Index (AQI).* Retrieved on Nov. 21, 2020 from: https://nepis.epa.gov/Exe/ZyPDF.cgi/P100FD3G.PDF?Dockey=P100FD3G.PDF

The LOJIC. (2020). Retrieved on Nov. 20, 2020 from: https://www.lojic.org/lojic-online

The USGS. (2020); Retrieved on Nov. 21, 2020 from: https://water.usgs.gov/wsc/map_index.html

Thomas, P. J., Baker, J. C., & Zelazny, L. W. (2000). An expansive soil index for predicting shrink–swell potential. *Soil Science Society of America Journal, 64*(1), 268–274.

URS Corporation. (2002). *Florida keys carrying capacity study*. https://www.monroecounty-fl.gov/DocumentCenter/View/84/2-Florida-Keys?bidId=

Wischmeier, W. H., & Smith, D. D. (1965). *Rainfall-erosion losses from cropland east of the Rocky Mountains, guide for selection of practices for soil and water conservation* (p. 282). Agriculture Handbook.

Chapter 15
Housing Analysis

Housing is one of the largest economic sectors in the society (Schwartz, 2015). Housing provides *shelters* to meet the basic needs of human beings, and therefore, requires high-quality standards to provide a safe, healthy, and quiet environment for people to rest and enjoy the home life. On the other hand, housing is often the largest expense of a household, and may provide certain investment benefits if housing price appreciates and housing transactions happen in a housing market. Housing costs comprise of two parts, land costs and structure costs. Land is a scarce resource and often has limited supply in urban areas, particularly in the *central business district* (CBD). Economic theory of supply and demand indicates that when the supply is lower than the demand, price increases to strive for the equilibrium. In general, land value has an uprising trend, although structures or buildings may depreciate or appreciate, depending on the condition of the structure or building. When the housing price increases and when there is a strong demand for housing, developers have more motivation to build more housing, and therefore the supply increases. When the supply surpasses the demand due to overbuilding, prices begin to decrease. The market then becomes vulnerable. When decreasing housing prices meet with certain trigger events, such as high unemployment rates, high inflation, or other conditions that make it difficult for mortgage borrowers to pay their mortgages, a housing crisis may occur and the crisis might trigger an overall economic recession.

In addition to its socioeconomic impact, housing also has profound environmental impact (Schwartz, 2015). Occupied housing needs drinking water, functional sewage and plumbing, solid waste disposal, electricity, or other types of fuels. All these generate a significant amount of carbon footprint. To mitigate the negative environmental impact of housing, developers, home goods producers, and planners alike strive to encourage energy efficient construction materials and home appliances. On the other hand, recycling and composting of wastes are encouraged by municipal service departments. Regardless, the environmental impact of housing is

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-030-93574-0_15].

Y. Li, S. Zhang, *Applied Research Methods in Urban and Regional Planning*,
https://doi.org/10.1007/978-3-030-93574-0_15

significant and low and zero carbon technologies are constantly explored to reduce such impact (Palmer et al., 2006).

Housing is usually less affordable in large cities and metropolitan areas, and in places where there is a strong demand, but a limited supply of housing. Therefore, housing price is not ubiquitously the same across places. The traditional "location, location, and location" of housing dictates differing housing prices. In places where housing is less affordable, providing affordable housing units to households or families who cannot afford market-rate housing thus becomes one of the priorities of urban planning.

This chapter introduces the analysis of housing market conditions, with a focus on housing needs assessment, and then introduces some important housing assistance and subsidy programs. Fundamentals of pro-forma analysis and cash flow analysis of real estate projects was covered in Chap. 13 Evaluation Research, and therefore, will not be covered in detail in this chapter.

Land Economics

The primary resource of housing construction is land. Therefore, it is imperative to introduce the demand and supply of land. ***Supply of land*** is inelastic, since land is scarce. When demand increases from population growth or in-migration, supply of land will not increase correspondently; on the contrary, the supply of land becomes scarcer when demand rises. Based on the basic economic theory of supply and demand, when demand is high while supply is limited, land price increases. Equilibrium is difficult, or impossible, to achieve regarding the supply of habitable land. This fundamental theory of supply and demand of land, therefore, dictates the spatial distribution of housing prices in different locations. Location of land, housing attributes, and socioeconomic characteristics all determine housing price, although certain factors have more influence than others. Multivariate regression models use equations with multiple independent variables to project or explain housing prices, and these models are called *hedonic housing price models*. The equation can be expressed as below (Ottensmann, Payton, & Man, 2008):

$$P = \beta_0 + \beta_H H + \beta_N N + \beta_L L + \varepsilon$$

Where P is a vector of house prices, H is a matrix of house characteristics, N is a matrix of neighborhood characteristics, and L is a matrix of one or more location characteristics. β_0 is the constant term vector, β_H, β_N, and β_L are matrices of the corresponding parameters, and ε is a vector of error terms. Housing prices are usually skewed, meaning not normally distributed, and therefore the logarithmic form (log) of housing prices is used as P. This equation is the most used hedonic housing price models by researchers and organizations. The model indicates that factors determining housing price are house characteristics (such as size, number of bedrooms, number of bathrooms, presence of fireplaces, etc.), neighborhood characteristics (such as median household income, racial and ethnic composition, housing value, vacancy rates, etc.), and locational characteristics (such as distance from parks, distance from the Central Business District, distance from the beach, etc.). Data for housing characteristics are often from local property appraisers' offices, neighborhood characteristic data are mostly from the Census Bureau, and the locational data are from GIS distance measures. Property parcel data and locations of major attractions or features from local appraisers' office and local GIS databases are combined with demographic data in ArcGIS, and then distance of property centroid from a location (depending on the research question) is measured and integrated into the dataset. Prepared and merged data are then imported into statistical analysis software for processing. In running multiple regressions it is imperative to make sure there is no significant and strong correlation among all the independent variables.

Because land supply is scarcer in urban centers than in suburban and exurban areas, it is natural that land price is higher in central business districts (CBD). The further away from the CBD, the more affordable land is. It is, therefore, not surprising that warehouses tend to locate in areas with cheaper land. Decreasing price of land in urban areas is explained by the *bid-rent theory* and *the monocentric land use model*, as formulated by Alonso (1964) and Muth (1969):

Different land uses respond differently to the decreasing land price from CBD, and therefore, spatial distribution of land uses differ. However, land use distribution is much more complex than as illustrated in Figs. 15.1 and 15.2, especially that many contemporary metropolitan areas are polycentric, meaning with multiple centers.

In choosing the best location to develop housing, developers and planners need to consider various factors, such as local housing needs, financial viability, and community impact. Once the site is determined, site feasibility analysis to analyze physical factors, legal factors, political factors, and off-site factors is imperative. Maps, drawings, land surveys, topography, soil tests, water tests, etc. are the foundations of site suitability and feasibility analysis. Detailed analysis of sites can be found from Chap. 11 Evaluation Research and Chap. 12 Urban Land Use Analysis.

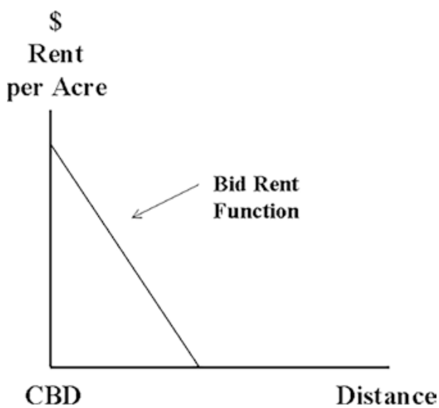


Fig. 15.1 Bid-rent theory

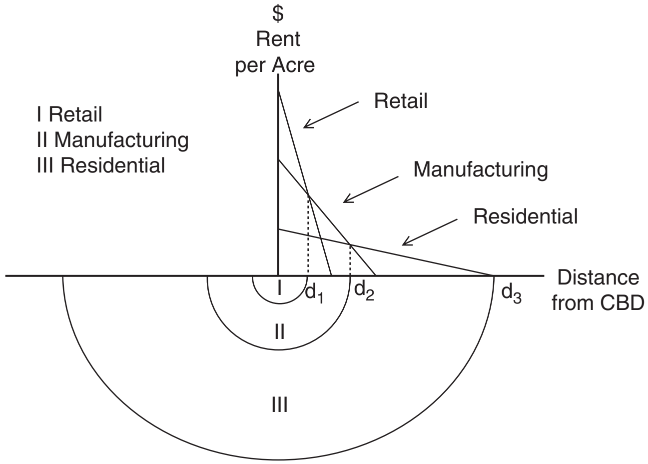


Fig. 15.2 Bid-rent theory and land use

Housing Needs Assessment

Housing Supply Analysis

The ***supply of housing*** is the total current or estimated future housing stock for an area or a housing market. For example, based on Quickfacts at the U.S. Census Bureau, the total number of housing units in Palm Beach County, Florida, in 2019 was estimated as 693,140. Among these housing units, 68.6% are owner occupied, and 31.4% are renter occupied (Table 15.1).

However, simply looking at the total units by tenure is not sufficient to estimate the supply characteristics of housing in an area. Other characteristics, such as housing units by affordability levels, may be more useful to gauge the overall affordability of housing stock. Figure 15.3 indicates affordable and available rental units per 100 renters in the West Palm Beach-Boca Raton Metropolitan Statistical Area. The chart indicates that there is a strong shortage of affordable housing for households making less than 120% AMI (Area Median Income) (Fig. 15.4).

When looking at overall surplus or deficit of affordable/available units by income, it is even more evident that housing units, regardless of ownership status, are in shortage for those making less than 120% AMI (Fig. 15.4). Figure 15.4 uses red to indicate shortage as "in the red" (budget deficit in accounting), and uses black to indicate over supply as "in the black" (profitable in accounting).

Vacancy and ***occupancy status***, and ***substandard housing*** are also some of the factors that need to be considered while analyzing housing supply. Data such as these can be easily obtained from housing statistics compiled by various agencies, such as the American Community Survey of the U.S. Census Bureau.

Table 15.1 Census quickfacts of housing statistics in Palm Beach County, FL

Housing	
Housing units, July 1, 2019, (V2019)	693,140
Owner-occupied housing unit rate, 2014–2018	68.6%
Median value of owner-occupied housing units, 2014–2018	$264,400
Median selected monthly owner costs -with a mortgage, 2014–2018	$1765
Median selected monthly owner costs -without a mortgage, 2014–2018	$667
Median gross rent, 2014–2018	$1320
Building permits, 2019	5550

Source: U.S. Census Bureau, Census QuickFacts, Palm Beach County, FL

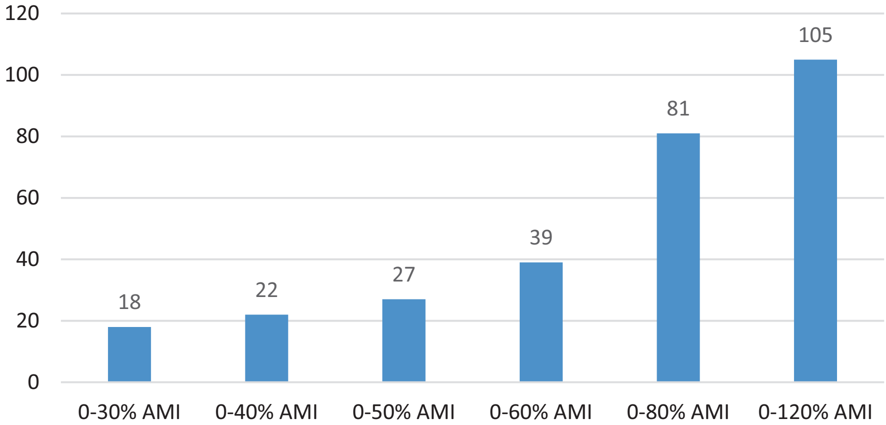


Fig. 15.3 Affordable and Available Rental Units, West Palm - Boca Raton MSA, 2014–2018. (**Source:** Shimberg Center for Housing Studies analysis of 2018 American Community Survey PUMS (Public Use Microdata Sample))

Housing is classified into various types but mostly housing covered in this chapter is residential, providing homes for residents. Commercial and office building analysis uses different datasets, but the fundamentals of supply and demand apply to all types of buildings and market analysis. Residential housing can be divided into different types, for example, single-family housing and multi-family housing. ***Single-family housing*** is housing with one unit on one lot/parcel of land, while ***multifamily housing*** includes multiple conjugating units, for example, duplex for two units, triplex for three units, and condominiums for multiple units, on one parcel of land. Counties or parishes in the U.S. often maintain a property parcel database, including the main characteristics of all parcels, such as uses, number of rooms, number of bedrooms, building size, lot size, appraised value, sales data and value, transaction date, taxation, etc. Table 15.2 indicates different types of

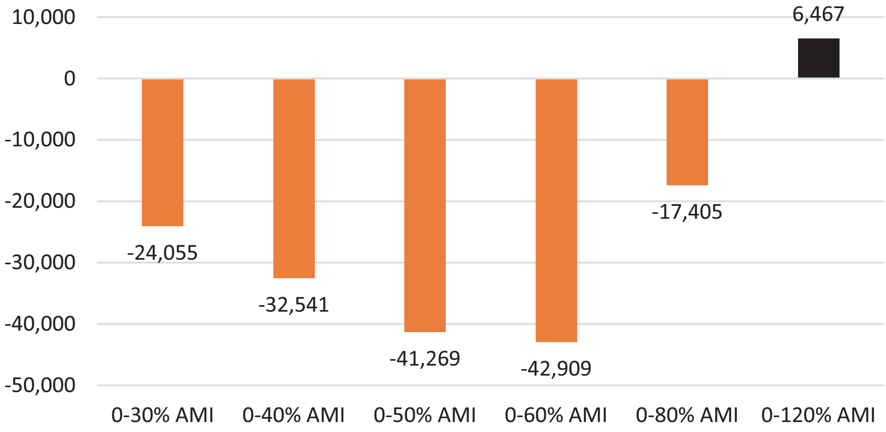


Fig. 15.4 Surplus/Deficit of Affordable/Available Units by Income, West Palm – Boca Raton MSA, 2014–2018. (Source: Shimberg Center for Housing Studies analysis of 2018 American Community Survey PUMS)

Table 15.2 Residential land use per Florida Department of Revenue

Land use codes	Specific uses
000	Vacant residential – with/without extra features
001	Single family
002	Mobile homes
003	Multi-family – 10 units or more
004	Condominiums
005	Cooperatives
006	Retirement homes not eligible for exemption
007	Miscellaneous residential (migrant camps, boarding homes, etc.)
008	Multi-family – Fewer than 10 units
009	Residential common elements/areas

Source: Florida Department of Revenue

residential housing based on Florida Department of Revenue land use code classifications (Florida Department of Revenue, 2018).

To understand the current supply of housing, it is imperative to have a clear inventory of the total housing units in a market. The total housing units include existing housing and new construction. Analyzing the supply also requires detailed breakdown of housing by type, tenure status (ownership and occupancy status), housing value (to better measure affordability), housing cost burden, among others. Housing cost burden is the ratio between total monthly housing cost and gross household income. If owning a house and paying mortgages, housing costs include mortgage principal repayment (P), interest rate (I), property tax (T), homeowners' insurance (I), utilities such as water, electricity, sewer, and waste management, homeowners' association fees, and maintenance and repairs. The first four types of costs are also called PITI. For renters, housing costs are usually rents and sometimes certain utilities. Data for housing units by income levels can be obtained from the Census Bureau in the U.S., and data for housing units by structural type can be from the Census Bureau and local property appraisers.

Provision of Affordable Housing

Housing affordability is a relative concept since households have different risk tolerance levels. However, for policy and planning purposes housing affordability is defined as households not paying more than 30% of their gross income on housing costs. For example, if the gross income of a household is $70,000, which is the income before taxes and before expenditures, the monthly income is about $5833.33. Therefore, the household should not pay more than 30% of $5833.33, which is $1750, in housing costs. Housing costs vary for different households. For renters, housing costs are usually rent and utilities, when certain utilities are not included in the rent. For homeowners with a mortgage, monthly housing costs are PITI, and sometimes mortgage insurance if downpayment is less than 20% of the purchasing price for conventional mortgages, maintenance and repairs, association fees, improvement expenses, etc. Homeowners without a mortgage do not need to pay for any mortgage related expenses, and property insurance is not mandatory. Therefore, for a household with $70,000 gross income, more than $1750 housing costs means the household is cost-burdened. If the household pays more than 50% of the monthly income, which is $2916.67, the household is severely cost-burdened.

In a market where housing costs are high compared to income, shortage of affordable housing, especially for those making less than 80% of the *Area Median Income* (AMI), i.e. low income households and families, is often an issue. Affordable housing, however, often extends up to housing for those making 120% AMI. Similar to public goods where market often fails, provisions of low income and affordable housing also needs government intervention in order to be successful. There are various supply and demand side programs to help provide affordable housing. From the supply side, the largest supply side programs of low income and affordable housing in the U.S. are the *Low Income Housing Tax Credit* (LIHTC) program and the public housing programs. Other supply side programs provide smaller portions of units. However, local communities also use regulatory stipulations and incentives for developers or landlords to provide affordable housing, for example, inclusionary zoning techniques, density bonus, accessory dwelling unit regulations, or micro housing ordinances. Nonprofit agencies are also some of the major players providing affordable housing, often with the assistance from foundations and government grants. The largest demand side affordable housing program is Section 8 Housing Choice Vouchers, where the federal government provides the gap between what a household can afford and the fair market rent of a locality. Affordable housing programs differ in different countries, for example, social housing is often the focus in Europe and government sponsored housing units were prevalent in Russia, China, and Singapore, before housing reforms to transition to market rate housing. Regardless of various programs and efforts in providing affordable housing, affordable housing shortage in hot markets, such as New York City, Tokyo, or London, continues to be a major public policy issue.

Fair Market Rent (FMR) and Income Limits

Fair Market Rent (FMR) is the payment standards set up by HUD for its tenant-based assistance programs, such as Section 8 Housing Choice Vouchers. FMR used to be determined by HUD at county levels, but in 2016 HUD began to experiment small area FMR based on zip code. FMR is the fair market rent associated with the number of bedrooms in the county or zip code level. For example, in Palm Beach County, FMR for 10 of the 73 zip codes for Fiscal Year (FY) 2021 is shown in Table 15.3.

Income Limits are the income standards to determine eligibility of tenant-based assistance programs. Income Limits are also determined by HUD at county levels. They are based on the size of households and come with different levels such as the income levels for 0 to 30% AMI (Area Median Income), 30% to 50% AMI, 50% to 80% AMI, and 80% to 120% AMI. Usually households with income higher than 120% are not qualified for public assistance. Table 15.4 provides an example of income limits for FY 2019 in West Palm Beach – Boca Raton Metropolitan Statistical Area (MSA), where the Palm Beach County is located. The FY 2019 AMI was $75,400.

Table 15.3 Small are FMRs for Palm Beach County, FL (FY 2021), example of 10 zip codes

ZIP code	Efficiency	One-bedroom	Two-bedroom	Three-bedroom	Four-bedroom
33,401	$940	$1130	$1410	$1910	$2290
33,402	$980	$1180	$1470	$1990	$2390
33,403	$870	$1050	$1300	$1760	$2110
33,404	$800	$960	$1200	$1620	$1950
33,405	$880	$1050	$1310	$1770	$2130
33,406	$1000	$1210	$1500	$2030	$2440
33,407	$840	$1010	$1260	$1700	$2050
33,408	$1000	$1200	$1490	$2010	$2420
33,409	$960	$1150	$1430	$1930	$2320
33,410	$1150	$1380	$1720	$2320	$2800

Source: The U.S. Department of Housing and Urban Development

Table 15.4 Annual income limits for West Palm Beach – Boca Raton MSA

Number of persons in household	Extremely low income (30%)	Very low income (50%)	Low income (80%)	Moderate income (120%)	Moderate income (140%
1	$ 17,600	$ 29,300	$ 46,850	$ 70,320	$ 82,040
2	$ 20,100	$ 33,450	$ 53,550	$ 80,280	$ 93,660
3	$ 22,600	$ 37,650	$ 60,250	$ 90,360	$ 105,420
4	$ 25,750	$ 41,800	$ 66,900	$ 100,320	$ 117,040
5	$ 30,170	$ 45,150	$ 72,300	$ 108,360	$ 126,420
6	$ 34,590	$ 48,500	$ 77,650	$ 116,400	$ 135,800
7	$ 39,010	$ 51,850	$ 83,000	$ 124,440	$ 145,180
8	$ 43,430	$ 55,200	$ 88,350	$ 132,480	$ 154,560
9	Refer to HUD	$ 58,520	$ 93,632	$ 140,448	$ 163,856
10		$ 61,864	$ 98,982	$ 148,474	$ 173,219

Source: Palm Beach County Government. https://discover.pbcgov.org

Low Income Housing Tax Credit (LIHTC) Program

Tax credits have been used by various government entities incentivizing individuals or businesses to increase financial assets, stimulate consumption, attract new businesses, or stabilize existing businesses. Tax credits work differently from tax deductions. Tax credits directly reduce the tax liabilities while tax deductibles reduce the income or profit bases that are used to calculate tax liabilities. For example, one agency owes $10,000 taxes to the taxing authority; however, the agency received a $1000 tax credit from the taxing authority. The total taxes to be paid now are $9000, $1000 less than the initial tax liability. Tax abatement and incentives are widely used financial incentives in local economic development (Bartik & Eberts, 2011).

To combat the decreasing public housing stock and increase housing affordability, Tax Reform Act of 1986 authorizes the ***Low Income Housing Tax Credit*** (LIHTC) program to help provide affordable housing. Since the inception of the program, LIHTC has provided over two million low income housing units and is so far the largest affordable housing program in the U.S. The tax credit can be 4% or 9%, and may shift away from 4% or 9% depending on market conditions and federal mandates. Land acquisition costs are not eligible to be considered when calculating total LIHTC credits. If a LIHTC is located in a difficult to develop area, meaning areas with high housing costs relative to the income levels, or in a Qualified Census Tract (QCT), the basis used to calculate the credit will receive a 130% boost. A Qualified Census Tract (QCT) is a census tract where at least 50% of the households have income below 60% of the AMI, or at least 25% poverty rate. The LIHTC credit lasts for 10 years and the property must be occupied by low-come households (usually less than 60% of the AMI) for at least 15 years, which raises the question of sustained affordability after 15 years. The amount of the credit depends on the cost and location of the LIHTC project. Each state qualifies for a total amount of credit per state population and at least 10% of the state LIHTC credit has to go to nonprofit organizations. At the state level, Housing Finance Agencies (HFAs) are the entities to administer and supervise the LIHTC program.

The following is the fundamental steps used to calculate LIHTC credits.

Step 1: Eligible Basis = Development costs – land costs – certain other costs
Step 2: Qualified Basis = Eligible basis × % low income units
Step 3: Basis Boost = Qualified basis × 130% (if the project is located in a “difficult development area” or a “qualified census tract”)
Step 4: Annual Tax Credit = Basis boost or Qualified basis x credit rate (9% or 4%)
Step 5: 10-year Total Tax Credits = Annual tax credit × 10

An example to demonstrate the calculation is as follows:

All 100 units of a project (100%) are for low-income families. The project costs $ten million to build, of which $1.6 million is for land and other costs ineligible for the LIHTC credits. The project is located in a difficult development area.

1. Eligibility basis = 10–1.6 = $8.4 million
2. Qualified basis = 8.4 × 100% = $8.4 million
3. Basis boost = 8.4 × 130% = $10.92 million
4. Annual credit = 10.92 × 9% = $982,800
5. 10-year total credit = 982,800 × 10 = $9,828,000

Once the developers receive the tax credit they can convert them into equity to pay for development costs. Organizations who purchase the credits use them as a tax shelter, reducing their tax responsibilities. The transaction of tax credit depends on demand and supply of the credit and therefore price varies by regions and by time. In a stable or booming economy, there is often more demand of the tax credit than during the bust or declining years. Therefore, $1 tax credit may be sold at a premium (higher than $1) in a bull market but may be sold at a discount (lower than $1) in a bear market. For example, if a developer receives $9,828,000 LIHTC credit and the tax credit is sold to a bank at $0.89 per credit, then the developer will receive $8,746,920 to be used in the development of the project.

$$\$9,828,000 \times 0.89 = \$8,746,920$$

When the converted equity is not sufficient for the development project, sometimes gap financing is provided by state and local governments through housing trust funds and block grants, or through nonprofit organizations that are interested in promoting affordable housing. State housing trust funds are often from real estate transaction fees, other types of fees, or donations. Major block grants that are often used in housing and community development are the Community Development Block Grant (CDBG) and HOME Partnership grant. The gap in sources of funds may be filled with debt as well, when developers borrow from financial institutions or other entities, paying an interest rate for the borrowing.

LIHTC is a major vehicle in providing affordable housing in the U.S. with greater flexibility than other housing subsidy programs. However, the program is complicated, and a substantial portion of the subsidy goes to transaction costs and investor profits. Other challenges include long-term affordability (after the expiration of the tax credit affordability period) and capital improvement in maintenance and upgrading dilapidated housing units. The program also offers minimum incentives to developers to build mixed-income housing, thus may contribute to the concentration of poverty. These are the same challenges facing many of the low-income and affordable housing programs. Another unique challenge to the LIHTC program is the volatility of equity conversion of tax credits. The volatile market may make it unstable for developers to receive adequate equity to be used in the development. Nevertheless, the LIHTC program has been a very important program in providing affordable housing in the U.S. and hopefully policies overcoming the challenges will help maintain its important role in the affordable housing market.

Public Housing Projects

Public housing is usually provided by the public sector and either rented to the resident at a low cost or no cost. Public housing stocks in different counties vary due to different housing policy outcomes. Data sources of public housing stocks are maintained by the public housing agencies. Eligibility threshold of staying in public housing projects are usually lower than other types of tenant-based housing assistance programs. In the U.S., income eligibility is usually from 0 to 60% per household size in different metropolitan statistical areas. Prior studies on public housing mostly focused on evaluating the health, economic, and emotional outcomes of the occupants (e.g. Newman & Harkness, 2001; Varady & Preiser, 1998), and strategies to improve management efficiency and policy outcomes of public housing programs (e.g. Goetz, 2012; Schill, 1993).

In the U.S., the public housing program was originated in 1937 and has gone through various reforms over the years. Public housing programs are usually managed by *Public Housing Agencies* (PHAs) and *Housing Authorities* throughout the country. Most public housing projects are rather small, and the projects are mostly financed through issuing bonds to be sold to the investors. Rental income and operating subsidies were used to cover operating expenses. Due to the long history and the lower quality of housing in the public housing program, after the 1980s most of the efforts have switched from the expansion of the program to preservation and redevelopment. The costs of maintenance and repair prove high. The Quality Housing and Work Responsibilities Act of 1988 established HOPE VI program to transform distressed public housing stock, usually high-rise towers, into smaller, mixed-income new urbanist communities. HOPE VI did not require one-for-one replacement of the housing units, resulting the loss of public housing stock after demolition. To overcome this limitation, the Choice Neighborhoods program established during the Obama administration requires one-for-one replacement of any demolished public housing units. Reforms of public housing are evolving and overcoming various challenges will help ensure the sustained operation of the public housing stock.

Section 8 Housing Choice Voucher Program

Another notable tenant-based assistance program in the U.S. is the **Section 8 Housing Choice Voucher** Program, which often nicknamed as Section 8 Housing. Section 8 Housing Choice Voucher program has evolved drastically over the years, shifting from project-based to mostly tenant-based, offering vouchers to qualified tenants so that they are able to relocate anywhere they wish if the vouchers are accepted by the landlords. The voucher pays the difference between the rent charged by the landlord and the rent the household is able to afford, usually between 25% to 30% of the monthly household income. However, the rent charged by the landlord cannot go over the Fair Market Rent (FMR) standards set by HUD in local areas, per the number of bedrooms, and excessive portion of the rent may not be covered by the voucher program.

The following is an example to calculate the rent split between the tenant and the Housing Authority. Suppose the tenant's income is $12,000 a year and the AMI for the area that she lives in is $65,000. The tenant is, therefore, in the extremely low income bracket (less than 30% of the AMI). She is a single mother with two young children, one boy and one girl, and she receives the Section 8 voucher. She is qualified for a three-bedroom unit wherever her family wishes to live. Through a real estate agent, she found a three-bedroom single-family home, and the monthly rent is $2100, exactly at the Fair Market Rent level of the zip code for a three-bedroom unit. Her monthly income is $1000, 30% of which is $300. Given her unique situation, the maximum affordable rent for her is $300. In this case, the Housing Authority will need to pay at least $2100–$300 = $1800 to the landlord and the tenant will be responsible for a maximum of $300. If she needs to pay utility costs herself, her own rent responsibility might be lower than $300. The subsidy from HUD, administered by local housing authorities, therefore, makes it possible for low income residents to find decent, sanitary, and affordable homes. Tenant income and eligibility is reexamined every 6 months to ensure fairness of the program. Annual inspections of the property are conducted to maintain housing standards for tenants using the vouchers.

The Section 8 Choice Voucher program is one of the largest affordable housing programs in the U.S. regardless of the long waiting list of households to be qualified. The lottery system to pick households within the income limits makes it a chance game for many families or households in need of sanitary and high-standard affordable housing. Once the qualified households receive the voucher the households have a limited amount of time to find a residence where the voucher is accepted. One of the purposes of the voucher program, specifically the Moving to Opportunity initiative, is to deconcentrate poverty; however, studies have found that

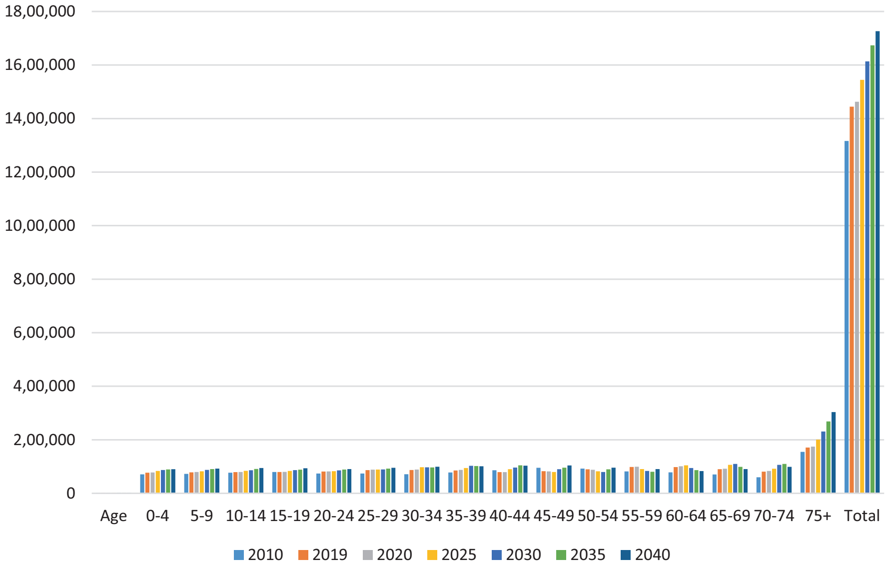


Fig. 15.5 Projected population by age cohort, permanent residents, in Palm Beach County. **Notes:** Counts refer to estimated or projected number of persons. (Source: Estimates and projections by Shimberg Center for Housing Studies, based on 2000 and 2010 U.S. Census data and population projections by the Bureau of Economic and Business Research, University of Florida)

the Moving to Opportunity initiative has only achieved very mild results of poverty deconcentration (Walter, Li, & Atherwood, 2015).

Housing Demand Analysis

The providers of housing units, regardless of private real estate developers, or public sectors, often need to understand housing demand. ***Housing demand analysis*** involves around demographic analysis, particularly current and future population and household estimation and projection. If affordability is a central focus, household projection needs to consider income levels as well. The following charts and tables provide some examples of population and household projection to help project the future demand in housing (Fig. 15.5, Tables 15.5, 15.2, 15.3, 15.4, 15.5, 15.6, 15.7, 15.8, and 15.9).

Demographic analysis is then compared to housing supply analysis to configure the gap in housing supply. One of the examples of conducting housing supply and demand analysis include the Comprehensive Housing Affordability Strategy (CHAS) and the Consolidated Planning process defined by the U.S. Department of Housing and Urban Development.

Table 15.5 Household projection by income levels, owner-occupied units, moderately cost-burdened (housing cost is greater than 30% income but less than or equal to 50%)

Household income	2010	2019	2020	2025	2030	2035	2040
Less than or equal to 30% of AMI	4625	5208	5284	5627	5944	6208	6438
Greater than 30% but less than or equal to 50% of AMI	10,784	12,141	12,319	13,120	13,858	14,473	15,009
Greater than 50% but less than or equal to 80% of AMI	13,182	14,841	15,058	16,037	16,939	17,691	18,346
Greater than 80% but less than or equal to 100% of AMI	7530	8478	8602	9161	9676	10,106	10,480
Greater than 100% of AMI	16,489	18,565	18,837	20,062	21,190	22,130	22,950

Data Source**:** Estimates and projections by Shimberg Center for Housing Studies, based on U.S. Department of Housing Development, Comprehensive Housing Affordability Strategy (CHAS) dataset and population projections by the Bureau of Economic and Business Research, University of Florida.

Table 15.6 Household projection by income levels, renter-occupied units, moderately cost-burdened (housing cost is greater than 30% income but less than or equal to 50%)

Household income	2010	2019	2020	2025	2030	2035	2040
Less than or equal to 30% of AMI	2009	2205	2233	2345	2426	2502	2579
Greater than 30% but less than or equal to 50% of AMI	9694	10,641	10,779	11,317	11,707	12,078	12,447
Greater than 50% but less than or equal to 80% of AMI	16,284	17,875	18,106	19,010	19,666	20,288	20,908
Greater than 80% but less than or equal to 100% of AMI	6681	7334	7429	7800	8069	8324	8579
Greater than 100% of AMI	5228	5738	5813	6103	6313	6513	6712

Data Source**:** Estimates and projections by Shimberg Center for Housing Studies, based on U.S. Department of Housing Development, Comprehensive Housing Affordability Strategy (CHAS) dataset and population projections by the Bureau of Economic and Business Research, University of Florida.

Table 15.7 Household projection by income levels, owner-occupied units, severely cost-burdened (housing cost is greater than 50%)

Household income	2010	2019	2020	2025	2030	2035	2040
Less than or equal to 30% of AMI	18,856	21,230	21,540	22,941	24,231	25,306	26,244
Greater than 30% but less than or equal to 50% of AMI	11,810	13,297	13,492	14,369	15,177	15,850	16,437
Greater than 50% but less than or equal to 80% of AMI	9887	11,132	11,295	12,029	12,706	13,270	13,761
Greater than 80% but less than or equal to 100% of AMI	2913	3280	3328	3544	3744	3910	4055
Greater than 100% of AMI	3133	3527	3579	3812	4026	4205	4360

Data Source**:** Estimates and projections by Shimberg Center for Housing Studies, based on U.S. Department of Housing Development, Comprehensive Housing Affordability Strategy (CHAS) dataset and population projections by the Bureau of Economic and Business Research, University of Florida.

Table 15.8 Household projection by income levels, renter-occupied units, severely cost-burdened (housing cost is greater than 50%)

Household income	2010	2019	2020	2025	2030	2035	2040
Less than or equal to 30% of AMI	22,782	25,007	25,331	26,595	27,513	28,384	29,252
Greater than 30% but less than or equal to 50% of AMI	15,555	17,074	17,295	18,158	18,785	19,380	19,972
Greater than 50% but less than or equal to 80% of AMI	6163	6765	6853	7195	7443	7679	7914
Greater than 80% but less than or equal to 100% of AMI	844	926	938	985	1019	1051	1083
Greater than 100% of AMI	436	478	484	509	526	543	559

Data Source**:** Estimates and projections by Shimberg Center for Housing Studies, based on U.S. Department of Housing Development, Comprehensive Housing Affordability Strategy (CHAS) dataset and population projections by the Bureau of Economic and Business Research, University of Florida.

Table 15.9 Housing cost burden by income (2012–2016, Miami, FL)

Income by cost burden (Renters only)	Cost burden >30%	Cost burden >50%	Total
Household income ≤30% HAMFI	30,955	26,395	41,505
Household income >30% to ≤50% HAMFI	19,475	8545	22,170
Household income >50% to ≤80% HAMFI	11,605	2350	19,785
Household income >80% to ≤100% HAMFI	2630	565	8200
Household income >100% HAMFI	3380	260	20,595
Total	68,045	38,115	112,260
Income by cost burden (owners only)	Cost burden >30%	Cost burden >50%	Total
Household income ≤30% HAMFI	6105	4805	8965
Household income >30% to ≤50% HAMFI	4400	2990	7505
Household income >50% to ≤80% HAMFI	3740	1845	8055
Household income >80% to ≤100% HAMFI	1295	370	4085
Household income >100% HAMFI	3305	815	20,735
Total	18,845	10,825	49,345

Data Source: The U.S. Department of Housing and Urban Development. https://www.huduser.gov/portal/datasets/cp.html

The Gap: Housing Affordability Analysis

There are different ways analyzing the affordable housing market. The first is the ***housing affordability indices*** created by different organizations. *Housing Affordability Index (HAI)*, calculated by the National Association of Realtors (NAR) in the U.S., measures whether a typical family earns enough income to qualify for a mortgage loan on a typical home at the national and regional levels, based on the most recent price and income data. For example, HAI is calculated as

$$\text{HAI} = \left(\text{median family income} / \text{qualifying income}\right) * 100$$

If the median family income for an area is $54,115, and the qualifying income to get a mortgage to buy a typical home is $39,456, then

$$\text{HAI} = (\$54{,}115 / \$39{,}456) / 100 = 137.2$$

Therefore, the higher the HAI, the more affordable an area is.

Another index, for example, *Housing Opportunity Index* (HOI), is created by National Association of Home Builders. Housing Opportunity Index is defined as the share of homes sold in a given area that would have been affordable to a family earning the local median income, based on standard mortgage underwriting criteria. Therefore, the higher the share of affordable homes (HOI) the more affordable the area is.

A third index, *Housing and Transportation Affordability Index* created by the Center for Neighborhood Technology (CNT), combines measures of housing and transportation affordability, since for a typical household transportation is oftentimes the second largest expenses following housing. The rule of thumb is that the combined housing and transportation costs should not be higher than 45% to ensure the household is not cost-burdened in housing and transportation.

The second method to measure housing affordability is to ***calculate the demand gap of affordable housing*** in an area. This method is more useful if the housing market is limited within the boundary of the area and it is less meaningful if the housing market is open. Therefore, calculating the affordable housing gap in a metropolitan area makes more sense than calculating the gap within a specific locality within the metropolitan area since the labor market is often fluid and workers commute among different jurisdictions for jobs. Available employment opportunities often have a mismatch with the housing affordable to the workers in that area. For example, in a wealthy city where housing price is high, service and retail workers working in the city will not be able to afford to live close to employment. Situations like this are called *job-housing imbalance*, where there is not enough affordable housing for workers within an area. Job-housing imbalance creates its own problems, such as excessive commuting and increased infrastructure and environmental pressure. Urban sprawling may be one of the most notable consequences of job-housing imbalance.

When calculating the supply and demand gap of affordable housing, the focus is on projecting or calculating those households making less than 120% of the AMI (Area Median Income), matching the housing units affordable to different income levels, and then identify the gaps. The following table presents a framework of calculating the gap (Table 15.10):

Calculation of the number of households and housing units is presented in the sections of housing demand and supply analyses. Data sources are usually the demographic and housing census.

As mentioned earlier, for a closed housing market, calculating demand gap of affordable housing units is a plausible method to measure housing affordability issues in a local market. However, in an open market, where households commute between different localities, demand gap for individual localities may not be the best method to measure locational affordability.

In an open market where households live and work in different localities, housing affordability can be measured using the gap in housing price/rent and households' affordability, based on income levels. For example, if the average rent for a two-bedroom apartment in an area is $1500, the required monthly income is at least $5000, and the annual income is $60,000. The Area Median Income (AMI) is $55,000. If a household makes less than 30% of the AMI (less than $16,500 a year) the household can only afford a two-bedroom unit for not more than $412.5 a month. There is a rent gap of $1087.5 ($1500–$412.5) for this household. Median income for all households at each income level and current average rent can be used as the calculation basis for affordability. A framework to calculate cost gap can be illustrated in Table 15.11.

Table 15.10 Affordable housing gap by income levels

Income level	Current/Projected # of households (Demand)	Current # of affordable housing units (Supply)	Affordable housing gap (Demand-supply)
<30% AMI			
30–50% AMI			
50–80% AMI			
80–120% AMI			
>120% AMI			

Table 15.11 Housing cost affordability gap calculation

Income level	Median household income/Month	30% of income	Current rent / Qualified income for a typical apartment or house ($)	Affordability gap (renters: 30% income – rent, or owners: median income – income qualified for a typical mortgage)
<30% AMI				
30–50% AMI				
50–80% AMI				
80–120% AMI				
>120% AMI				

This framework can be repeated based on household size and number of bedrooms to measure more specific locational affordability based on housing costs and income levels. For owners, median household income subtracting income qualifying for a typical mortgage to buy a typical house, instead of rent, will be used to calculate affordability gaps.

When making decisions on housing planning, planner and public policy makers should consider combining the demand gap at the metropolitan level and the housing cost affordability gap at the local level to come up with the best decisions. Giving the close relationship between housing and transportation, collaborations among different localities are critical to make housing planning successful, especially among trip origination and destination communities. In the U.S., employment flows are data derived from the decennial census prior to 2010, or from the American Community Survey (ACS) data after 2000. CTPP (Census Transportation Planning Products) derived from the census data provides detailed information about commuting patterns. An online application, Census on the Map (https://onthemap.ces.census.gov/), provides aggregated mapping and data needs for employment flows in different municipalities.

Box 15.1 Example #1 of a Housing Needs Assessment

Fundamentally a housing needs assessment should assess the housing demand, housing supply, and then identify the gap between the supply and the demand, focusing on the gap of affordable housing units for households with income lower than 120% AMI (Area Median Income). When conducting the assessment, prior studies for the locality, secondary data on demographics and housing, and possibly interviews and surveys as primary data sources should be analyzed to identify the housing needs. Studies may be structured differently per different localities and authors. For example, In *Montgomery County Housing Needs Assessment* (HR&A Advisors, Inc., 2020), two strengths of the report include: (1) Geographical trend of housing supply and demand, and the geographical distribution of cost-burdened households; (2) Future projection of housing needed annually per projected household growth. The report is structured as the following:

- **Background**. This section introduces who is conducting the study and what data and methods are utilized.
- **Geographic Study Area**. This section uses maps to delineate the study area.
- **Demographic Trends and Current Housing Needs**. This section describes the housing market conditions and demographic trends and affordability. Charts and tables are used to describe these indicators.

- – overall household and population growth
- – growth of housing units
- – trend of residential building permits
- – geographical locations of the highest household growth
- – tenure status
- – homeownership affordability (household income required to afford the median home value)
- – change in number of owner-occupied housing units by household income
- – change in number of owner-occupied housing units by age of householder
- – change in number of housing units by race by tenure
- – change in households per square mile by tenure
- – years lived in units by household income and tenure
- – change in number of households by income bracket
- – share of low-income households in the region
- – percentage households by income levels
- – income growth by geographical areas
- – per square mile gain and loss in rental units and in household earnings
- – trend of cost-burdened households
- – geographical areas with rising cost-burdened renter households
- – geographical areas with extremely cost-burdened renter households
- – housing supply gap by household income levels (at <30% AMI, 31%–50% AMI, 51–65% AMI, 66%–80% AMI, 81%–100% AMI, and 101% – 120% AMI)
- – housing supply gap by geographical areas
- – shift in renter households by household size
- – shift in renter-occupied housing units by number of bedrooms, unit structure, and unit size
- – number of renter households by housing unit size and household size to measure crowding conditions
- – shift in owner-occupied housing units by number of bedrooms, unit structure, and unit size

- **Future Housing Needs.** This section projects the future housing needs by household growth. Charts and tables are used to describe these indicators.
  - – household forecast and net new housing needed annually
  - – housing forecasts by type and tenure
  - – household forecasts by household income
  - – household forecasts by AMI segments
- **Appendix.** Any other tables and charts that are important but not essential are in this section.

Box 15.2 Example #2 of a Housing Needs Assessment

As mentioned in Box 15.1, housing needs assessment has different structures, but the fundamental principle is to analyze the housing supply, demand, and then identify the gap between housing supply and demand. Future trends should also be part of the assessments. Another report, *Broward County Affordable Housing Assessment* (The Metropolitan Center, Florida International University (FIU), 2018), presents slightly different structure than the report demonstrated in Box 15.1. One strength of this current report is that it includes the demographic and housing profiles of municipalities and unincorporated areas. The structure of this report is as follows:

- **Executive Summary**. This section summarizes the purpose, methodology, and key findings of the report.
- **Introduction and Methodology**. This section defines affordable housing, introduces the indices measuring affordability (National Association of Realtors (NAR) Index, Housing Opportunity Index, and Housing and Transportation Affordability Index), explains the link between economic growth and housing needs, and describes the methodology and the scope of the study.
- **Housing Supply Analysis**. This section analyzes the housing inventory by type, development and market trends, and single family, condominium, and rental markets.
  - inventory of single-family and multi-family units
  - owner and renter-occupied units
  - housing vacancy status
  - development trends (trend of building permits, new rental development activity)
  - single-family home market: median sales price, distressed market, and sales price distribution by sales price
  - condominium market: existing condominium units sold, distribution of sales price, new and existing units sold over time,
  - rental market: average monthly rent, average monthly rent by sub-market
  - home foreclosure activity: top 5 cities with the highest foreclosure rates
- **Housing Demand Analysis**. This section introduces the labor market and economic base, housing demand, and housing affordability and cost-burdens.
  - labor market and economic base: employment by industry
  - employment and housing demand
  - household composition and household income: household income, and housing cost-burden analysis

  - single-family market affordability analysis: affordability index by number of bedrooms (gap between median sales price and affordable home price at median household income), new single-family sales by major sub-market
  - condominium market affordability analysis: affordability for existing and new condominiums in major sub-markets
  - renter and owner market affordability analysis: rent trend, rent affordability by household income categories, owner affordability by household income categories

- **Future Housing Supply and Demand**. This section describes the industry and employment growth projections and the projected housing needs.
- **Conclusions**. This section concludes the shifts in housing demand and supply, growing housing affordability gaps, worker resident impacts, and housing and transportation costs.
- **Municipal/unincorporated Area Profiles and Housing Supply/Demand Analysis**
- **Appendix**
- **List of Tables**
- **Table of Figures**

Housing Finance

Analysis of the housing market should always focus on from the angles of supply versus demand, and housing finance is no exception. Housing finance from the demand and consumer side is mostly to measure housing costs, and whether the households are cost burdened when purchasing or renting the housing unit. The cost of owning a home is usually associated with the initial purchasing costs, maintenance costs, homeowners' association fees, costs of utilities, and improvement or remodeling costs. Maintenance costs, utility expenses, improvement costs, and other costs are usually easy to understand, but the mechanism of a mortgage, the loan used to purchase a house when needed, needs some explanations. For renters, the major costs are rents, and sometimes costs of utilities. Housing finance from the supply side, the provider of housing, is much more complex compared to household level housing financial analysis. Financial feasibility analysis, mostly the cash flow and pro-forma analysis, must be conducted before planning for a housing project. Then financing of the project needs to be secured, and later selling or renting the units will involve another layer of financial analysis. Housing finance is complicated, and the following section will provide a brief introduction to mortgages, rents, project financing, and basic cash flow and pro-forma analysis, to offer a glimpse of the basic concepts in housing finance.

Mortgages and Secondary Mortgage Market

So, what is a ***mortgage***? A mortgage is a legal term associated with borrowing money from a financial institution when purchasing a real estate property. A mortgage is a loan that the borrower secures to pay for the housing to be purchased. In the United States, between 62% to 68% of the homeowners have mortgages. Therefore, mortgages are critical to most people when purchasing a home. Traditionally, financial institutions typically loan the borrower 80% or less of the purchase price, so that the borrower has to leverage at least 20% of the purchase price. The 20% out-of-pocket funds is called down payment. The 80% loan is the outstanding mortgage balance when purchasing the house. Therefore, 80% is called *loan-to-value* (LTV) ratio. Different mortgage products or different borrowers may have different LTVs.

Since homeownership has proven to be beneficial to households, children's school performance, and neighborhood stability, many countries strive to promote homeownership. In order for many ordinary households to afford a home, various government policies have been enacted to achieve the goal of providing decent home for every household. Reforms in mortgage market brought lax underwriting standards during the 1990s in the U.S. Underwriting a mortgage means careful due diligence based on certain standards, such as income, credit scores, credit history, and assets, before issuing the mortgage. Relaxed standards include reduced down-payment, in turn a higher interest rate, and various "exotic" mortgages as defined by Schwartz (2015, p. 87–88):

- Various forms of adjustable mortgages with an initial "teaser rate" but then higher interest rate.
- Interest-only mortgages during an initial period of time, creating payment shock later on.
- Balloon mortgages with interest-only payment or low payment, followed by a lump-sum principal payment at the end of the mortgage term.
- Payment option mortgages where borrowers can request lower monthly payment, potentially creating payment shocks later on.
- Stated-income loans where the mortgage principal is based on stated-income of the borrower, instead of actual income.

Some of the exotic mortgages are also labeled as subprime loans, which are often characterized as much higher interest rates than prime mortgages. When exotic and subprime loans go overboard the practice becomes illegal, and called predatory lending. Subprime loans and predatory lending often target neighborhoods with concentrated low-income minority population. In order to promote lending, some lenders forfeit the necessity of verifying income and credit scores, and issued mortgages without necessary due diligence. The process is called green-lining of mortgages, in contrast to the illegal *red-lining* practice before the civil rights movements in the 1960s. Starting the 1930s mortgage lenders usually red-lined certain neighborhoods, often low-income minority neighborhoods, to prevent borrowers from these neighborhoods to obtain a mortgage. In the 1990s, the green-lining of mortgages targeted nearly the same type of neighborhoods. Red-lining perpetuates poverty concentration in certain areas, and deprives certain households of access to credit. Fair Housing Act of 1968, The Community Reinvestment Act of 1977, and Home Mortgage Disclosure Act of 1975 outlawed the practice of redlining; however, the effect of redlining in neighborhood decline and concentration of poverty lasts. *Green-lining* provides easy access of credit to low-income households and families; however, the effect of green-lining often causes families or households losing their homes through foreclosure. Mortgages under the practice of green-lining often are exotic mortgages or subprime loans, since lenders do need to consider risks when loaning to a borrower with lower income, and/or low credit scores. To compensate for the potential risks of delinquencies, meaning the borrower may not be able to pay the mortgage on time, lenders usually charge higher interest rates, and with special terms and conditions attached. When the mortgage payment is not received for more than a few times, the process is called default. When borrowers default, lenders have the right to take back the property and put it on market for auction or for sale. The foreclosure crisis during the Great Recession 2007–2009 was attributed to subprime loans and unemployment due to increased default risks.

Contrast to *subprime loans, prime mortgages*, or sometimes called conventional mortgages, are usually mortgages issued to borrowers with higher income and higher credit scores. Before the recent housing crisis between 2007 and 2009, conventional mortgages sometimes also offer high LTV loans, usually 90%, which means that the borrower only needs to pay 10% down payment. Lower down payment will require the borrowers to carry a mortgage insurance, to be added to the regular PITI (*Principal, Interest, Tax, Insurance*) payment. Here tax means property tax, and insurance means property insurance. Calculations of PITI can be easily carried out through Microsoft Excel formula PMT (rate, nper, pv, [fv], [type]), or an online calculator. The "rate" in the PMT function indicates mortgage interest rate, and needs to be divided by 12. For example, if the effective mortgage interest rate is 4.0% for a mortgage, the "rate" should be 4.0%/12 = 0.33%, which is 0.033. "nper" is the term of the mortgage in months. So if the term is 30 years, which is usually for most of the mortgages, "nper" will be 30 × 12 = 360. "pv" is the initial loan amount, and "fv" is the future value of the loan as 0. Since most mortgages are fixed-rate mortgage, meaning the interest rate remains the same throughout the duration of the mortgage, "type" does not need to be specified.

The monthly scheduled payment of principal payment will reduce the overall balance of the loan over time, although initially interest payment dominants the P + I payment but gradually decreases, while principal payment increases. Overall, principal decreases with each payment. The procedure of decreased principal balance is called amortization, which can be illustrated through Table 15.12 and Fig. 15.6. The amortization assumes a $300,000, 30-year, 4.25% interest rate mortgage originated in June, 2020. The table demonstrates part of the payment schedule for 2020 and 2021. The chart indicates the decreasing principal balance over the course of the mortgage.

Table 15.12 Amortization schedule of a mortgage ($300,000, 30-year, 4.25% interest rate, originated in June 2020) for the years of 2020 and 2021

Date	Interest	Principal	Balance
Jul, 2020	$1063	$413	$299,587
Aug, 2020	$1061	$415	$299,172
Sep, 2020	$1060	$416	$298,756
Oct, 2020	$1058	$418	$298,338
Nov, 2020	$1057	$419	$297,919
Dec, 2020	$1055	$421	$297,498
Jan, 2021	$1054	$422	$297,076
Feb, 2021	$1052	$424	$296,652
Mar, 2021	$1051	$425	$296,227
Apr, 2021	$1049	$427	$295,800
May, 2021	$1048	$428	$295,372
Jun, 2021	$1046	$430	$294,942
Jul, 2021	$1045	$431	$294,511
Aug, 2021	$1043	$433	$294,078
Sep, 2021	$1042	$434	$293,644
Oct, 2021	$1040	$436	$293,208
Nov, 2021	$1038	$437	$292,771
Dec, 2021	$1037	$439	$292,332

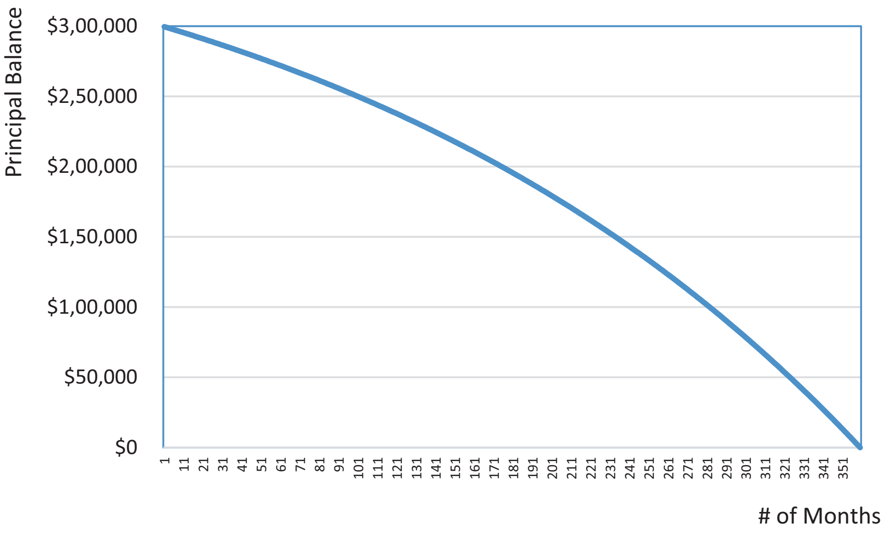


Fig. 15.6 Amortization schedule of a mortgage

In addition to subprime and prime mortgages, some other types of mortgages are often backed by government entities. A few notable examples in the U.S. include FHA-insured mortgages and VA-insured mortgages. FHA means Federal Housing Administration, which was created through the 1934 Housing Act. VA means Veterans Affairs. These two types of mortgages are backed by HUD (the U.S. Department of Housing and Urban Development) and VA (the U.S. Department of Veterans Affairs). With these two types of mortgages, there is usually a lower downpayment, higher interest rates, and therefore have higher default risks compared to conventional mortgages. Other programs, such as HomeStep and HomePath, are programs provided by Freddie Mac and Fannie Mae, when home purchasers purchase homes owned by these two government sponsored enterprises (GSE). These two programs sometimes only require as little as 3.5% downpayment, in turn a slightly higher mortgage interest rate compared to mortgages in conventional markets. A notable exception of these programs is that they are not subprime loans, and they do not require a mortgage insurance when downpayment is less than 20%.

Freddie Mac and Fannie Mae are two of the most important GSEs, along with Ginnie Mae. The major function of these organizations is to purchase mortgages from financial institutions, bundle them together, and sell these mortgage-backed securities, similar to stocks, mutual funds, futures, forwards, etc. to investors. Therefore, the market to securitize mortgages is called secondary mortgage markets. In this sense, the initial lenders do not own these mortgages anymore and they often act as a pass-through entity to pass your monthly payment to these three organizations. Then they will pass the gains or dividends to the investors in the secondary mortgage markets. Default risks are also passed through to investors. Due to the complicity of the secondary mortgage market, this book will not go into more details about the financial calculations associated with these mortgage-backed securities.

Project Financing

The basic financing mechanism for any privately owned projects is debt financing, equity financing, or hybrid financing. However, many nonprofit private entities also receive grants as one of the financing sources, which is often the case in providing affordable housing. Chap. 11 explains financial feasibility where the NPV (Net Present Value) of the case flow of a project needs to be at least zero (breakeven) or greater than 0 (profitable). Chap. 11 also explains the basic mechanism of cash flow analysis for any project, sources of funds and uses of funds, etc. This current chapter mostly focuses on the sources and uses of funds, budget, and cash flows of a housing project. Project budget and cash flow analysis are called real estate pro-forma analysis.

Pro-forma and Cash Flow

Real estate pro-forma analysis is a type of cash flow analysis. The initial step of cash flow analysis is to establish estimated operating expenses and major assumptions based on local housing market and area standards. Operating expenses cover various expenses related to payroll and benefits, maintenance, utilities, etc. (see Tables 15.13, 15.14, and 15.15). Major assumptions include vacancy, revenue, other income, other expenses, total mortgages, mortgage loan-to-value (LTV) ratio, etc. Acquisition cap ratio is the capitalization rate, which is the ratio of Net Operating Income (NOI) to current property value. For example, if a real estate property has a NOI of $80,000 and the current property value is $800,000, then the cap ratio is $80,000/$800,000 = 10%. If the pro forma is accurate, the cap ratio is the return on income for real estate investment. A cash flow statement is then generated based on the operating expenses and major assumptions (Table 15.15). The cash flow statement measures the differences between income and expenses over a certain time period. In Table 15.15 before-tax cash flow was the value that is calculated as the difference between income and expenses, before paying property, income, or corporate taxes. When discounting these cash flow data into current values they are converted to Net Present Value (NPV) (see Chap. 11 for formulas and methods of calculation).

Analyzing Housing Market Conditions

Housing market conditions vary across different locales and analyzing the conditions requires analyzing the supply and demand conditions, but mostly the supply conditions in relationship to demand. For example, the annual "The State of the Nation's Housing" by the Joint Center for Housing Studies of Harvard University

Table 15.13 Example of operating expenses

	Total	Per unit
Payroll and benefits	$225,000	$563
Utilities	215,000	538
Repairs and maintenance	195,000	488
Insurance	35,000	88
Landscaping	20,000	50
Marketing	60,000	150
Administrative and general	75,000	188
Legal and accounting	15,000	38
Subtotal	$840,000	$2100

Table 15.14 Cash flow calculation assumptions, an example

Major Assumptions		**2020**
Property taxes/unit/year		$1500
Inflation rate:		
Revenues		4.0%
Expenses		4.0%
Management fee		4.5%
Acquisition cap rate		9.0%
Vacancy rate		6.0%
Bad debt expense		1.0%
Other income/occupied unit/year		$100
Replacement reserve/unit/year		$200
Rental revenue mix:		
1 bedroom, 1 Bath	**100**	$825
2 bedroom, 2 Bath	**200**	$975
3 bedroom, 2 Bath	**100**	$1125
	400	
Financing		
LTV:		0.8
Amount financed:		$23,456,444
Mortgage term:		30
Mortgage rate:		8%
Mortgage constant:		0.088827
Mortgage payment:		$2,083,576
Holding period		5

Table 15.15 Example of cash flow analysis

	2020	2021	2022	2023	2024
Gross potential income:					
Studio, 1 Bath	$990,000	$1,029,600	$1,070,784	$1,113,615	$1,158,160
1 bedroom, 1 Bath	2,340,000	2,433,600	2,530,944	2,632,182	2,737,469
2 bedroom, 2 Bath	1,350,000	1,404,000	1,460,160	1,518,566	1,579,309
Subtotal	4,680,000	4,867,200	5,061,888	5,264,364	5,474,938
Vacancy allowance	280,800	292,032	303,713	315,862	328,496
Bad debt expense	46,800	48,672	50,619	52,644	54,749
Net potential rent	4,352,400	4,526,496	4,707,556	4,895,858	5,091,692
Other income	37,600	39,104	40,668	42,295	43,987
Effective gross income	4,390,000	4,565,600	4,748,224	4,938,153	5,135,679
Operating expenses:					
Payroll and benefits	234,000	243,360	253,094	263,218	273,747
Utilities	223,600	232,544	241,846	251,520	261,580
Repairs and maintenance	202,800	210,912	219,348	228,122	237,247
Insurance	36,400	37,856	39,370	40,945	42,583

(continued)

Table 15.15 (continued)

	2020	2021	2022	2023	2024
Landscaping	20,800	21,632	22,497	23,397	24,333
Marketing	62,400	64,896	67,492	70,192	72,999
Administrative and general	78,000	81,120	84,365	87,739	91,249
Legal and accounting	15,600	16,224	16,873	17,548	18,250
Management fees	197,550	205,452	213,670	222,217	231,106
Replacement reserve	80,000	83,200	86,528	89,989	93,589
Property taxes	600,000	624,000	648,960	674,918	701,915
Total operating expenses	1,751,150	1,821,196	1,894,044	1,969,806	2,048,598
Net operating income	$2,638,850	$2,744,404	$2,854,180	$2,968,347	$3,087,081
Debt service	$2,083,576	$2,083,576	$2,083,576	$2,083,576	$2,083,576
Before-tax cash flow	**$555,274**	**$660,828**	**$770,604**	**$884,772**	**$1,003,506**
Debt service coverage ratio	1.2665	1.3172	1.3698	1.4246	1.4816
Operating expense ratio	0.3712	0.3712	0.3712	0.3712	0.3712
Breakeven cash flow ratio	0.8129	0.7959	0.7795	0.7638	0.7487
Operating return on asset	0.0900	0.0936	0.0973	0.1012	0.1053
Gross rent multiplier	6.2151	5.9761	5.7462	5.5252	5.3127

covers the housing markets (new construction, for-sale inventories, home price growth, and housing affordability), demographic drivers, homeownership, rental housing, and housing challenges nearly every year.

Housing Construction

Housing construction usually involves around exploring the number of new permits issued, number of new constructions, and the condition and characteristics of the newly built units. Trend over the years and outlook to the future should be explored to help the audience and readers understand the longitudinal characteristics of the housing construction industry. A typical analysis should include at least the following items:

- Number of new permits issued for the study period
- Housing conditions, number of new units, size, and price, etc., of different types of development, such as residential, office, commercial, and industrial, comparing with demand conditions
- Locational characteristics of new construction
- Construction of affordable housing units

Housing Price

Housing price and change in housing price reflect not only the state of the housing market, but act as the barometer to measure the overall health of an economy. Housing price can be measured with different indicators.

Median housing value is a variable used in the U.S. census to measure the median housing value of a census unit, based on the survey respondent's estimated housing value for owner-occupied housing, if the property were offered for sale. Since this is a rough estimate, there may be a significant number of errors when being compared to the actual sales price of a comparable unit in the neighborhood.

Housing price is the transaction price of properties, which is more accurate and reliable than estimated housing value. Based on historic transaction data a housing price index can be constructed to reflect the dynamics of housing price change in a locale. In the U.S., Federal Housing Finance Agency (FHFA) calculates the *Housing Price Index* (HPI) based on repeat sales of properties in a geographic area. HPI also controls the effect of inflation so that the index is a fair representation of real housing appreciation/depreciation.

Median housing value and housing prices vary greatly depending on specific housing market conditions in a geographic area. Consistent appreciation or brief depreciation during a recession but bouncing back quickly after the recession usually indicate a strong or hot housing market, which also indicate the strong economic prospect in the area. Housing price change at neighborhood levels also indicate stable, declining, or improving neighborhoods.

Housing price is determined by various factors, such as the attribute of the property itself and the location of the unit. Property attributes that strongly predict housing prices are unit size, number of bedrooms, and number of bathrooms. Locational variables are often relating to the distance from CBD (Central Business District), the distance from major highways or lightrails, the distance from the ocean, or amenities in the neighborhood such as presence of parks, waterbodies, transit stops, etc. Therefore, ***hedonic housing models*** usually use these contributing factors to measure housing price in a linear (straightline) or non-linear multiple regression model.

Housing Tenure

Housing tenure is the financial arrangement of living in a housing unit, indicating owning versus renting. Homeownership rate is used as an important indicator for neighborhood stability, although the rate differs greatly among different places. In general, higher homeownership rate is associated with increased neighborhood stability due to lower rates of residential turnover and higher vested interests in the neighborhood by homeowners. However, large metropolitan areas often saw a lower homeownership rate and the neighborhood effect of homeownership may not be highly applicable in urban neighborhoods.

Owners tend to be wealthier, older, and family households, while renters tend to be less wealthy, younger, and households without children. There is a higher percentage of renters with housing cost burdens, meaning paying more than 30% of the household income in housing. Rental housing also tends to have a higher vacancy rate in owner-occupied housing, and thus neighborhoods with a higher percentage of renters have higher vacancy rates than neighborhoods with higher homeownership rates.

Another important segment of housing tenure analysis is analyzing the characteristics of the homeless population. In the U.S., data on the homelessness are usually from the Point-in-Time (PIT) count of sheltered and unsheltered people experiencing homelessness on a single night in January.

Housing Vacancy Rates

Housing vacancy rate is measured as the percentage of vacant units, long-term and excluding those for sale or vacation homes, among all housing units in a geographic unit. Housing vacancy always exists in the housing market, due to mismatch of market information. Markets with strong real estate absorption rates, such as a hot market, tend to have lower vacancy rates, and therefore housing price or rent is higher due to tighter housing supply. Conversely, markets with sluggish or declining economic growth would have a higher vacancy rate. Overall, for a real estate investor, when the vacancy rate of a real estate portfolio is higher than 8%, the investor should watch for the warning signs of declining profits due to the high vacancy. Excessive vacancy due to housing abandonment should be closely monitored by local communities to prevent its negative impact on neighborhood stability.

Housing Affordability Analysis

As covered in some of the previous sections in this chapter, housing affordability analysis mostly focuses on the gap between supply and demand of housing units for households earning less than 120% AMI. Housing affordability indexes across different metropolitan statistical areas also allows for comparison among different geographic areas. Meanwhile, analyzing the housing cost burden and percentage household income used in housing may also provide a glimpse of the affordability issues in an area. When analyzing housing cost burdens, it is imperative to combine with analyzing transportation costs as well, since housing and transportation are proven the largest household expenses.

Using GIS to Analyze Housing Market Conditions

Since the invention of geographical information systems, GIS has been widely used in various disciplines and fields, ranging from businesses to government agencies. The following housing and real estate analyses often use GIS to aid in data processing and analysis:

- Site suitability analysis. GIS is widely used in analyzing site conditions to explore the feasibility of a housing or real estate project. Onsite and offsite conditions, such as vegetation, vista views, traffic and transportation network, environmentally sensitive areas, etc., can be easily mapped with the assistance of GIS software.
- Target market analysis. GIS can not only incorporate demographic data into various geographic units and generate analysis of population segmentation characteristics, it can also map out the potential market capture areas (trade areas), based on drive time, ring buffers, or walk time.
- Simulation analysis. Simulation analysis includes simulating housing price change, future outlook of the housing market, etc.

References

Alonso, W. (1964). *Location and theory of land rent: Toward a general theory of land rent*. Publications of the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University. Harvard University Press.

Bartik, T., & Eberts, R. (2011). The role of tax incentives and other business incentives in local economic development. Brooks, N., Donaghy, K., & Knaap, G-J. (Ed.), *The Oxford handbook of urban economics and planning*. Oxford, England: University of Oxford Press.

Florida Department of Revenue. (2018). *Production guide and data record layout for the comma delimited file format.* https://floridarevenue.com/property/Documents/2018prodguide.pdf

The Metropolitan Center, Florida International University (FIU). (2018). *Broward county affordable housing needs assessment.* https://www.broward.org/PlanningCouncil/Documents/2018AffordableHousingNeeds.pdf

Goetz, E. G. (2012). The transformation of public housing policy, 1985–2011. *Journal of the American Planning Association, 78*(4), 452–463. https://doi.org/10.1080/01944363.2012.737983

HR&A Advisors, Inc. (2020). *Montgomery county housing needs assessment.* https://montgomeryplanning.org/wp-content/uploads/2020/07/MoCo-HNA-July-2020.pdf

Muth, R. F. (1969). *Cities and housing*. University of Chicago Press.

Newman, S. J., & Harkness, J. M. (2001). The long-term effects of public housing on self-sufficiency. *Journal of Policy Analysis and Management, 21*(1), 21–43.

Ottensmann, J. R., Payton, S., & Man, J. (2008). Urban location and housing prices within a hedonic model. *The Journal of Regional Analysis & Policy, 38*(1), 19–35.

Palmer, J., Boardman, B., Bottrill, C., Darvy, S., Hinnelis, M., Killip, G., Layberry, R., & Lovell, H. (2006). *Reducing the environmental impact of housing*. Consultancy study in support of the Royal Commission on environmental Pollution's 26th report on the urban environment. Environmental Change Institute, University of Oxford.

Schill, M. H. (1993). Distressed public housing: Where do we go from here? *The University of Chicago Law Review, 60*(2), 497–554.

Schwartz, A. F. (2015). *Housing policy in the United States* (3rd ed.). New York.

Walter, R. J., Li, Y., & Atherwood, S. (2015). Moving to opportunity? An examination of housing choice vouchers on urban poverty deconcentration in South Florida. *Housing Studies, 30*(7), 1064–1091. https://doi.org/10.1080/02673037.2015.1009004

Varady, D. P., & Preiser, W. F. (1998). Scattered-site public housing and housing satisfaction: Implications for the new public housing program. *Journal of the American Planning Association, 64*(2), 189–207.

Web Resources

U.S. Department of Housing and Urban Development (HUD).: https://www.hud.gov/.
HUD Exchange.: https://www.hudexchange.info/.
HUD User.: https://www.huduser.gov/portal/home.html.
American Housing Survey.: https://www.census.gov/programs-surveys/ahs.html.
American Community Survey.: https://www.census.gov/programs-surveys/acs
Federal Housing Finance Agency.: https://www.fhfa.gov/.
Federal Financial Institutions Examination Council (FFIEC).: https://www.ffiec.gov/.
U.S. Consumer Financial Protection Bureau. Home Mortgage Disclosure Act (HMDA): https://www.consumerfinance.gov/data-research/hmda/.
Center on Budget and Policy Priorities.: https://www.cbpp.org/topics/housing
National Housing Conference.: https://nhc.org/.
National Low Income Housing Coalition.: https://nlihc.org/.
National Coalition for the Homeless. https://nationalhomeless.org/.
National Urban League.: https://nul.org/.
Urban Institute.: https://www.urban.org/.
Urban Land Institute.: https://uli.org/.
NeighborWorks America.: https://www.neighborworks.org/home.
Enterprise Community Partners.: https://www.enterprisecommunity.org/.
U.S. Green Building Concil.: https://www.usgbc.org/.
Brookings Institution.: https://www.brookings.edu/.
National Association of Realtors.: https://www.nar.realtor/.
Cities Alliance: Cities without Slums.: https://www.citiesalliance.org/.
United Nations Development Programme.: https://www.undp.org/content/undp/en/home.html.
UN – Habitat.: https://unhabitat.org/.