BUSINESS SITE SELECTION, LOCATION ANALYSIS, AND GIS

RICHARD L. CHURCH
ALAN T. MURRAY

WILEY
JOHN WILEY & SONS, INC.

This book is printed on acid-free paper. ♾

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Church, Richard L.
Business site selection, location analysis and GIS / Richard L. Church, Alan T. Murray.
p. cm.
Includes index.
ISBN 978-0-470-19106-4 (cloth)
1. Industrial location–Geographic information systems. I. Murray, Alan T. II. Title.
HD58.C48 2009
658.2'1–dc22

2008022806

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To our families (Carla, Steven, and Karen, & Patty, Ryan, and Natalie)

CONTENTS

PREFACE

Location theory is rooted in the disciplines of geography, engineering, mathematics, and economics. The science of "where should it be" is truly multidisciplinary and continues to be of interest to practitioners and researchers alike representing a variety of fields, ranging from business to operations research to computer science. The contributions of modern pioneers such Hakimi, Vasonyi, Cooper, and ReVelle helped to expand the underlying theory, as well as to formulate relevant models for application. Since the first location algorithm was proposed in 1937 by Weiszfeld (Vasonyi), this field has relied on the computer to solve and analyze location problems. This reliance has been strengthened even more with the emergence of capable commercial modeling software packages as well as geographical information systems. We consider GIS and optimization techniques as equally important tools in this maturing field. In fact, scholars and practitioners today need an understanding of both areas (GIS and optimization).

Existing books on location tend to discuss the topic from a perspective of GIS or optimization, but not both. But the science of *where* has matured to the extent that this dichotomy is no longer ideal, so students need to be firmly grounded in both subjects. This need is what inspired us to write this text.

There is no required background for this text, simply an interest in spatial analysis. The level of discussion and technical sophistication is for both undergraduates and graduates, with all chapters including introductory and advanced material. The 12 chapters in the text can be taught over a 10-week quarter or 16-week semester.

Maintaining a reasonable length for a textbook often means that choices need to be made regarding what is covered and what is not. This book is no exception. The reader will find that all chapters, except Chapters 2 and 3, attempt to introduce location modeling as well as relevant GIS issues. Each topic is approached so that it is not necessary to have had a prior course in either area, although that might be helpful or preferable. The first three chapters are introductory and should be covered by all readers. The second part of the text is based on using models to analyze existing systems (marketing and distribution), as well as the siting of a single facility—be it point, line, or area. The third part of the text deals with the design of a system while making multiple simultaneous siting decisions. This final section addresses models for covering, dispersion/noxious facilities, and median/plant location. Although some topics were not able to be covered, such as hub location or hierarchical system design, the text is structured so that it can easily be supplemented by assigned readings associated with these more specialized topics.

Practitioners will find this text helpful as a reference given the comprehensive introductory material as well as advanced topics.

ACKNOWLEDGMENTS

This text is of our own design, and we alone are responsible for any critical omissions. We are, however, indebted to a number people and wish to express our gratitude to them for their help and support. We want to thank staff and students in the Center for Urban and Regional Analysis at OSU, especially Tim Matisziw and Wenqin Chen, and Matt Niblett (UCSB) in helping set up examples, produce figures, and proofread drafts.

We also wish to thank our wives, Carla and Patty, for their encouragement, support, and inspiration. Without their help, this book would have not been possible.

Finally, we wish to acknowledge Chuck ReVelle, Mike Goodchild, and others in this field who have both inspired and assisted us in this endeavor.

Rick Church
Alan Murray

CHAPTER 1

INTRODUCTION

1.0 MOTIVATION

Did you ever wonder why a hospital, fire station, or shopping mall was built in a certain location? Identifying the best site for a particular facility is no easy task. Planners must consider economic and demographic factors, while addressing political realities. This task is as old as our world. From early prehistoric times man has had the task of determining the best location in which to site a place to dwell, for hunting food, harvesting natural resources, and so on. As civilizations evolved, cities developed, trade routes emerged, and roads were constructed. When observing the remains of aqueducts, roads and buildings of the Roman Empire, it is easy to wonder how such systems evolved and how critical siting decisions were made. One can conclude that location decisions have always been essential. For the hunter and gatherer, a good location decision might have meant survival. Today, such decisions might not be a life and death issue, but they are nonetheless very important.

When we look at our current urban systems, we can see specific types of order and systematic arrangement. For example, older cities were located along rivers, close to the fall line—a good place to generate mechanical power and, later on, electricity. Many major cities have port facilities, as they serve as gateways to hinterlands away from the coast or navigable waterways. Some cities have developed close to major natural resources, where, for example, they can convert coal and iron ore to steel. As one looks closely at vast areas of fertile soil and adequate rainfall, a spatial arrangement of villages, towns, and cities emerges that appears almost like a regular pattern of specific geometric shapes. Christaller (1932) reasoned, "Just as there are economic laws which determine the life of the economy, so are there special economic-geographic laws determining the arrangement of towns." A contrarian might suggest that poor choices in location will be punished economically. But no one disputes the importance of location in business, manufacturing, retail, and public services.

There are three ways in which we can study location. The first approach is to *map* what we see by positioning an object's location in space by its locational coordinates. We can use such positional coordinates to accomplish tasks like navigating across an ocean, mapping all land parcels in a county, or calculating the amount of deforestation. Collecting spatial data and capturing its geographical position is an important task. The major function of **geographical information systems (GIS)** is to provide a means to collect, store, retrieve, map, and analyze such spatial data. GIS allows us to store a number of different themes of data across space, including elements such as soil type, land cover, precipitation, elevation, and so on. It is with this richness of data types and the ability to manipulate, model, and analyze spatial data in complex ways that the real value of GIS becomes apparent. In fact, it is upon this seamless fabric of multiple layers of data that today's location decisions are made. The main theme of this book involves how to integrate spatial analysis and logic with GIS.

The second approach to location analysis is an attempt to describe arrangements that have emerged and try to explain why certain decisions were made. For example, we may look for spatial clusters of a specific industry, where an individual firm appears to have made a similar siting decision based on some agglomeration property like nearness to a uniquely qualified work force. We call this a **descriptive approach**. The descriptive approach attempts to explain what emerges over time, or where a certain animal is likely to be found, as an example.

The third approach to location analysis is to identify the best location for an activity, or maybe the best set of locations for a system of activities. We call this type of modeling **prescriptive** or *normative*, in that we are trying to determine which location is best, rather than why certain location patterns have emerged. That is, the inherent feature of normative modeling is to help make decisions for the present and future. Although many models have been developed from a purely theoretical viewpoint, many others have been developed with a definite bent towards the notion of supporting a specific application, such as locating a store, locating an alignment for a road, or siting a warehouse for a distribution system.

Both pure and applied location analysis have emerged into a field called **location science**. The growth of the field of location science has occurred somewhat independent of the development of GIS. This book arises from a need for addressing specific components of location conditioned with GIS. The integration of location modeling with GIS allows one to accomplish more, do it more efficiently, produce better data and model representations, develop better solution approaches, provide new insights, and aid in the visualization

of location alternatives. Our central focus will be on how practical, real-life problems involving a "siting decision" can be supported through the use of this integrated approach.

You have probably heard that there are three important issues in real estate: "Location, location, location!" There is no denying that this is true to some extent in many areas such as commerce, natural resource management and conservation, resource extraction, manufacturing, and product distribution systems, but it is important to understand that every activity has a place or location, and often it impacts the function or role of that activity. Although it is not the only consideration in the planning process, the "location decision" is obviously important, and is the central focus of this book.

In the remainder of this chapter, we will introduce what we consider to be the theoretical roots of normative location modeling, along with a set of *first principles* of geographical location theory. We also will introduce several historical developments in which modeling and quantitative assessment have become the preferred approach to resolving siting decisions. We acknowledge that some siting decisions are made in an ad hoc fashion, and they will continue to be, but as our urban and commerce systems become more complex, the ad hoc and "back-of-the-envelope" approaches fail to capture all of the important problem complexities and are more likely to fail in identifying the best alternative. Worse yet, these simple approaches may highlight or suggest an inferior alternative to what actually exists.

Within the context of spatial planning, one may well think, "There are no easy problems left." It is important to recognize that *greenfield* problems, where we start from scratch, do not exist in most applications. Often, we have parts of a system that are already in operation and we need to locate one or more facilities in order to serve or operate in conjunction with the existing elements. This fact alone demands a greater concern for characterizing what exists, as well as locating within the context of the built-up or altered landscape. Consequently, the role of location scientists must include capturing and representing relevant landscape elements, from transport linkages to sources of raw materials, so that models of location activities include all relevant features such as barriers, connections, zones of demand, and so on. It is important to handle such complexities rather than ignore them or assume they are unimportant. In fact, the merger of location science with GIS will be a necessity in order to handle the complexities of the real world.

1.1 HISTORY

Location science deals with siting one or more activities or facilities in such a manner so as to optimize one or more objectives. One of the first problems of location science was proposed by Pierre de Fermet in the early 1600s. Fermet suggested the following problem: Given three points on the plane, find the fourth point such that the sums of the distances to the three given points is a minimum. This simple, geometrically inspired problem dealt with finding a central *median* location.

Weber (1909) described a three-point problem associated with an industrial setting as follows: Find a location for a manufacturing facility that receives raw material from two point sources and ships its final product to a point-specified market. Thus, Weber's problem suggests finding a fourth point (the factory) among three points (two raw-material points and one market) in order to minimize weighted distance (see Figure 1.1). Weights here represent per-unit product amounts (i.e., the amount of raw material 1 per unit product, the weight of raw material 2 per unit product, and the weight of the final product).

The objective of the classic Weber problem is to locate the most efficient site for the factory, assuming that transport costs are a function of Euclidean (straight-line) distances. Fermet defined an unweighted problem, whereas Weber defined a weighted point location problem that represented the best location of an industrial activity. This problem is usually called the **Weber problem**, given the realistic application orientation, rather than a problem of geometrical curiosity. Weber's analysis was the first to pose a problem of industrial activity in terms of optimal placement. Weber's original work has been expanded over time in two problem domains: networks and surfaces (planar, spherical, inclined planes, etc.).

In 1826, Johann H. von Thünen proposed a model for analyzing agricultural patterns. In his explanation, he assumed that there was an isolated city surrounded by a hinterland that was uniform and unbounded in any direction. The city receives agricultural products from the surrounding hinterland only and nothing from other areas. The hinterland is occupied by farmers who wish to maximize their profits, and who adjust automatically to the market's demands. There is only one mode of transport: horse and wagon (hey, it was 1826!). It is assumed that transportation costs are borne solely by the farmers

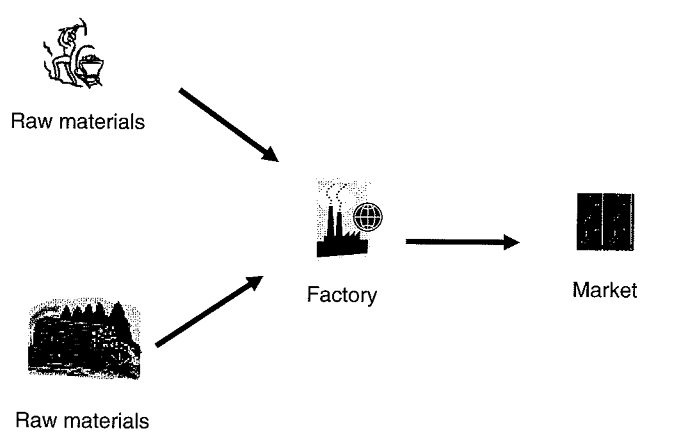


Figure 1.1 Weber's problem of locating a factory to minimize costs.

and are proportional to the distance to the market. Within this framework, von Thünen considered three interrelated factors: the distance farms are from the market (i.e., city), the prices received by the farmers for their goods, and economic rent of the land.

Land rent represents the difference between revenues obtained from the land and the costs of working that unit of land, and is the surplus left after all costs have been deducted.

Figure 1.2 depicts land rent as a function of distance from the city for two commodities. Each function is linear and decreases as a function of transport distance to the city. Notice that the cost of shipping commodity 2 is higher per unit distance than that of commodity 1. But at the market, commodity 2 commands a higher price than commodity 1. Thus, farmers near the market should plant commodity 2 instead of commodity 1. Beyond distance X, commodity 1 provides a greater net land rent than commodity 2. Thus, von Thünen reasoned that there were effectively rings around the city corresponding to different crops, reflecting the different values of uses for land. Thus, a given farmer would select those crops that would generate the greatest land rent, depending on the distance to the market and whether the demand for a given commodity has been met by farmers who are closer to the market. The von Thünen model is a simple economic construct that describes the process of land use allocation. This was the first attempt to systematically address land use allocation.

Hotelling (1929) wrote about pricing stability between two vendors of water. In his initial description, Hotelling fixed the vendors at sites of artesian wells along a linear market (see Figure 1.3). These two vendors paid nothing for the water, including pumping costs, as they were artesian sources. Hotelling's analysis showed how these two vendors would engage in **competition** for market share by setting prices, given that each attempted to maximize profit. Hotelling demonstrated that the two vendors acted like a two-person game, each responding, in turn, to the competitor's price until an equilibrium was reached. In his discussion of this model, Hotelling allowed competitors to

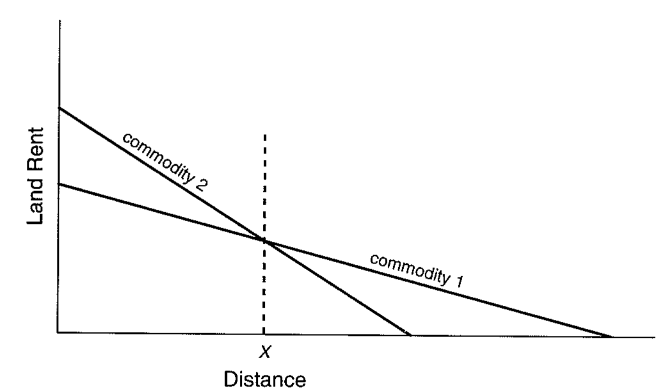


Figure 1.2 von Thünen's analysis of land rent for agricultural commodities.

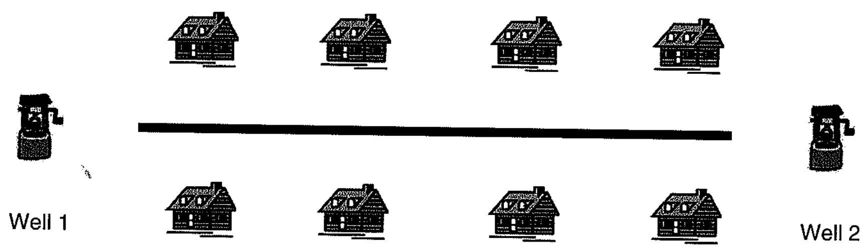


Figure 1.3 Hotelling's competitive location problem.

move positions along the linear market. He reasoned that the two competitors would, in attempting to maximize profit over a series of moves, end up co-locating at the center of the market, sharing the market equally between them. Hotelling's analysis demonstrated the notion of competing retail strategies, gaming, and resulting location decisions.

In 1933, Walter Christaller attempted to describe the arrangements of the villages, towns, and cities of southern Germany. His work became the foundation of **central place theory**. Christaller's premise was that there are natural laws that could explain a systematic arrangement of towns. Christaller assumed that, given an unbounded fertile farming region, certain systematic and geometrical arrangements of retail centers would occur. He viewed the purchase of goods from retail centers based on several properties. First, all goods have a **range**, which is the furthest distance that consumers are willing to travel to buy the good. If the good is a refrigerator, for example, people are willing to travel a fairly long distance to make a purchase, because refrigerators are not purchased very often. In contrast, if the good is something like bread, which is purchased frequently, then a consumer will not be willing to travel far.

Goods with low distance ranges are classified as *low-ordered goods*, and goods with high distance ranges are *higher-ordered goods*. Christaller reasoned that centers offering low-ordered goods would be more plentiful and spread out, whereas centers offering higher-ordered goods would be less numerous.

Christaller also introduced the notion of **threshold**. The threshold is the distance from a center at which the demand for the good is large enough to satisfy the requirements for a vendor to remain in business. It is simple to see that if the threshold of a good exceeds the range of the good, at a given location, then that location is not profitable for offering that good.

Christaller assumed that each customer would travel to the closest retail center offering the good of interest, as retailers of a given type were considered to be equal in all ways. He reasoned that excess profits occurred when the market range exceeded the threshold. He further stated that entrepreneurs would attempt to compete so that excess profits would be kept to a minimum. He postulated that the ideal arrangements of centers offering the lowest ordered goods would be points on a triangular lattice, carving out hexagonal market

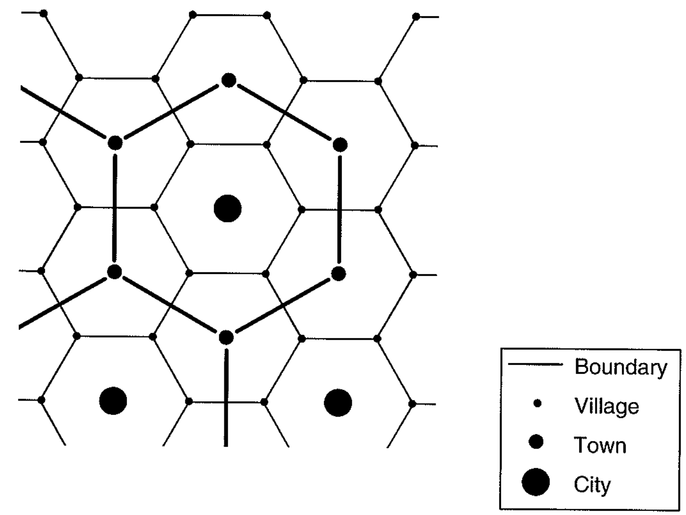


Figure 1.4 Christaller's central place hierarchy.

areas, where each customer assigns to his or her closest center. The next higher-ordered center would be located among a subset of the lowest-ordered centers. Thus, he argued that a hierarchical arrangement existed where high-ordered centers offer all goods that are offered by lower-ordered centers. Each set of higher-ordered centers, therefore, formed hexagonal-shaped market areas. A set of arrangements representing Christaller's marketing principle is given in Figure 1.4.

Geographers have had a long-term interest in Christaller's work. In the 1960s and early 1970s research focused on the geometrical arrangements reminiscent of that postulated by Christaller (Berry, 1967). Altogether, Weber, von Thünen, Hotelling, and Christaller represent founding fathers of location science, where each work laid the foundation for different areas of location research, retail competition, spatial layout, and industrial location. Much of the early work took place before computers were available. With the availability of computers, location science took a decided bent toward the use of computation and the solution of more complex location problems. Some of the leaders in the beginning of the computational modeling era of this field are Walter Isard, Leon Cooper, Charles ReVelle, Michael Teitz, and Louis Hakimi. The contributions of many of these individuals will be described at greater length in different chapters of this book.

1.2 FIRST PRINCIPLES

Some time ago, Tobler (1970) examined urban growth in an urban region. In the development of a model of population distribution, he discusses the nature of population from a global perspective. After identifying difficulties involved in specifying a model of global scale relative to impacts on a particular urban area, Tobler invoked what he termed the *first law of geography*:

> "Everything is related to everything else, but near things are more related than distant things."

With this principle, Tobler crafted a model for the urban area he was working with, Detroit, that ignored most of the world. It is important to understand that each field of science is based on a set of principles or laws, under which elements are analyzed, new laws are identified, and the field is advanced. It is natural to ask just what are the first principles of location science, and how would such principles guide us toward further study and scientific advancement?

In this spirit, we identify three **laws of location science (LLS)**. The first law of location science can be stated as follows:

LLS1—*Some locations are better than others for a given purpose.*

Even though this statement may seem obvious, the impact of this law is significant. If some locations are better than others for a stated purpose, then given a stated purpose, what is the best location at which to site a given activity? As an example, if we wish to place a fire station in order to serve a given area, it would be unlikely that the station would be located outside of the service area. Thus, one can begin the process of identifying a region, zone, or polygon within which a station should be located. The bottom line, of course, is that location matters.

A corollary to LLS1 is the following: *Efficient system locations tend to beat inefficient ones*. Holding all other factors constant, an efficient location pattern will tend to persist longer for a given use than inefficient ones. For example, if two stores are competing in a market that can support only one store, assuming that both offer the same goods and services at the same price, customers will go to the closest store. The store in the least efficient location, therefore, will have fewer customers and thus go out of business first. As a result, without other compelling factors, one should seek the best location(s) for an activity.

The second law of location science is this:

LLS2—*Spatial context can alter site efficiencies.*

Suppose you are going to locate a dry-cleaning business and you have identified an area to open a store in. The site that is the closest to your estimated customer base may not be as good as a site next to the local grocery

store, as an example. That is, people can drop off their dry cleaning on their way to the grocery store, as it would be convenient and they visit the grocery store often. Even though the customer base of the grocery store may include a number of people that do not use dry-cleaning services, and that a better place to serve just those in need of dry cleaning might exist, locating next to the grocery store may be more efficient as the spatial context of other goods and services have altered travel patterns and behavior, thereby altering the landscape of efficient locations. Therefore, spatial context is important.

Often, sites where goods and services are provided tend to concentrate and build up around a given place, providing economies of agglomeration. For example, law offices are often selected close to courthouses, so that lawyers can minimize time traveling to and from the courthouse in the process of doing their business. Even though a customer base for a given legal group could be quite some distance from the courthouse, the cost of doing business may still be minimized by locating near the courthouse and ignoring the distances traveled by the clients. If a legal office is located at a great distance from a courthouse, then you will find that it tends to work in a specialty area in which the location of the local courthouse is not as important. Another example is the location of parts manufacturing in relation to assembly points for automobiles. If a given parts manufacturer only supplies parts to a single assembly plant, then the best location may well be close to the assembly plant itself. It is easy to see that the second law of location science is important in that there are dependencies between the location of other services that can be important determinants in finding the best location for a given activity.

The third law of location science is associated with investment in multiple facilities. Such problems are often termed *multifacility location problems*. Here is the third law:

LLS3 *Sites of an optimal multisite pattern must be selected simultaneously rather than independently, one at a time.*

Up to this point, virtually all of our examples have dealt with the location of a single store, office, or manufacturing location. But suppose that several locations are needed. For example, consider a simple problem of locating several pizza stores. Assume that only pizza delivery is provided, and there is guaranteed delivery of pizza in 30 minutes or less to any place in town or the pizza is free. If 15 minutes will be taken up by making and baking each pizza, this means that each area of the town needs to be within 15 minutes of at least one pizza store, or a lot of customers will be getting a free pizza.

Thus, we need to locate our pizza stores so that each neighborhood in town is within 15 minutes of at least one pizza store. Doing this means that the delivery areas of each store must be no greater in size than 15 minutes' driving time from a store, and that together the delivery areas of all stores must cover the town. Suppose that this can be achieved by the location of four pizzerias. The solution for this cannot be determined without locating the stores as a system (i.e., in concert with one another); otherwise, we risk leaving a neighborhood outside the necessary 15-minute delivery time. The third law highlights the fact that multiple facilities cannot be sited efficiently without taking into account service provided by each facility as an integrated whole. The best location of one facility is dependent on the location of others, and the best system can be found only when all facilities are located simultaneously.

The first law of location science (LLS1) easily leads us to the notion that a search for the optimal or near-optimal locations should be made. Otherwise, we risk locating in inferior places, which may lose out to a competitor's site. The second law (LLS2) means that spatial context is not only important, but that it is necessary to collect and analyze appropriate data for the purposes of making a site selection. Finally, the third law (LLS3) suggests that the search for the best pattern or configuration of sites cannot be achieved by solving a series of one-facility location problems. The impact of LLS3 is enormous, as the complexity of modeling location problems increases as the number of entities being located increases. Except in unusual circumstances, such problems cannot be decomposed into a series of simple independent problems. Thus, it is important to develop relevant multifacility location models to address complex spatial problems.

1.3 PLANNING CONTEXT

The process of planning as done by county and state public agencies and most private companies has evolved over time. For example, most public agencies now have a planning process that involves public participation. For example, in developing an *environmental impact statement* (EIS), the process starts at the outset with an invitation for the public to comment on what it considers to be elements or criteria that must be considered or addressed. Before an EIS is finally approved, the process includes a period of public review and comment, unlike master planning in the 1960s by engineers and planners with little or no public input. Planning by public agencies now includes significant public input. Such an open process requires that decisions be made with a focus on improvement, while mitigating impact as measured by metrics that communicate in a quantitative manner exactly what a project or plan is expected to produce. For example, building a county park may require a traffic study to measure the impact on local traffic levels and their impact on nearby residents, a forecast of expected attendance, a cost–benefit analysis to determine if it is economically viable, and an environmental impact review. The *do-nothing* alternative must always be considered as one possibility for final selection. Comparing alternatives, performing an EIS, and performing a cost–benefit analysis all require a quantitative framework in which to calculate

metrics such as forecasted attendance and traffic flow. Some elements are not easily measured, but such intangibles (e.g., the loss of three acres of native bunch grass) need to be listed and addressed as well. With the exception of such intangibles, measuring and comparing alternatives usually can be presented in a format where each element is reported in a quantitative manner (e.g., expected attendance is 200 people per day, or traffic flows will impact traffic at the intersection of Elm and Main Streets by decreasing level of service B to C in the evening rush hour, or an additional recreational worker will be required).

It is important to note that legislation introduced in the late 1960s and early 1970s associated with the environmental movement goaded and required public agencies to be more open, involve the public, and address issues such as environmental impacts. For example, the National Environmental Policy Act (P.L. 91-190) in 1969, the Endangered Species Act (P.L. 93-205) in 1973, the Clean Water Act of 1972 (P.L. 92-500), the Resource Conservation and Recovery Act of 1974 (P.L. 92-580) and the Forest Management Planning Act of 1976 (P.L. 94-588) promulgated regulations that virtually required a quantitative approach to planning and environmental impact analysis. For example, both the Clean Water Act and the Resource Conservation and Recovery Act require that planning not be isolated and local, but that it be expanded to basinwide water management planning and regional planning for the disposal of solid waste. The cradle-to-grave requirements for hazardous materials of the Superfund Act (P.L. 96-510) require systems of quantitification, modeling, and monitoring. Thus, the era of environmentalism brought about major changes in the planning process, with a resulting need to better quantify impacts, benefits, and costs.

Another trend that emerged was the use of large-scale models called the planning revolution. In the 1960s and early 1970s, it was thought that large-scale comprehensive modeling could support and help resolve the needs of the planning process and environmental impact analysis. Although this supposition was true, there were impediments to a reliance on such large-scale models at the outset. Even though research was supported to develop large and comprehensive planning, monitoring, and management models, something was missing. That major ingredient was the technology to support such models.

First, computers were very expensive and very limited in their capability to handle large, comprehensive planning data sets. As an example, computer memory in most applications was less than 250 kilobytes, and processors were *very* slow compared to today's standards. Overall, the cost of computer use was considerably higher than it is today. The IBM personal computer was not available until 1983. Second, software for data management was limited, and GIS did not exist beyond simple raster-based applications. Software for handling spatial data was very limited, and in most instances was not sold commercially but was often written to support a specific application. Computer-driven plotters for producing maps did not exist, because most maps were produced by hand. Computer-produced maps were limited to crude shaded maps produced by repeatedly striking character keys. Color monitors did not exist, and the media for writing programs was usually limited to punch cards and paper tape. Third, inexpensive software did not exist to solve models and optimization problems representative of planning problems. Software available in the late 1960s could solve problems with a few thousand variables and constraints, but required the complete use of an expensive computer for considerable computational time. This meant that only a few agencies, such as the Air Force or companies like AT&T and Exxon (then Standard Oil), could afford to buy and apply such technology. Finally, technology to collect spatial data and store it was limited. Laser distance measures, GPS units, and satellite imagery did not initially exist. Consequently, the planning process needed technology that did not exist. But the story obviously does not end there, as improvements were made on all fronts: hardware (like computers, storage devices, monitors, plotters, etc.), database software, GIS software, modeling software, statistical software, and so on. Even new algorithms and heuristics were developed to solve model sizes that dwarfed those solved in the past.

Now it is commonplace to talk about large-scale models, such as global climate change models or large-scale models to optimize treatments within a national forest. Even linear programming software is ubiquitous; it is now a component of Microsoft Excel. Computational costs have consistently decreased over time, and now most maps are not drawn by hand but are produced using computer software. Thus, the revolution in planning, technology, and modeling has evolved over the last 40 years. We must now consider how technology can be used to support the continued quantitative revolution, supporting the planning process, optimizing the design and operations of systems involving business, natural resource management, and public services. It is with this perspective that this book has been written. Now we must, as a matter of course, consider just how to approach problems such as location modeling in an integrated manner using GIS and incorporating available census data, satellite imagery, and so on, along with data especially collected for a location model application.

1.4 ROLE OF GIS

The fields of location science and GIS have developed almost independently. The reasons for this are fourfold. First, models structured in the early period of location science were simple and structured as geometric problems (like those of the Weber and/or Fermet). Second, many models in location science utilize elements of operations research (OR). This field involves modeling for decision making where techniques are equally applicable in spatial and nonspatial domains. The field of OR has evolved since the 1940's and many of the models discussed in this text are solved using OR-based techniques. Third, the field of GIS did not develop to support location science, but, rather,

because of the need to support a variety of uses and services. Geographical information systems were developed to collect, store, manage, manipulate, display, and analyze spatial data. Such systems are designed to present spatial data in the form of a map (e.g., thematic map), and are designed to retrieve data in a form that can be useful for analysis. As such, GIS was not developed per se to solve location models but to support a variety of needs, from mapping to spatial queries, and from visualizing a terrain to supplying data to models and statistical tools. That is, the goals in developing GIS transcend specific needs of location science, because the application domain is much broader than location science. Finally, the number of professionals working in both fields has been relatively small, and until now work in one area has been somewhat independent of work in the other. We can depict these three fields with a Venn diagram, as shown in Figure 1.5. Each field is illustrated as an ellipse. Notice that the ellipses overlap, reflecting that they are interdependent in some ways.

It is important to recognize that certain issues addressed in location science are actually spatial planning problems that can be resolved in GIS without knowledge of **operations research** or location science. Also, certain issues in location science can be addressed within a theoretical framework that does not involve actual data or specific techniques in operations research. However, many problems of location, from retail store siting to biological reserve site design, involve the need to characterize an application domain complete with spatial data of considerable detail (e.g., road network, census tracts, population estimates, etc.), and rely on a combination of functionality, from GIS to models and algorithms based on operations research.

As problem applications become more sophisticated, the spatial data needed in their application must be supported in one form or another by GIS. Thus, the role of GIS in location modeling ranges from central to peripheral data support, recognizing that complex spatial manipulation, query, and computation may be necessary. As an example, locating cell-phone towers requires characterizing the terrain along with characterizing surface clutter, which are elements that tend to reflect, bend, or obscure cell-phone signals (e.g., buildings and vegetation).

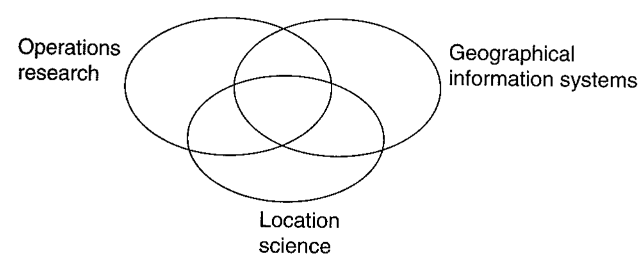


Figure 1.5 Overlapping fields of location science, operations research, and geographical information systems.

Modeling terrain is a central element of many GIS packages, and keeping track of ground cover through data attributes helps to estimate clutter height. Thus, GIS keeps track of data needed to estimate the area coverage of a potential cell-phone site. Simply put, an antenna reception model can be easily integrated with a GIS data model in order to create map coverages of potential sites, whereas such a model would require significant database development and data collection without using a general purpose GIS.

What we have attempted to show is that as these three areas of modeling have matured (GIS, OR, and location science); there is a convergence and burgeoning overlap between these fields based upon the demand for better and more accurate spatial data, the demand for better models characterizing real-landscape problem domains, and the demand to map and visualize solutions in order to support decision making across a wide-ranging set of scales, from the warehouse floor to harvest areas in a large forest plantation to the infrastructure of pipes and pumps, reservoirs, and tanks of a water supply system. Whether a water tower or a retail shop is being located, future applications are likely to be intimately linked with GIS, relying on a wealth of spatial data and spatial operations and utilizing models that characterized the problem domain in a manner that is as close to ground truth as possible. This is not only the future of business location decision making, but location science applications in general. Looking beyond theoretical location constructs and focusing on the resolution of actual siting problems will result in the production of new models, data constructs, algorithms, and theoretical principles. Hopefully, this book will be the first of many that help lead the way in this multidisciplinary field.

1.5 SUMMARY

In this chapter, we have attempted to describe the evolving fields of location science and GIS. We have also attempted to demonstrate that the future of location science and GIS will evolve and overlap, whereby many new applications and models will be developed based on a fusion of these two fields of science. This book represents an attempt to show how this fusion and integration can be accomplished. This book is meant as a textbook as well as a reference. We recognize that many students, and experts, for that matter, will understand one field and not the other. We begin with two introductory chapters: one an introduction to GIS and the other an introduction to mathematical modeling. These two chapters will form the basis of different areas of location analysis and GIS presented in subsequent chapters. Many of our examples will utilize ArcGIS, along with an optimization package called LINGO. In subsequent chapters, our examples will build on the use of ArcGIS, LINGO, and special-purpose routines.

This is not a computer programming book, but a book that presents detailed models and applications, and demonstrates how such problems can be solved with a fusion of tools, relying on concepts that fall within location science and/or GIS.

1.6 TERMS

descriptive approach
prescriptive approach
geographical information systems (GIS)
location science
Weber problem
efficiency
competition
central place theory
range
threshold
laws of location science (LLS)
operations research
planning revolution

1.7 REFERENCES

Berry, B. 1967. *Geography of market centers and retail distribution.* Englewood Cliffs, NJ: Prentice-Hall.

Christaller, W. 1966. *Central places in Southern Germany.* Trans. by C.W. Baskin. Englewood Cliffs, NJ: Prentice-Hall. Originally published in German in 1933.

Hotelling, H. 1929. Stability in competition. *Economic Journal* 39:41–57.

Isard, W. 1956. *Location and space economy.* Cambridge, MA: MIT Press.

Losch, A. 1954. *The economics of Location.* New Haven, CT: Yale University Press. Originally published in German in 1939.

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

von Thünen, J. H. 1826. *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie.* Hamburg: Frederich Perthes. Translated by Peter Hall as *Von Thünen's Isolated State* (Oxford: Pergamon Press, 1966).

Weber, A. 1909. *Uber den standort der Industrien* (Tubingen). Translated by C.J. Friedrichas as *Theory of the Location of Industries* (Chicago: University of Chicago Press, 1929).

1.8 EXERCISES

1.1. In your local town or region, can you list business locations that are struggling, strip malls or shopping centers that appear to have high vacancy rates, or offices with high tenant turnover? How is this explained by the first law of location science?

1.2. Give an example of how a service you rely on is well located. What about one that is poorly located?

1.3. What is a good or service that you would prefer to be co-located, or near, an activity you regularly frequent? Why?

1.4. Many towns of 50,000 to 100,000+ people have experienced growth of retail concentrated at first in regional malls, and now in "big-box" retail shopping centers. Often, such centers have decimated the classic Main Street mercantiles of town.

(a) Has this happened in your area, or can you think of an area where this has happened?

(b) What types of stores have moved onto Main Street after the traditional retail establishments relocated to the mall (e.g., antique stores)?

(c) Can you estimate the threshold for a mall by observing the sizes of cities that have malls?

1.5. Classify the following goods or services within the context of order: low, medium, or high:

(a) Plastic surgery

(b) Convenience store

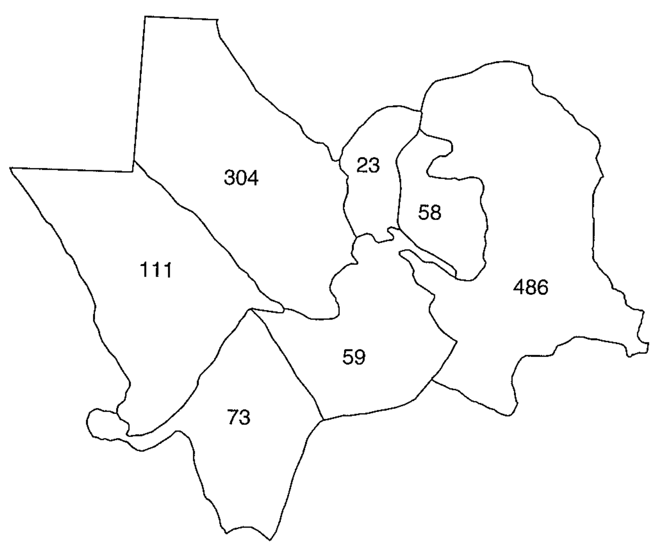
(c) Home appliance and furniture store

(d) Dry-cleaning shop

(e) Videostore

(f) Ferrari dealership

1.6. The neighborhoods depicted in the following figure indicate the number of children under age 10 in a town. Identify the best locations for two day-care centers. Why are these good sites?



1.7. During a time of budget crunch for the city of Los Angeles, the fire department considered closing up to 10 fire stations out of the city's 105 existing stations.

(a) Estimate how many different plans for closing 10 stations exist.

(b) Would you want to close all stations in a given neighborhood? Would this be equitable?

(c) Can you think of one or more metrics that can be used to measure how good one plan is, compared to another?

1.8. In making your decision in problem 1.6:

(a) Did you assume that there were no competitors?

(b) Did you assume that they could serve only those inhabitants within a neighborhood?

(c) Did you consider locating on a boundary between two or three neighborhoods?

CHAPTER 2

GIS

2.0 INTRODUCTION

A geographical information system (GIS) is a particular form of information system that combines geographically (spatially) referenced data, as well as nonspatial attribute data. Information systems, such as Access, FoxPro, Paradox, and Oracle generally process and manage only attribute data that are not explicitly located in space. As an example, suppose we had a spreadsheet of data associated with each of the U.S. states. Aspatial information might include state name, total population, number of households, number of businesses, and so on. However, what about where each state is or which states neighbor (or border) other states? This is a role for GIS, as it is capable of storing and managing the political boundaries of each state, in addition to the attribute information. This enables one to identify, and view, each state on the surface of the earth. In addition, GIS analysis functionality allows for neighboring states to be identified through boundary evaluation. This is the major contrasting feature of GIS, because geographic extent is an explicit and important component of all information being stored, managed, and processed. Thus, GIS may be considered a hybrid information system that structures data and summarizes features based on the inherent characteristic of the data being managed—geographic extent.

Here is a formal definition of a geographic information system (GIS):

> GIS is hardware, software and procedures that support decision making through the acquisition, management, manipulation, analysis, and display of spatially referenced information.

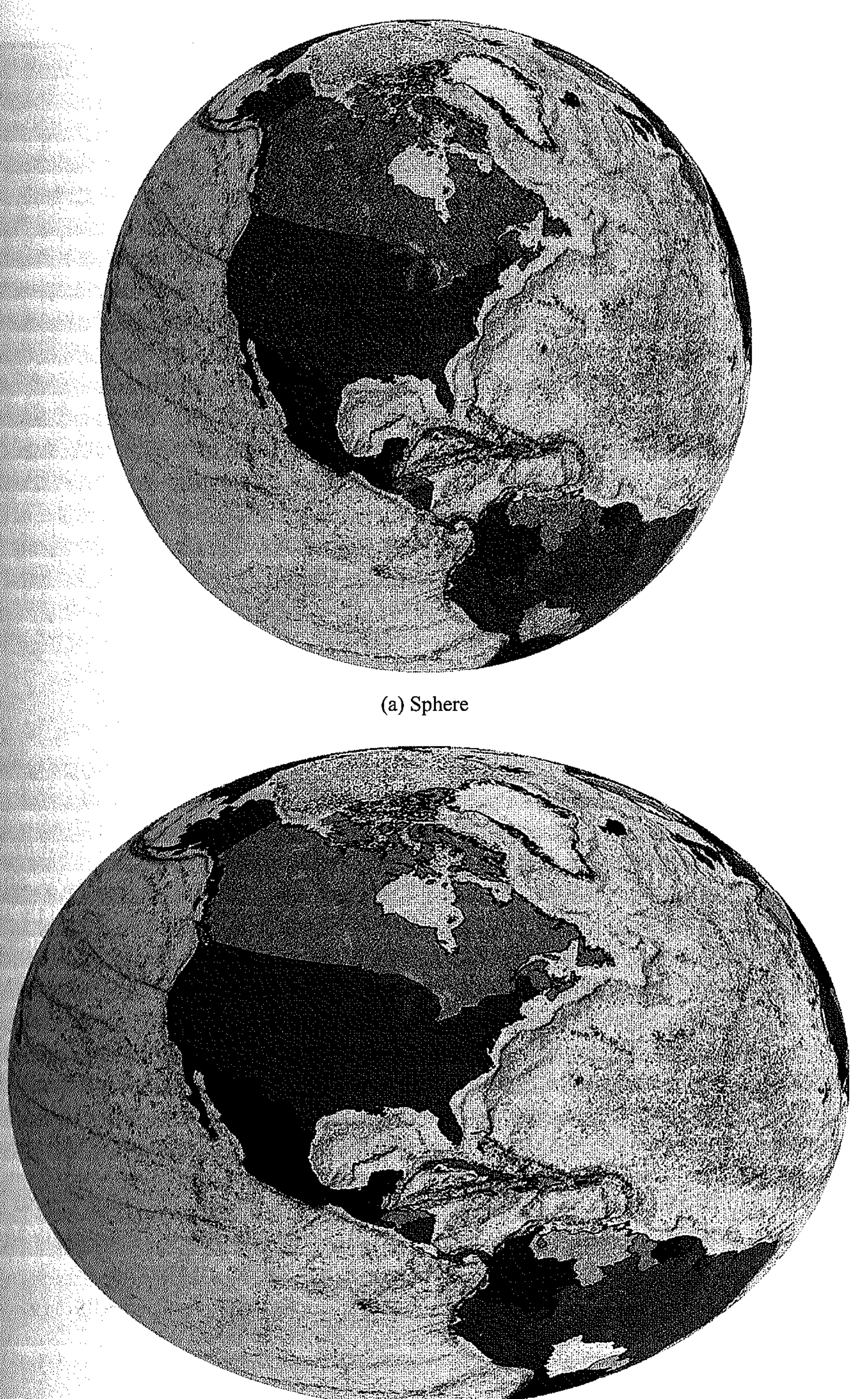
This definition implicitly, if not explicitly, assumes that the real world is being represented in a digital environment. Given this, the earth, out of necessity, is approximated in a number of ways. First, there is the actual shape and form of the earth and how this is represented digitally. Second, Earth is approximated in how objects and themes are represented digitally. Finally, there are selective inclusion, or sampling, issues. We now discuss each approximation issue in turn, but driving representational simplification is the fact that functional concerns for data storage and computational processing are paramount for a successful and responsive GIS operating environment.

Earth is actually an irregularly shaped, three-dimensional object. A practical requirement in a digital environment has been to use a reasonable "regular" representation of Earth's surface in order to support an easy-to-use location referencing system.[1] Perhaps the most prominent assumption is that Earth is a perfect sphere. Of course, knowing that there is much topographical variability on the surface of the earth, it clearly is not a perfect **sphere**. In fact, a better approximation of the shape of Earth is as an oblate ellipsoid or **spheroid**. Better still is the **geoid**. These approximations are illustrated in Figure 2.1, and reinforce the notion that all of them only approximate the actual shape and size of Earth. As a result, there is undoubtedly representational error of some sort. Fortunately, such representational error has proven to be manageable, enabling significant data storage and computational efficiencies to be realized.

The second digital approximation issue raised is how objects and themes are represented on the earth's surface. This is also referred to as *data modeling*. Formally, data modeling has to do with a field versus object perspective. In practical terms, there are basically two spatial representations relied on in commercial GIS: raster and vector. A **raster** representation delineates space as a collection of grid cells and fills the entire space. On a more technical level, a raster layer is defined by first specifying the coordinates on the earth's surface, (x, y), where the initial raster cell is located, as well as its orientation and size. In addition, the number of cells in both the x and y direction must be given. Each cell in a simple raster has one attribute value. As an example, Figure 2.2 indicates ozone (O_3) readings for each 3×3 km grid cell of this layer. If there are multiple attributes, then additional layers would be included for each attribute. Such additional layers for Figure 2.2 could indicate other air pollutants, such as carbon monoxide, hydrocarbons, particulates, and so on. Further, additional layers could contain information on soil types, vegetation content, land use, and so on.

An alternative to raster is the **vector** representation of geographic space. Features in a *vector* representation are described by boundaries and do not

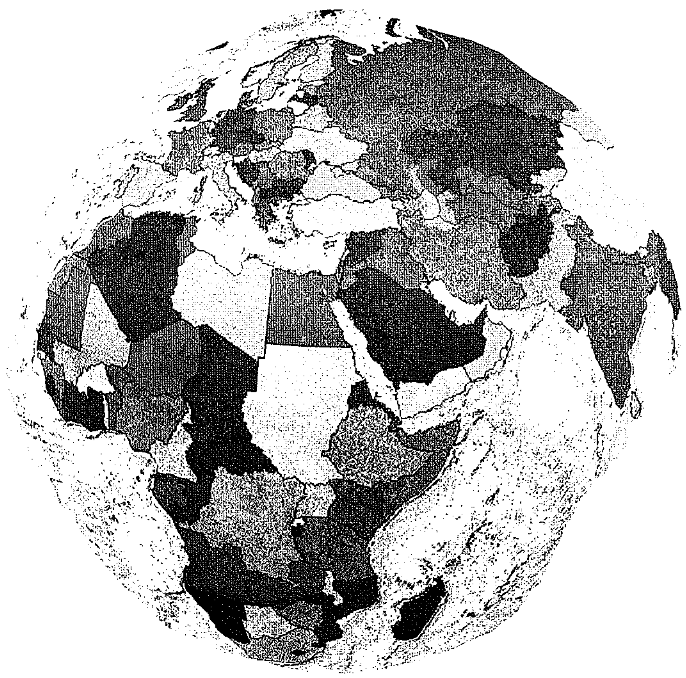
[1] A global referencing approach relies on latitude and longitude measures in degrees from the equator and prime meridian, respectively, to identify any location.



(a) Sphere

(b) Ellipsoid

Figure 2.1 Popular approximations of Earth.



(c) Geoid

Figure 2.1 (*Continued*)

necessarily fill the entire space. Thus, features in a vector representation may be thought of as objects. As an example, Figure 2.3 shows cities, freeways, and parks. Nothing else is defined or identified in areas not corresponding to these features. The vector representation is characterized by three different objects: **point, line** and **area.**[2] Attached to each of these objects could be a range of attribute information. Formally, a point is defined as a coordinate pair, (x, y). Each point in the vector representation is unique and independent of other points. The line object builds on the point because it is mathematically defined using a beginning point and an ending point, as follows: $\{(x_1, y_1), (x_2, y_2)\}$. Of course, there is a functional form describing the shape of the line connecting these two points. Often, points are assumed to be connected by a straight line; a polyline approximates a line that is not straight.[3] The final vector object is an area (or **polygon**). An area may be thought of as a collection of points and

[2] Object orientation has begun to change this characterization somewhat as systems like ArcGIS can handle other features like circles, arcs, etc.
[3] A polyline may be mathematically defined as a series of points, $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\}$, where each consecutive point is connected by a straight line to the previous point.

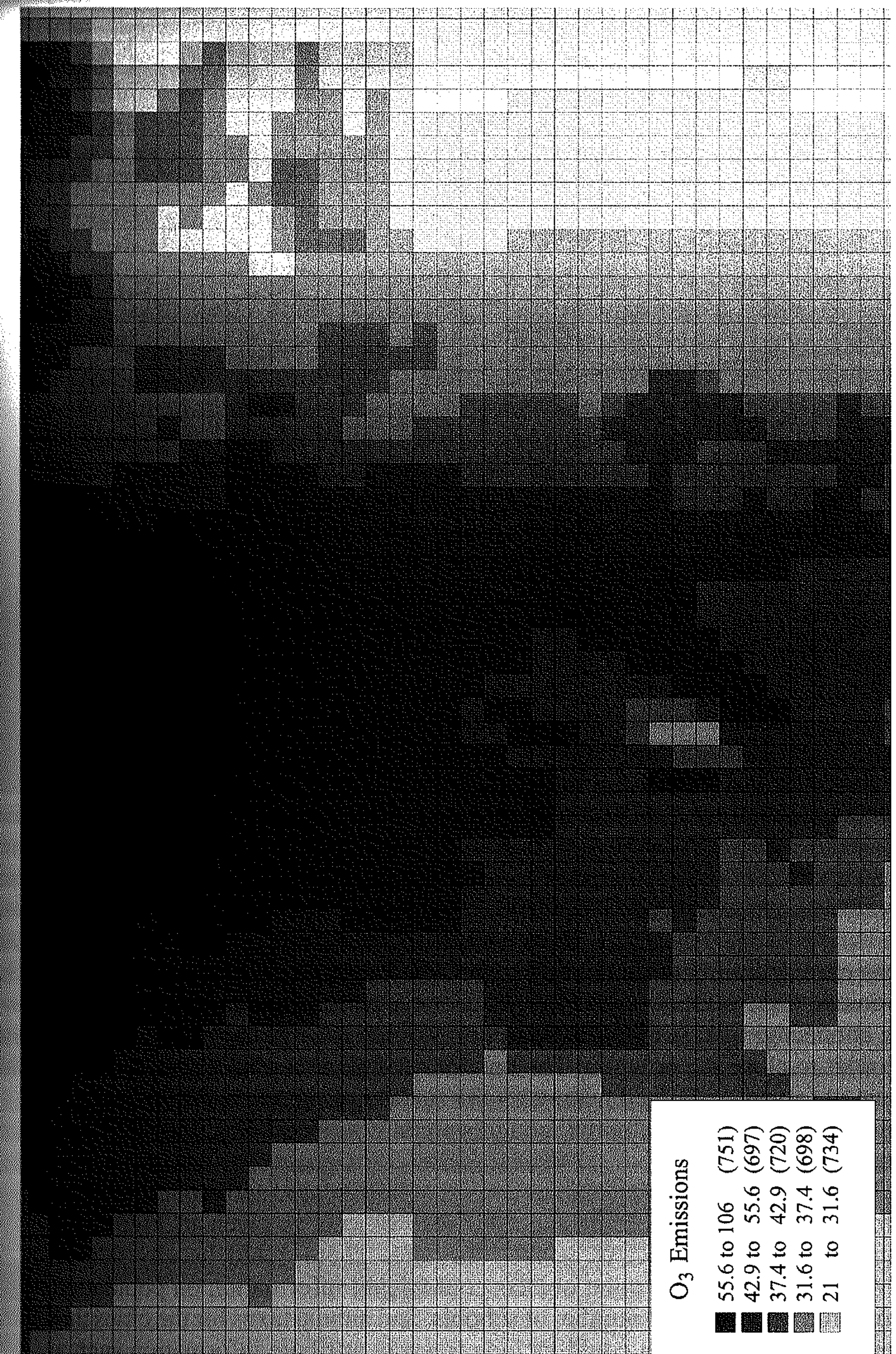


Figure 2.2 Raster representation of space.

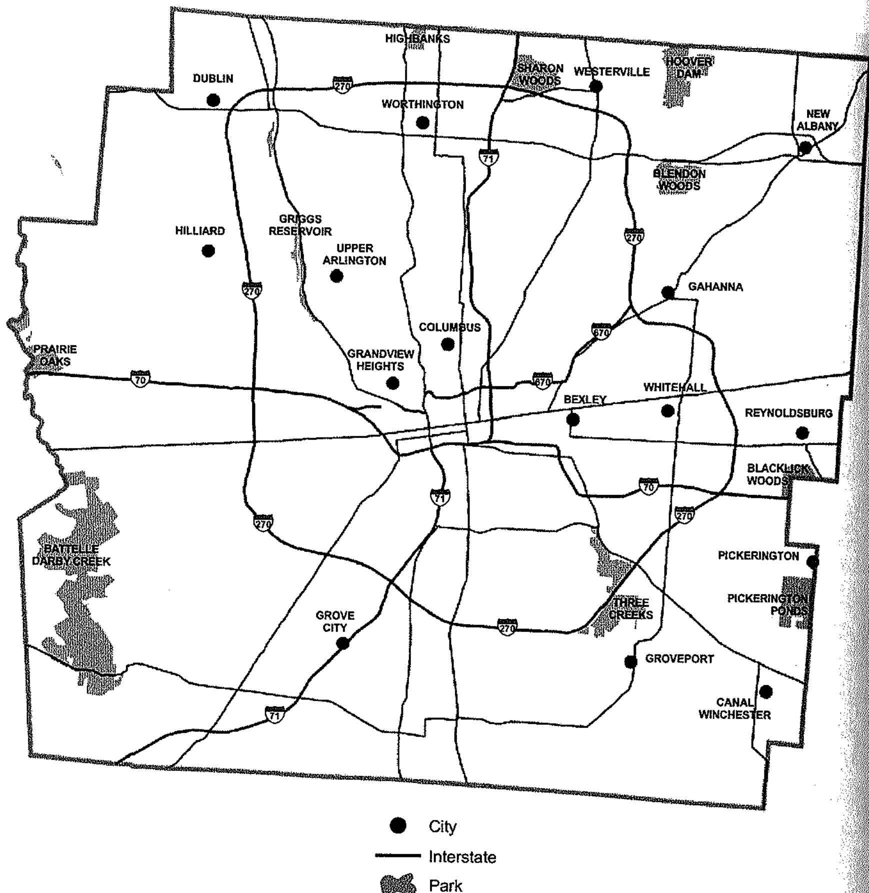


Figure 2.3 Vector representation of space.

lines corresponding to its boundary. In GIS, a polygon is defined by a series of points, $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_m, y_m)\}$, where consecutive points are connected by a straight line and (x_1, y_1) and (x_m, y_m) are the same point in order to close the polygon. The boundary of an area object corresponds to some observed or defined feature on the surface of the earth. As an example, areas could delineate suburbs, congressional districts, lakes, forests, states, countries, and so on.

The point, line, and area objects consist of both spatial and aspatial attribute information. An example of each of these objects is given in Figure 2.3. Attached to each object is a label attribute. There could be any number of other attributes. The point objects could detail information on population, economic performance, number of schools, and annual taxes for the cities they identify. The line object could include freeway attribute information such as the number of lanes, speed limit, year constructed, and so on. The area object

could identify quantities of fauna and flora, recreational amenities, hours of operation, and so on for each park. Thus, a range of attribute information may be attached to each vector object.

There are advantages to be gained from both the raster and vector representations, depending on the context of the analysis being conducted. The raster representation is a simple data structure with many organizational benefits. It is easy to understand and is useful for summarizing environmental characteristics of geographic space. On the technical side, the raster representation does not allow for point and line objects to be incorporated efficiently or accurately. Although data volume may be high due to its space filling property, it is efficiently stored because positions are relative (e.g., represented by row and column indices). For multiple raster layers, certain computational processing can be very efficient. Alternatively, the vector representation is a complex data structure, but it is extremely well suited for handling spatial objects (points, lines, polygons). The data volume can be relatively low for the vector representation, although positional data are required for every object. Computational processing for a vector layer(s) can be cumbersome.

The final digital representation approximation issue was **selective inclusion**. GIS was defined as a system that could be applied to analyze a specific region or subregion in almost any context. Of course, this can only happen if needed data are acquired or created. Selective inclusion has to do with the fact that information in a digital environment is the result of someone making a decision about what to include in a database, or study, relative to what actually exists on or near the surface of the earth. That is, spatial sampling is taking place. Although we commonly conceive of geographic information in terms of layers, not unlike the layers depicted in Figure 2.4, an important question

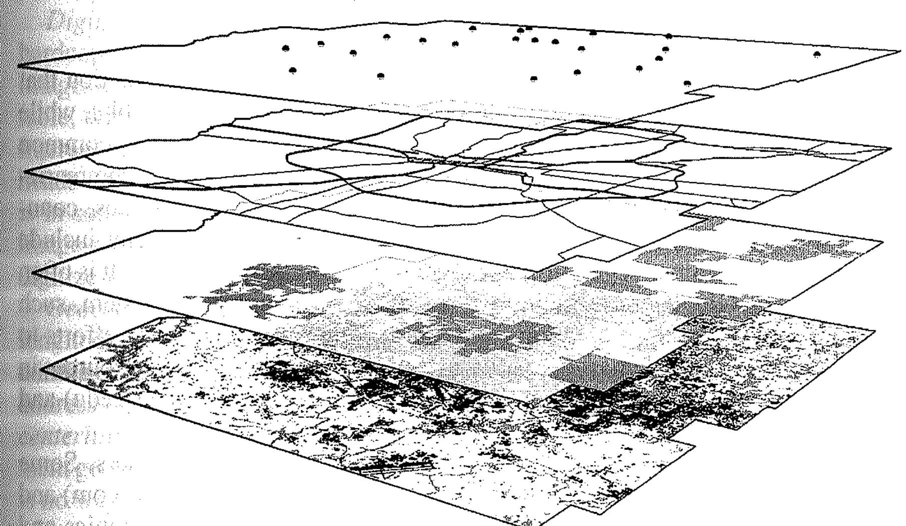


Figure 2.4 Layers of geographic information.

is whether all possible layers of information about Earth are represented. What about within a given layer: Have we omitted any objects or features? Also relevant to the issue of selective inclusion is the temporal dimension of geographic information. Do layers reflect changes across time, or are they simply cross-sectional? Given time constraints, fixed or limited budgets, and knowledge/understanding limitations, selective inclusion is a reality in any application of GIS. However, best-practice use of GIS would suggest that any project would ensure that relevant objects were included and appropriate layers considered.

This section has provided an introductory perspective of GIS. In the sections that follow, we return to the defining components of GIS: acquisition, management, manipulation, analysis, and display. Each component is discussed in detail, establishing a basis for the location analysis and modeling that follows in subsequent chapters of the text.

2.1 DATA ACQUISITION

If not apparent to begin with, a potential GIS user quickly finds out that needed spatial information in a digital form is a critical issue. It is virtually impossible to do any meaningful analysis without supporting spatial information. Possibilities for getting the spatial information needed for any project or study include public data repository (or government agencies), private vendors, collecting and organizing it yourself, or paying (or hire) someone to collect and organize it. In the remainder of this section we review data acquisition sources and/or approaches, enabling the use of GIS.

2.1.1 Existing Sources

There are actually many existing sources of digital spatial information that can be used in a GIS. Some are free or readily available to the public, while others can be purchased from a private vendor. Perhaps the most common source in the United States is Census data (www.census.gov), summarizing at varying scales population characteristics of the nation, states, counties, cities, and towns. Other data produced by government agencies include 100-year floodplains, vegetation classifications, and transportation. It is often the case that GIS vendors include some basic data with their system, such as political boundaries of nations and states. There continue to be efforts to bring together, free of charge, readily available spatial information sources in digital libraries, such as Project Alexandria (www.alexandria.ucsb.edu) and Geospatial One-Stop (geodata.gov).

Alternatively, there are many commercial spatial data providers. Some specialize in transportation data, such as TeleAtlas (www.teleatlas.com) and NAVTEQ (www.navteq.com), while others focus on geodemographics and market research, such as, Claritas (www.claritas.com). Given the importance and significance of digital spatial information in the use of GIS, it is not surprising to see the data provider and service industry emerge and thrive, currently exceeding $5 billion in annual sales and expected to grow substantially in the coming years. Thus, commercial data providers are making profits through the sale and distribution of digital spatial information to users. The point is that while spatial information exists, it may be expensive to acquire.

2.1.2 Semiexisting Sources

By semiexisting sources we mean that the spatial information is not necessarily in a GIS-compatible digital format. As an example, one might have a map from the mid-1800s showing locations where gold has been found in California. Technically, the information about where gold was found exists, but it is contained in a paper map, not as digital information. Another example is a spreadsheet of addresses indicating the residential locations of consumers who have purchased a particular product. Again, the information exists, and in this case it is in a digital format, but it is not amenable for use per se in GIS in order to be evaluated geographically. In both cases, however, it is possible to process this information in order to produce a digital form that can be used in a GIS. We now discuss three basic approaches for processing semiexisting sources of information: scanning, digitizing, and geocoding.

Scanning is the process of converting a hardcopy map into a digital image. Most people are likely aware of scanners for converting text on a page into a digital form. A similar type of process is done for maps, where the scanner is used to detect the presence of information on the map pixel by pixel (or raster cell by raster cell). The resulting scanned image is a digital version of the map, and can be accessed in some way by most GIS.

Digitizing is the process of capturing, or creating, vector objects from hardcopy maps or other geographic information source (e.g., photographs and images). Manual digitizing relies upon a piece of equipment called a digitizing table, where a cursor or puck is used to trace points, lines, or polygons of interest on the map when it is attached to the table. Another approach is heads-up (or on screen) digitizing, when the geographic information source is in some digital form like an image or photograph. Given a scanned map image, as an example, one could import the image into GIS and then manually digitize vector objects in the image.

Geocoding is the process of converting a street address to a latitude and longitude measure on the surface of the earth. A requirement is a database containing records of street segments, the geometry of each street segment, and the address ranges on each side of the street segment. This is often called a *street centerline database*, because the centerline of streets represents the street segment geometry. If a street and address are not found in the database, then there is no way to identify the associated latitude and longitude of the address. An address is successfully geocoded when the street is found in the database and the address location is estimated, thereby identifying the latitude and longitude of that address. With the latitude and longitude, a point on the earth's surface can be found, and it corresponds to the street address. Most commercial GIS packages offer functionality to geocode addresses, and there are commercial vendors specializing in geocoding services. However, successful geocoding is a function of the street centerline database relied on, so difficulties can arise if this database is out of date, inaccurate, or of poor quality.

2.1.3 Surveying and Airborne Approaches

A final data acquisition approach to be discussed is through surveying and airborne approaches. These are included together because they are increasingly interrelated and/or interdependent. We now discuss the following basic approaches for generating spatial information: surveying, GPS, areal photography and remote sensing.

Surveying is an approach for creating vector-based spatial information (points, lines, polygons) by measuring angles and distances from known positional locations. Key here is *known* positional locations, or reference points. The traditional approach to surveying relies on transits and theodolites to measure angles and on measuring tapes to obtain distance. It requires two people to do this. With changing technology, total stations are now more likely to be used to measure angles and distances in surveying. In general, surveying ensures high positional accuracy—down to the millimeter level in some cases. However, it is a time-consuming approach.

GPS (global positioning system) is a satellite navigation system operated by the U.S. Department of Defense, originally designed for military use. It is a constellation of satellites orbiting Earth at about 20,000 km. The satellites have atomic clocks that transmit highly accurate radio signals that can be read by handheld or mounted receivers. This enables position on the earth's surface to be determined, as well as velocity and time, assuming that the receiver is in view of a sufficient number of satellites. Thus, vector information can be created. For example, a receiver could be used to identify the location of a bus stop (point), record the delivery route of a FedEx vehicle (line), or demarcate the catchment area of a watershed (polygon). With differential GPS, signal errors can be corrected, and positional accuracy to the centimeter level is possible. This is done through the use of ground reference stations to adjust GPS readings.

Areal photography is taken from above the surface of the earth, possibly in an air balloon, plane, or helicopter. This produces a digital image (or possibly a photograph that is subsequently scanned into a digital image). A digital image can be georeferenced and then used to derive features or attributes on the earth's surface. For example, heads-up digitizing could be used to create vector objects, such as roads, lakes, rivers, buildings, fields, or forests. Often, high positional accuracy can be achieved, down to the fraction of a meter.

Remote sensing is generally used to create raster-based spatial information. Specifically, sensors mounted on a satellite are used to measure solar energy (electromagnetic radiation), though sensors can also be mounted on planes or helicopters as well. This enables physical, chemical, and biological properties to be derived on or near earth's surface, but requires processing and interpretation of the sensor readings. Spatial and temporal resolution can vary substantially, with some platforms capable of producing measurements for a raster cell a few meters or less in size and other platforms giving a raster resolution of up to 10 km or more in size for an individual cell.

2.2 DATA MANAGEMENT

An important issue for GIS is the capability of efficient storage and quick access to both spatial and attribute information. Storage and access issues have long been the concern of database management systems (DBMS) more generally, but the added complication of managing information in terms of spatial location and proximity has been challenging. As has been said previously, there are effectively two basic data models relied on in GIS, raster and vector. They are, in fact, fundamentally different conceptual and organizational approaches to spatial information management. Thus, we discuss each in turn below.

2.2.1 Raster

As defined earlier in the chapter, the raster data model organizes space through the use of a tessellation, often a regular square cell (or pixel), though others are possible (e.g., triangle, parallelogram and hexagon). For illustrative purposes, Figure 2.5 shows a raster with rows and columns labeled. The raster

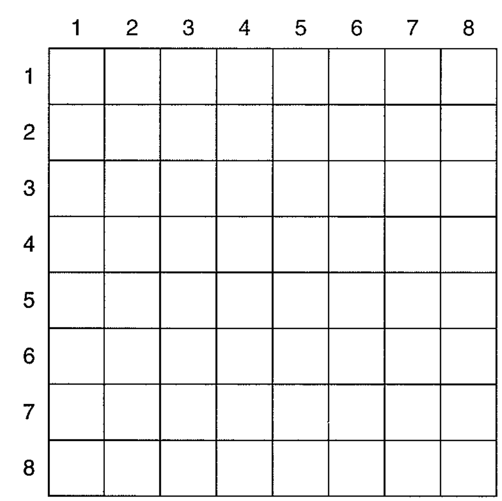


Figure 2.5 Basic raster boundary structure.

layer is positioned through the placement of the cell in row 1, column 1. Thus, we need basic information about cell size (e.g., 30 × 30 m), orientation (e.g., north), and coordinates of this cell. Beyond this, the layer is based on some number of rows and columns of cells, so this must be given as well. With this information, a raster representation for an area of interest is defined. Referring back to Figure 2.5, there are eight rows of cells extending out to eight columns. Given the geographic location on the earth's surface of the initial cell, as well as orientation and cell size, it is simple to determine the geographic location of any of these 64 cells. What is noteworthy about the raster structure is that there is no need to store each cell boundary individually, because there is spatial regularity. Knowing the location of the initial cell, and other supporting details like cell size and number of rows and columns, it is possible to easily construct the spatial geometry of the raster model. This enables considerable storage efficiencies to be achieved.

Of significance with the raster representation is that we know not only where each cell is located, but also the attribute value associated with each cell. Attribute values could be categorical, binary, integer, or real. The attribute values shown in Figure 2.2 correspond to ozone levels in each raster cell, and are reported as real valued. The values of each raster cell must obviously be managed by the GIS, and this is where coordinated linkage with cell boundaries is important. It turns out there are a number of alternatives for an attribute data structure to store and manage raster information. Some are more economical in terms of storage space, while others are more efficient in terms of access and processing speed.

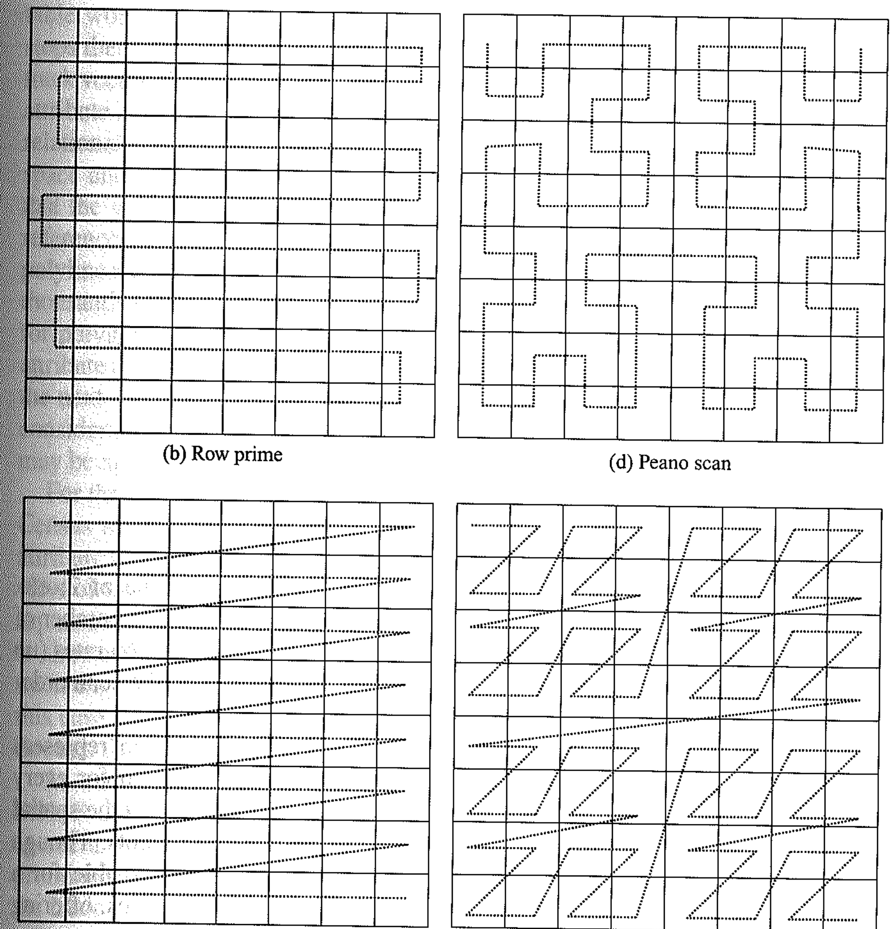
Storing attribute data for a raster model is based on a **scan order** through the raster cells. Four common scan orders are shown in Figure 2.6. A simple scan order is row by row, illustrated in Figure 2.6a. Beginning in the upper-left-hand corner, or row 1, column 1, attribute data would be stored according to the order of the scan. Thus, for row by row, the attribute for the row 1, column 1 cell, would be listed, followed by the attribute for the row 1, column 2 cell, then the row 1, column 3 cell, and so on. If the raster attribute is soil type (see Figure 2.7), then the data would be stored as follows for the row by row order:

1 1 3 3 2 2 2 2 1 1 2 2 2 2 2 2 . . .

The Morton order (see Figure 2.6c) differs in the scan of the cells, beginning with row 1, column 1, then row 1, column 2, next to row 2, column 1, followed by row 2, column 2, etc. For the soil type example in Figure 2.7, this would result in the following attribute data storage:

1 1 1 1 3 3 2 2 3 1 3 1 1 1 2 2 . . .

The reason that alternative scan orders can be important is that some offer greater data compression potential than others, reducing storage needs. This is



(b) Row prime

(d) Peano scan

(a) Row by row

(c) Morton order

Figure 2.6 Alternative scan orders (beginning at row 1, column 1).

achieved through an encoding scheme, such as run-length, block, or wavelet. As an example, run-length encoding replaces the same consecutive attribute value with a count of that attribute. Thus, 1 1 3 3 2 2 2 2 1 1 2 2 2 2 2 2 . . . for the row by row order becomes:

2 1, 2 3, 4 2, 2 1, 6 2, . . .

Attributes, then, are stored in a pairing format, where the first piece of information is an integer count and the second is the actual attribute value for those cells. The run-length encoded row by row scan order, as an example, is interpreted as follows: 2 1 specifies that the first two cells have an attribute value of 1; 2 3 indicates that the next two cells have an attribute value of

1	1	3	3	2	2	2	2
1	1	2	2	2	2	2	2
3	1	1	1	2	2	2	2
3	1	2	2	3	3	2	2
3	3	1	2	3	3	3	3
3	1	1	2	1	3	3	3
3	1	1	2	1	1	3	3
1	3	2	1	1	1	3	1

Figure 2.7 Soil type (1 = clay, 2 = sand, 3 = rock).

3; 4 2 identifies the next four cells as having an attribute value of 2; etc. This is how data compression is achieved, by reducing the actual number of attribute values stored. It is not uncommon to see high compression rates (up to or exceeding a factor of 40) for a raster layer, depending on the scan order and encoding scheme.

The raster data model is useful and important because it can represent a study region in a complete way, as an attribute value is given for every location. Further, it can do this in an efficient manner by handling the spatial location and geometry of cells with minimal storage requirements. Though not discussed here, other data management approaches are possible for a raster layer, such as quadtrees and other hierarchical data structures, offering potential storage and access efficiencies as well.

2.2.2 Vector

In general terms, there are a number of DBMS approaches, including relational, network, hierarchical, object and object-relational. The prominent data management approach in vector-based GIS has been **relational DBMS**, though this has arguably evolved to more of a hybrid object-relational DBMS. Again, issues of efficient storage and quick access to information are critical for the success of any developed system. In fact, such issues have been responsible for the evolution of DBMS in GIS. In the remainder of this section, we review object-relational DBMS to support the spatial representation of vector features.

To begin, a relational DBMS structures information using linked tables. As an example, one table could provide information on the attributes of bus stops, another table could track bus stop maintenance information, and a final table would give the positional coordinates (latitude and longitude) of the stops themselves. The object-relational DBMS differs because of the need to track vector objects. On the one hand, the database generally has considerable attribute information, and it makes sense to manage this using well developed relational DBMS functionality. On the other hand, dealing with geographic space and spatial objects complicates things. One of the reasons for this is that the geometry of vector features varies. For example, a point is always referenced by two pieces of information, a latitude and longitude. However, a polyline may be defined by as few a two points or as many as hundreds, or even thousands of points. It depends on the line feature. A similar situation exists for polygons. The object-relational DBMS approach has evolved to handle attribute data and feature geometry in an integrated way. Though discussed in the next section, the object-relational DBMS approach goes beyond storage considerations as a primary consideration is access to data and operations that may be applied.

For the management and storage of data in a vector GIS, consider the U.S. Census tracts shown in Figure 2.8. These are polygons, and each contains attribute information about the area it represents. Figure 2.9 provides a logical view of the object-relational DBMS approach for managing and storing the aspatial and spatial information in this case. In Figure 2.9a, the attribute table is shown, indicating the various attribute values reported for each Census tract. Figure 2.9b also reports information about each Census tract, but in this case it is the polygon geometry. The linkage of the two files is possible through the use of the feature ID, which is given in each table.

As mentioned previously, the relational DBMS and object-relational DBMS approaches are not the only possible ways to approach the management and storage of information in GIS. However, these have been the prominent approaches to date. Given this, we have limited the focus here to the basic relational/object-relational approaches.

2.3 DATA MANIPULATION

The previous discussion of data management could have included issues of data manipulation, as this is generally considered part of DBMS. We have opted to separate management (storage and access) from manipulation in our overview of GIS because of the unique needs that exist in dealing with spatial information, and the implications for data manipulation. In particular, there are some basic data manipulation approaches for dealing with spatial data. In this section we focus on conversion, aggregation, overlay, and interpolation.

2.3.1 Conversion

A primary feature of GIS is its ability to integrate layers of spatial information. This is only possible by first identifying a common coordinate system. Thus,

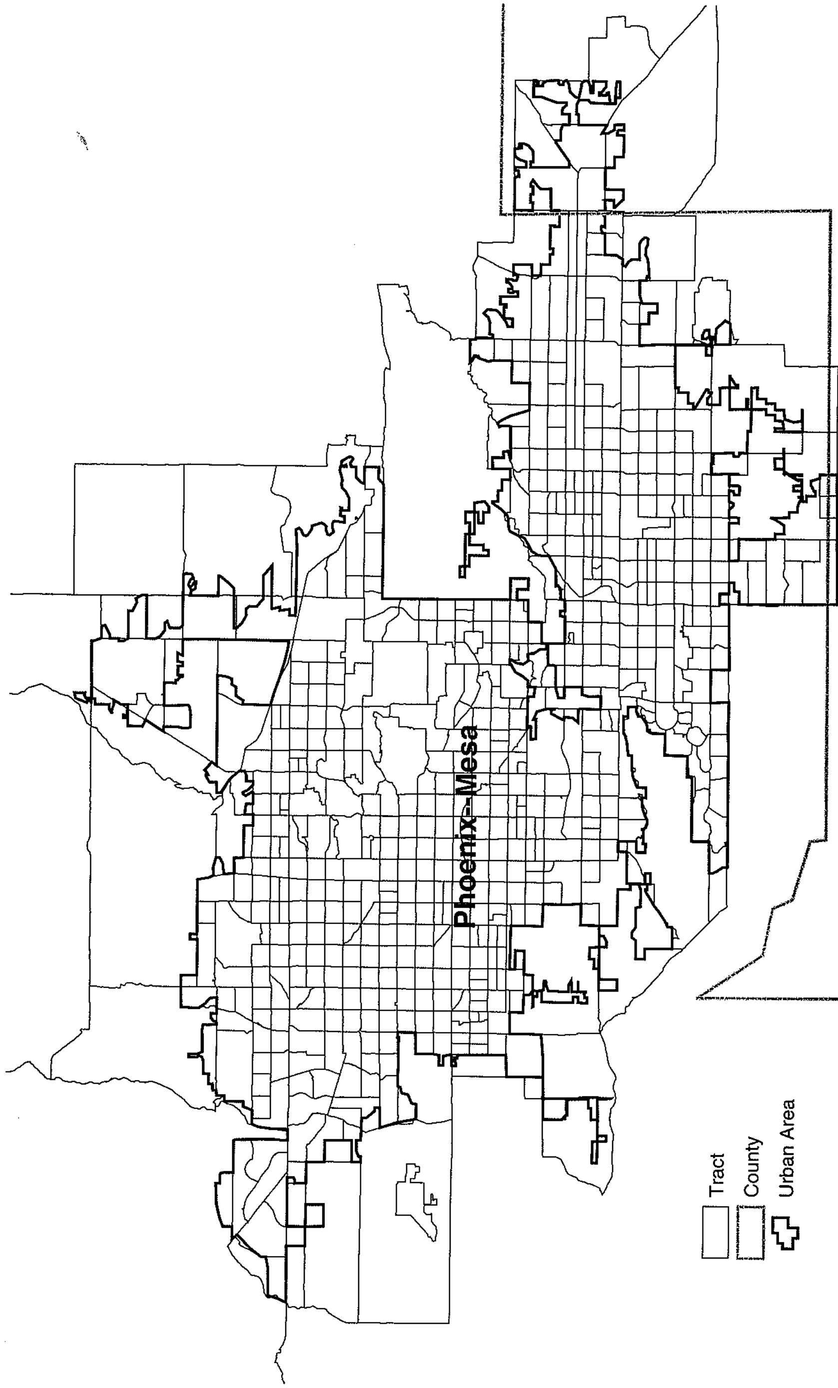


Figure 2.8 Census tracts (polygons) in Phoenix, Arizona.

ID	Shape	Point Data
106200	Polygon	-112.05243 33.59837, -112.03894 33.59571, -112.048 33.58829, -112.06182 33.581, -112.03786 33.59694, ...
106100	Polygon	-112.08128 33.59253, -112.06412 33.58784, -112.09963 33.5821, -112.09573 33.60035, -112.06523 33.60709, ...
106300	Polygon	-112.02458 33.5893, -112.03786 33.59694, -112.01325 33.59692, -112.01309 33.58246, -112.03894 33.59571, ...
105400	Polygon	-112.03465 33.57871, -112.04998 33.585, -112.03894 33.59571, -112.03268 33.58814, -112.0484 33.58224, ...
105300	Polygon	-112.07382 33.57509, -112.06493 33.5826, -112.06509 33.58039, -112.08246 33.57296, -112.08254 33.5822, ...
105200	Polygon	-112.05665 33.57476, -112.0484 33.58224, -112.05003 33.5677, -112.0651 33.58176, -112.06182 33.581 , ...
105102	Polygon	-112.09072 33.57801, -112.08254 33.5822, -112.09964 33.57485, -112.0996 33.58079, -112.09963 33.5821, ...
104501	Polygon	-112.09146 33.57076, -112.08246 33.57296, -112.09961 33.5676, -112.09962 33.57248, -112.09964 33.57485, ...
104502	Polygon	-112.01485 33.55113, -112.02335 33.57219, -112.01599 33.56722, -112.00322 33.56831, -112.03605 33.56369, ...
104700	Polygon	-112.05224 33.5613, -112.04714 33.57112, -112.04329 33.55568, -112.044 33.55529, -112.04805 33.57124, ...
104600	Polygon	-112.07382 33.56028, -112.0652 33.56799, -112.07385 33.55282, -112.08259 33.55291, -112.08254 33.56753, ...
104801	Polygon	-112.09109 33.56028, -112.08254 33.56753, -112.08259 33.55291, -112.09967 33.55309, -112.09961 33.5676, ...
104802	Polygon	-112.05642 33.54579, -112.04864 33.55418, -112.04746 33.53825, -112.06504 33.53828, -112.06521 33.55273, ...
103615	Polygon	-112.09107 33.54574, -112.08259 33.55291, -112.09489 33.53856, -112.09549 33.53857, -112.09967 33.55309, ...
103700	Polygon	-112.07384 33.54557, -112.07385 33.55282, -112.06504 33.53828, -112.08245 33.53836, -112.08259 33.5529, ...
...		

(b) Object definition table

ID	WHITE	BLACK	AMERI_ES	ASIAN	HAWN_PI	OTHER	MULT_RACE	HISPANIC	...
106200	3295	16	11	53	2	37	39	136	
106100	4499	162	99	47	2	225	138	564	
106300	4468	190	110	74	1	310	158	823	
105400	3423	60	37	34	0	106	69	275	
105300	4643	90	63	64	3	619	162	1402	
105200	5243	184	183	124	3	662	216	1569	
105102	4032	33	20	134	4	44	66	216	
104501	2853	134	145	353	10	938	170	2429	
104502	3027	126	218	262	22	1628	291	3144	
104700	6258	226	147	66	3	812	211	2608	
104600	3101	146	107	32	3	588	192	1432	
104801	3574	28	26	62	4	118	87	383	
104802	5817	96	41	124	8	106	120	409	
103615	3572	265	166	126	1	1018	225	2505	
103700	6656	185	84	224	6	367	143	880	
...									

(a) Attribute table

Figure 2.9 Logical view of information management and storage.

a major need in GIS is the capability to carry out **coordinate transformation** tasks, or *conversion*, such as translation, scaling, rotation, and reflection. This means that one or more of the information layers (e.g., Figure 2.4) may not correspond exactly to the common coordinate system being used, and as a result would need to be converted/transformed in some way. That is, the spatial coordinates of features represented would need to be altered through a common translation, rotation, etc. to be consistent with the other data layers.

A second category of data conversion in GIS is *rubber-sheeting*, where a data layer component would need to be stretched and/or pulled in various ways to correspond to or spatially match up with other data. Consider, for example, the integration of the roadway networks in two adjacent counties, where each county is responsible for the creation and management of its network. If our study spans the two counties, then we need to integrate both networks into the same road layer. Rubber-sheeting would be the process of matching up the intersection of streets along the common or shared edge(s), assuming they do not match exactly.

A third major category of data conversion in GIS is *projection*. This is the process of taking the three-dimensional Earth and projecting it into a two-dimension coordinate system—flattening it. There are many possible approaches for projection, but all result in some distortion or error. The reason for this is that it is impossible to represent a three-dimensional object in two dimensions without distortion/error of some sort. More discussion on this topic is left for Chapter 6.

Other data conversion approaches in GIS worth mentioning include clipping, raster/vector conversion, and line thinning and smoothing.

2.3.2 Aggregation

An important feature of GIS is the ability to spatially aggregate information, sometimes referred to as reclassification. In a raster layer this could be accomplished by increasing underlying raster cell size. An example of such **aggregation** is shown in Figure 2.10. The original data for land use type is given in Figure 2.10a, where the cell size is 25 m. The cells are aggregated using GIS, and the result is shown in Figure 2.10b, where cell size is now 125 m. A similar process exists for a vector layer, although it depends on the type of objects represented. For point objects, scale often dictates that some level of aggregation is important. For example, consider the various cities in the Columbus, Ohio, region shown in Figure 2.11a. This view of the different incorporated entities of Columbus makes sense at this scale, but in the context of cities across the eastern United States, as shown in Figure 2.11b, an aggregate representation of the city is more meaningful. Thus, aggregation in this case is the combination of neighboring points. If the data layer were polygon based, aggregation likely would take the form of merging neighboring polygons, or rather dissolving interior boundaries of aggregated polygons. In all cases, however, aggregation involves taking input objects and creating one new output object, combining in some way the associated attribute information.

2.3.3 Overlay

A third major feature of data manipulation in GIS is **overlay**. For convenience, it is assumed that the scale of the spatial layers is the same. Overlay is interpreted a little differently in raster and vector GIS. In a raster system, overlay takes a number of raster layers, each representing a different attribute, and applies logical or arithmetic operations (e.g., and, or, xor, addition, subtraction, multiplication, division, assignment) to produce an output layer. Thus,

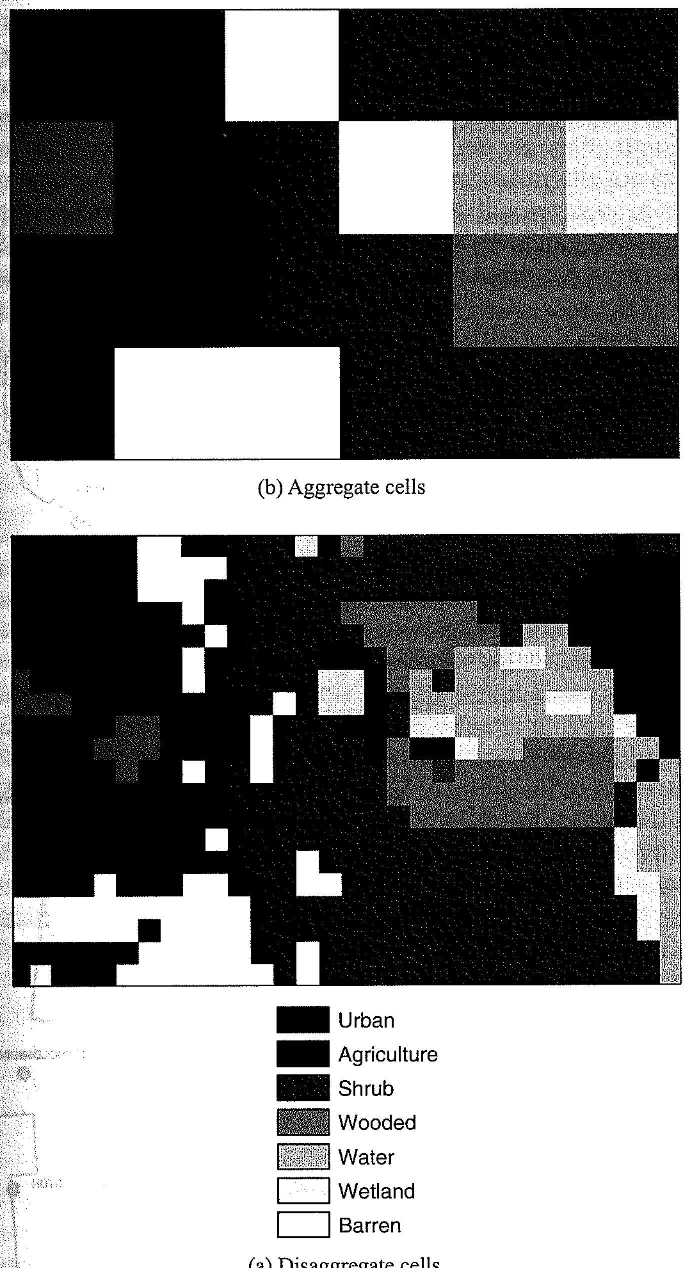
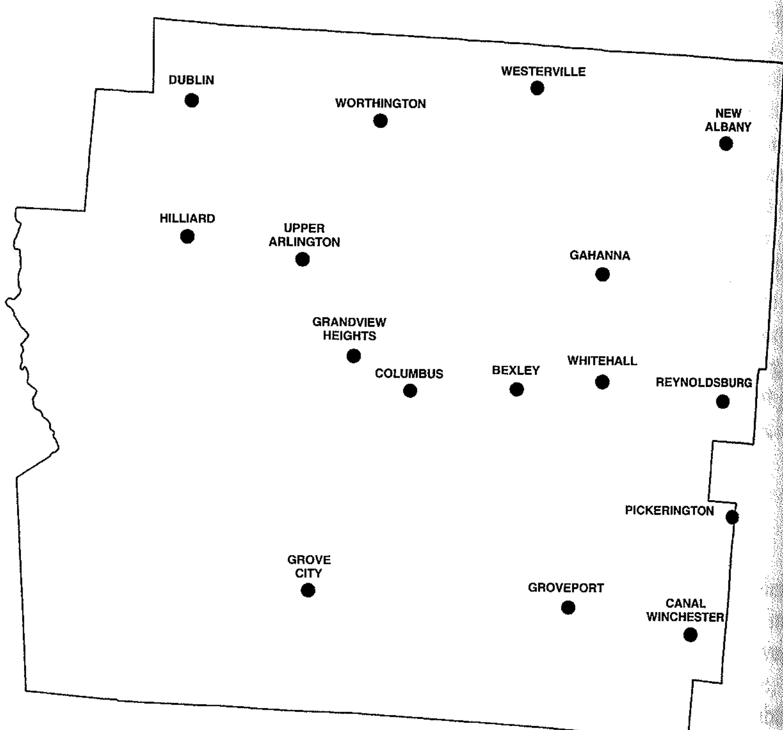


Figure 2.10 Raster cell aggregation.

(b) aggregate Columbus area information



(a) disaggregate local area information (Columbus)

Figure 2.11 Point aggregation.

overlay in a raster system is conditioned upon having the same underlying raster grid cell structure for all input layers. Overlay in raster is equivalent to *map algebra*. We will return to this topic in Chapter 5.

For vector GIS, overlay is more general, often call topological overlay. The objects of input vector layers are considered to not coincide, thereby raising a need to define new objects in a new layer representing the lowest common denominator of the two inputs. An example of this process is depicted in Figure 2.12, where there are two input layers containing different, noncoincidental polygons. The resulting overlay layer will consist of each unique polygon shown in Figure 2.12.

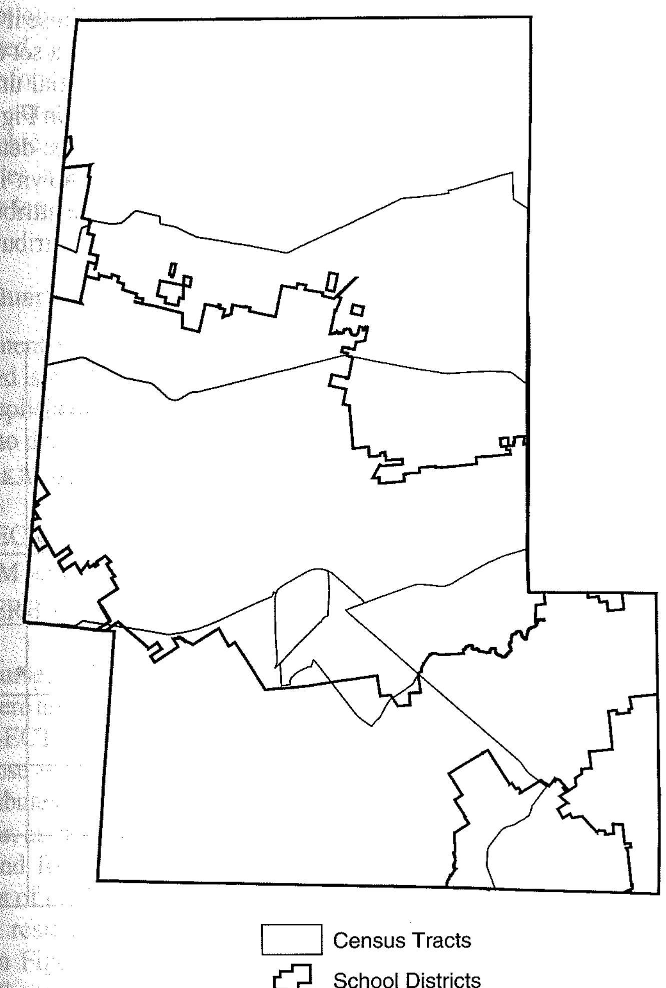


Figure 2.12 Topological overlay.

2.3.4 Interpolation

The final major data manipulation feature of GIS to be discussed is **interpolation**. The definition of interpolation is *a process of estimating attribute values in areas where information cannot be directly acquired*. One classification of interpolation differentiates point and areal. Another classification makes a distinction between exact and approximate interpolation.

A *point interpolation* method estimates the attribute value at specific locations using observed values at other points. As an example, Figure 2.13 gives two layers of information. One is a raster layer for which we wish to know the air quality across space. The second layer is the point-based air-quality monitoring stations. The challenge is to use the sample information (monitoring station readings) for interpolating what the air quality is for each raster cell.

Areal interpolation is different from point interpolation because it basically estimates polygon attributes values (e.g., total population) for a set of spatial units (e.g., school districts) using as input another set of spatial units (e.g., Census tracts). An example is the school districts layer shown in Figure 2.12, where the total population by district is not known and must be determined. Population is, however, known for the Census units layer shown in Figure 2.12. Thus, *areal interpolation* is the process of estimating the attribute value for each polygon in the school district layer using the known attribute values for the polygons in Census unit layer.

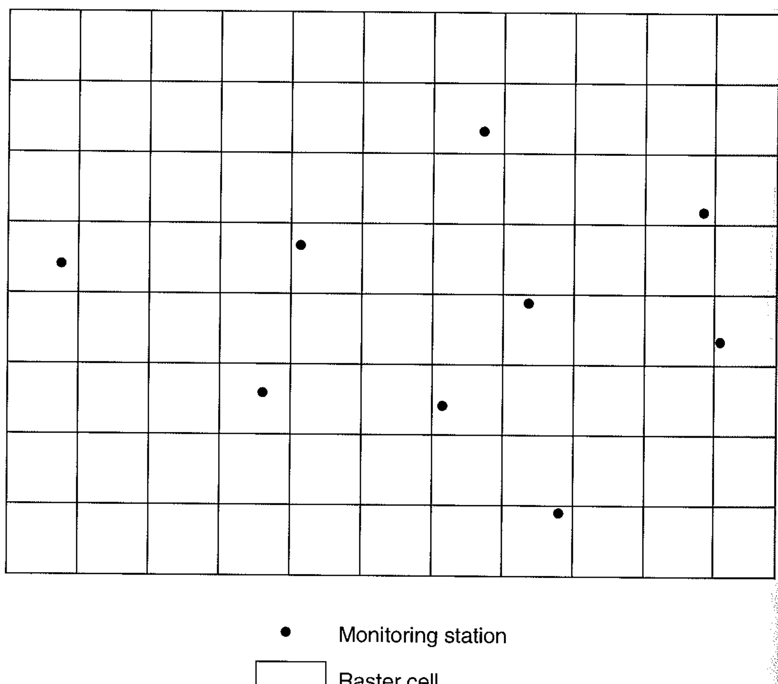


Figure 2.13 Interpolation from known sample points.

A second classification is an exact versus approximate interpolation method. An *exact method* always returns observed values at sample locations (point interpolation) or maintains total polygon attribute values (areal interpolation). *Approximate interpolation* methods, by contrast, may not return or maintain the same observed values. There are many relied upon exact and approximate interpolation approaches, including proximal, inverse distance weighting, kriging, trend surfaces, dasymetric mapping, and pycnophylactic methods.

2.4 DATA ANALYSIS

A perception about GIS is that data analysis capabilities are rather simple and limited, because the visual analysis possible in presenting and mapping spatial information is often very powerful in and of itself. The reality is that GIS actually supports many forms of analysis in addition to visual or cartographic output. In this section, we review the range of data analysis capabilities using GIS, including query, proximity, centrality, and service zone. We leave for the next section an overview of visual analysis.

2.4.1 Query

A prominent DBMS analytical technique applied to attribute data has been query, and is well developed in terms of its use. Specifically, *query* in a DBMS is accomplished using **SQL** (structured query language). In GIS, SQL has evolved to deal with both aspatial and spatial queries. The syntactic structure of SQL is as follows:

```
SELECT <columns>
FROM <tables>
WHERE <conditions>
```

Of course, access to and operational syntax depends on the system being used. There are many operators possible in such queries: AND, OR, EQUALS, INTERSECTS, TOUCHES, WITHIN and CONTAIN, to name a few.

Suppose we were interested in identifying those households with particular attributes, perhaps reflecting potential consumers of a new product. In particular, assume that we want to identify all homes with at least three bathrooms and four bedrooms. This would be an aspatial query because only attributes of residential parcels are of interest. The result of this query would be those residential parcels that meet all imposed conditions, and they are shown in Figure 2.14 for these indicated conditions (parcels with three or more bathrooms and four or more bedrooms).

A spatial query, by contrast, accounts for some spatial relationship. Such a query could be to identify all commercial properties near a freeway onramp,

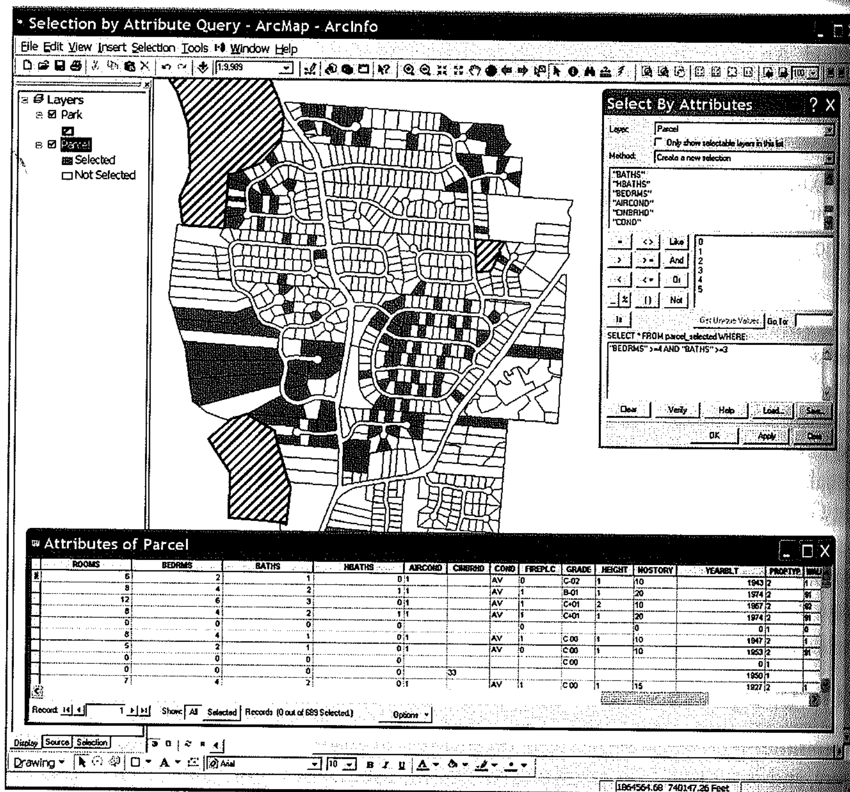


Figure 2.14 Selection by attribute query.

finding all residential properties close to a lake or river, or highlighting in the database all objects in a specified geographic area. An example of a user-defined spatial query is shown in Figure 2.15, indicating those parcels within 700 feet of parks.

2.4.2 Proximity

Given that GIS is specifically designed to manage information corresponding to something on the surface of the earth, it seems intuitive that an analytical capability in GIS would be to determine the proximity between locations. This includes distance but also topological relationships, such as adjacency, contiguity, intersection and connectivity, and object buffers.

Distance is a measure of proximity between two given locations, and can be defined mathematically. In GIS, we can conceive of two fundamental types of distance: global and planar. A global distance measure takes into account the curvature of the earth (three-dimensional), whereas a planar measure is limited to two-dimensional space.

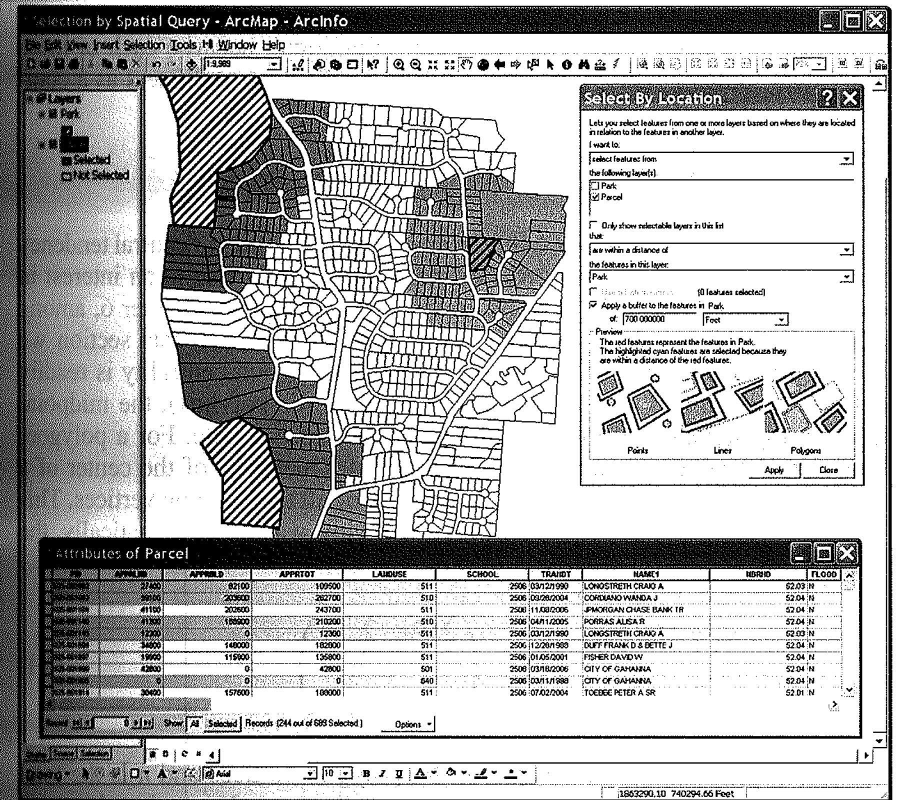


Figure 2.15 Selection by spatial query.

Important topological relationships are adjacency, contiguity, intersection, and connectivity. Two objects are adjacent if they are next to each other. For example, if two polygons share a common boundary, they are said to be adjacent. An area is contiguous if it is possible to travel/move between all parts of the area without leaving the area. Such an area can comprise a number of polygons, so the question in this case is whether the polygons are mutually adjacent to each other. Intersection is whether two objects meet or overlap in some way, and where this occurs. As an example, do two roads (or line segments) intersect? Connectivity is somewhat related to contiguity in the sense that it reflects movement between two locations. Often, connectivity is considered in networks, and whether a path exists from one node to another node. Finally, a buffer is a topological transformation of an input object (point, line, polygon, or raster cell) using a specified width or travel time, and can be regular or irregular in shape. As an example, a regular width buffer for a point is a circle having a radius equal to the width of the buffer. Similar conceptual extensions result when buffering lines, polygons, or raster cells.

Admittedly, the formal definitions of proximity-based data analysis approaches have been brief. This has been intentional, because each of these individual topics is elaborated on in different chapters of the text as a component of GIScience issues in location modeling.

2.4.3 Centrality

Centrality is often associated with statistics and the notion of central tendency in a distribution. Spatial information necessarily gives rise to an interest in central location, not unlike its use in statistics. Although Chapter 6, among others, will get more to the heart of central locations, in this section we focus on the centrality of an object. For a point object, centrality is readily defined. It is simply the point itself. For a line (or polyline), the midpoint is generally considered the center, and is easy to compute. For a polygon, however, the center is not so straightforward. One view of the center of a polygon is its *centroid*, or, rather, the average of all the polygon vertices. This is fairly easy to compute and intuitive in many ways. Mathematically, the centroid is

$$(\bar{x}, \bar{y}) = \left(\frac{\sum_i x_i}{n}, \frac{\sum_i y_i}{n} \right) \tag{2.1}$$

where (x_i, y_i) are the coordinates of the polygon vertices and n is the total number of vertices. Often, a centroid is inside the polygon, but this need not be the case in general. Furthermore, the density of vertices defining the polygon greatly influences the location of the centroid. For these reasons, other views of centrality for a polygon are needed. One alternative is the **skeleton**, or **medial axis**. Formally, the skeleton of a polygon is the locus of the centers of circles that are tangent to the polygon at two or more points, with all such circles being contained in the polygon. For the polygon shown in Figure 2.16a, the associated skeleton is given in Figure 2.16b. It is worth noting that the skeleton has proven useful in location analysis, with important spatial properties.

2.4.4 Service Zone

A *service zone* is an identified, or identifiable, area that has some significance to a location-specific entity. Here we review three types of service zones: proximal area, catchment area, trade area, and viewshed.

A *proximal area* refers to the geographic space served by its closest facility. GIS supports the identification of catchment or trade areas through creation of **Thiessen polygons**, also referred to as a **Voronoi diagram** (as well as Dirichlet and proximal polygons). Formally, Thiessen polygons, given a set

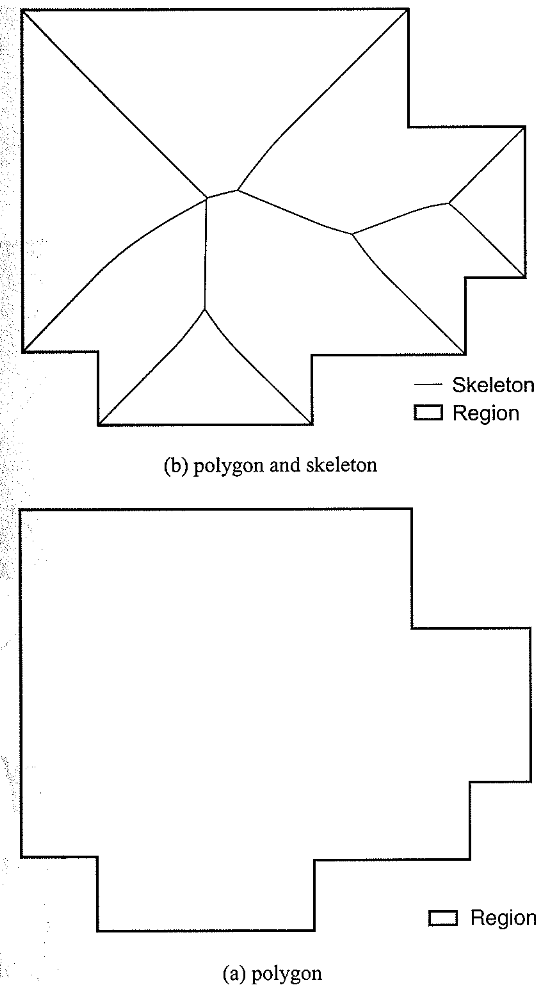


Figure 2.16 Skeleton of a polygon.

of generator points, are those polygons defined such that any location (x, y) is associated with its closest generator point. Polygon boundaries are defined by the half-plane of neighboring generator points. As an example, Figure 2.17 shows the Thiessen polygons for the indicated set of generator points.

A *catchment area* is the area serviced by a facility (e.g., school), but could also be the area inhabited by some species (e.g., spotted owl, kit fox, goshawk, etc.). In the context of facility service, it could be equivalent to the proximal area, but might differ as well. A common interpretation of a catchment area in physical geography is spatial extent associated with natural boundaries such as a watershed.

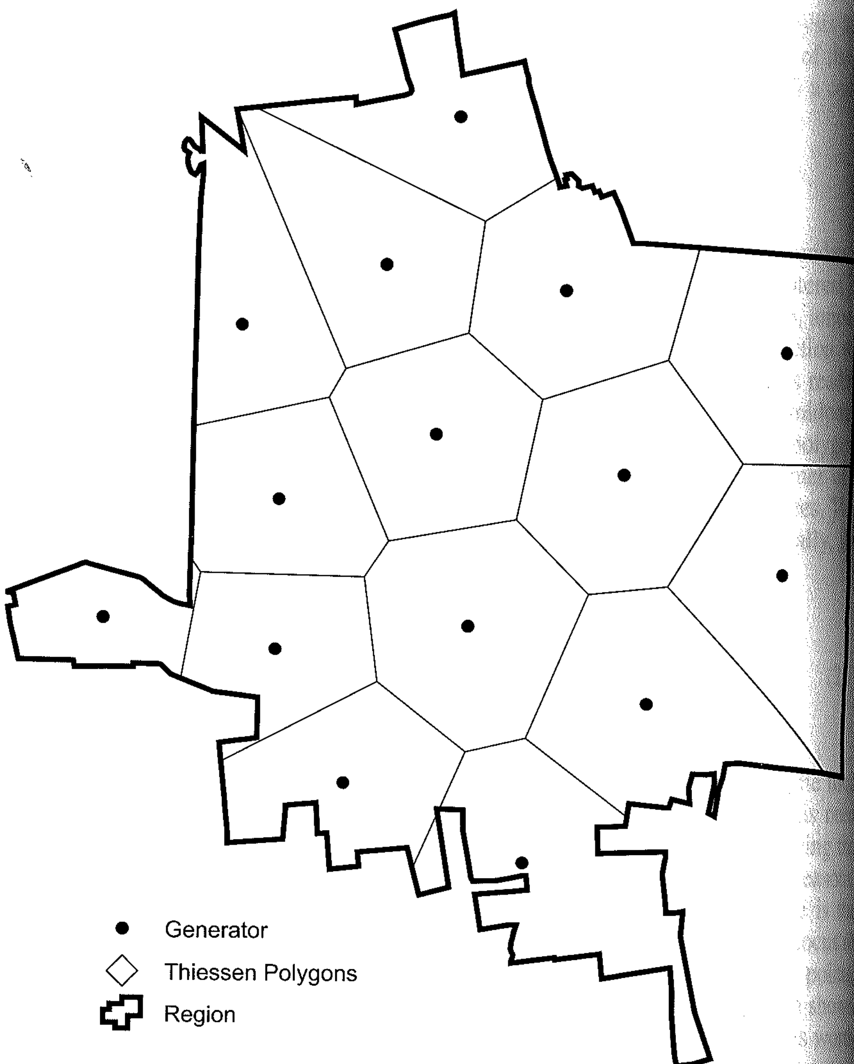


Figure 2.17 Thiessen polygons for the set of generator points.

A related service zone is the *trade area*. Commonly referred to in business and retail geography, a trade area is the geographical area containing customers (or potential customers) served by a business, firm, retailer, or shopping center. There are many methods proposed for identifying and/or delineating trade areas, including radial rings, drive time, Thiessen polygons, and gravity-based models. Further discussion is left for Chapter 4.

Another type of service zone is the *viewshed*. A viewshed is the surface area visible from one or more observation points. As a service zone, the

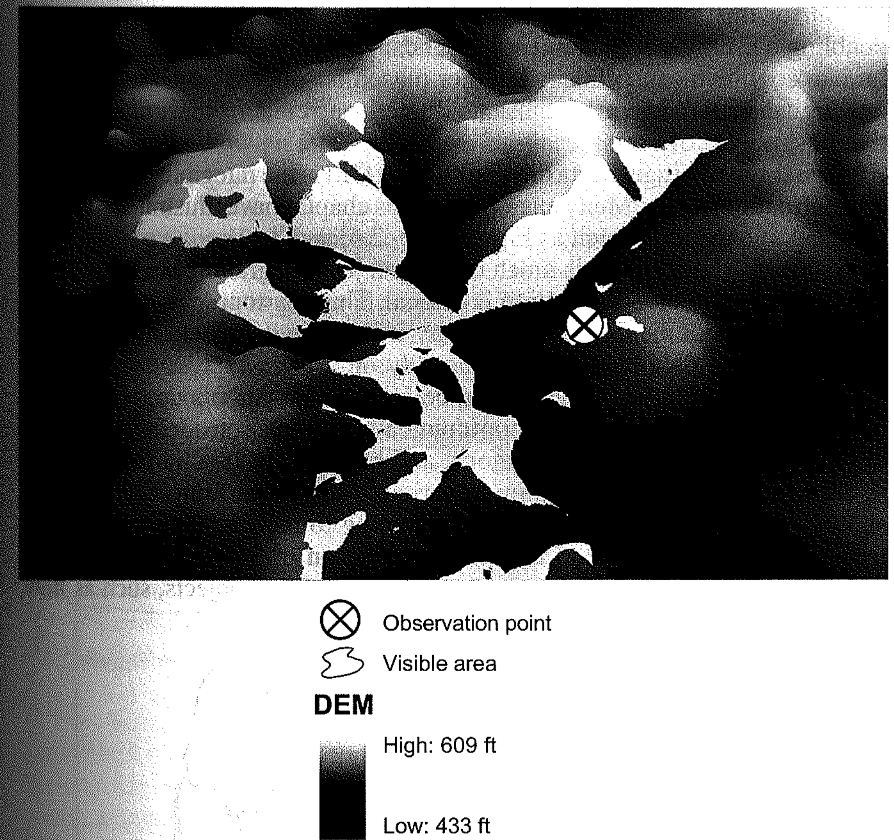


Figure 2.18 Viewshed from an observation point.

viewshed could represent the area capable of observation from a lookout tower, or watchtower, or even surveillance equipment. If the lookout tower is used for forest-fire spotting, as an example, then the viewshed is the area capable of being monitored. Figure 2.18 shows the viewshed associated with the indicated observation point.

The service zone is a particularly important GIS analysis capability that supports location analysis and modeling. As we will see throughout the text, many models require potential service zones as input.

2.5 DATA DISPLAY

As mentioned previously in this chapter, data display has been a defining characteristic of GIS. In particular, GIS has come to be synonymous with the visualization and output of spatially referenced information. Part of the reason for this is that visual display of geographical data is very powerful, enabling a user to observe and infer trends, relationships, and so on based on human cognitive capabilities. Another reason is simply that GIS technology and digital spatial information has facilitated map creation and output by even the most novice of users.

Arguably, the biggest data display asset of GIS is the integrated presentation of information layers. Many examples in this chapter have illustrated this layer integration capability, such as Figures 2.3, 2.4, and 2.11b.

Of course, data display is much more than just integration of spatial layers because effective communication is critical. Thus, cartography is particularly important in data display, which is why it is its own subfield in geography. Cartography in data display relies heavily on artistic expression. Considerations such as visual representation, color choice, symbols used, and so on are important factors in good communication. Such considerations impact the display of points, lines, polygons, and raster cells. It is, in fact, possible to depict point-based information in various ways. For example, Figure 2.19 shows symbols varying in size, corresponding to their relative total population. This can be contrasted to their presentation in Figure 2.11b. It is also possible to show variation relative to attributes of line objects, such as flow,

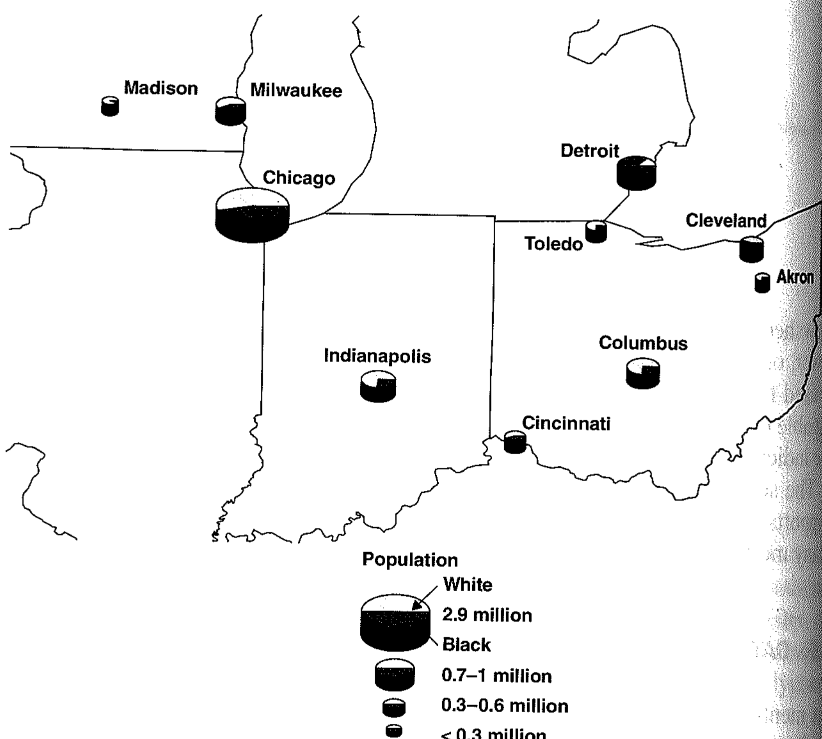


Figure 2.19 Visual display of point attributes.

direction, and congestion. As an example, Figure 2.20 illustrates the amount of flow along transportation segments of an interstate system.

Not surprisingly, polygon objects, or areas, can also communicate associated attribute information in various ways. As with points and lines, visual display can show how an attribute varies in space. Perhaps the most prominent area based display of spatial variation is through the use of a **choropleth map or output**. A choropleth map attempts to classify attribute values of polygons into a specified number of simplifying classes. Often, 4 to 7 classes are recommended for effective comprehension by a user. An example of a choropleth display is given in Figure 2.21. In this case, five classes are shown for total population (2000) in each Census tract. The classes, then, are meant to summarize variation and facilitate understanding. There are, in fact, a number of standard choropleth classification approaches, including natural breaks, quantile, equal area, equal interval, and standard deviation.

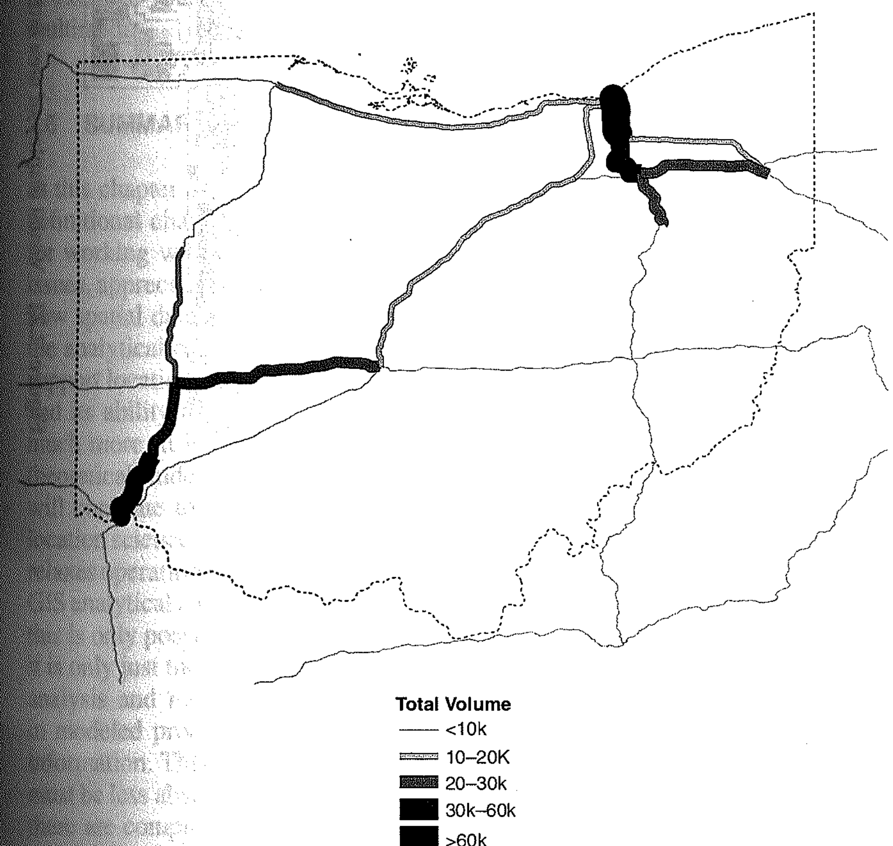


Figure 2.20 Visual display of line attributes.

Figure 2.21 Visual display of area attributes.

All are based on taking the attribute range and dividing it up in some way, where each class is then defined by so-called break points along the attribute range. The natural breaks approach defines a prespecified number of classes based on minimizing within group variance. The quantile approach creates a predetermined number of classes such that each contains an equal number of observations. The equal area approach defines classes so that each class contains approximately the same total area. The equal interval approach divides the attribute range into so that each class has an equal width. Finally, the standard deviation approach defines classes by standard deviation from the mean attribute value.

Our discussion of data display would not be complete without giving some attention to current trends, and many have potential implications for location analysis and modeling. Perhaps the more general trend is the notion of **GeoVisualization**, where the idea is to expand on how human and physical information is communicated by enabling exploration, synthesis, presentation, and analysis. This can involve effective 3-D representation, the ability to examine space–time relationships, animation of change, and the use of multimedia.

2.6 SUMMARY

In this chapter we have attempted to provide an overview of GIS based on definitional characteristics. In doing this, the intent was to establish a basis for working with spatial information, recognizing geographical abstraction issues, appreciating data reliability and quality considerations, understanding how spatial data is managed and manipulated, and having knowledge about the analytical capabilities of GIS. Thus, while GIS offers much potential to support location analysis and modeling through the use of the data it manages and its ability to display and visualize modeling results, GIS represents so much more. It is only through the understanding of this technology and its theoretical underpinnings that business site selection and location analysis will continue to mature and advance. In fact, much of current research in location science explores the changing theoretical structure of certain models, relaxes operational assumptions, addresses issues of data uncertainty, and uses GIS analytical capabilities to aid in the solution of some location models. Yet, this is only possible through the integration of location science and GIS, and it is only just the beginning. Of course, GIS too has implications for location analysis and modeling. One issue is that GIS gives rise to greater detail in modeled processes because of the availability of a finer level of spatial information. This increases user and decision maker expectations, so models must be less abstract. Such increased detail comes at a cost, however, because there are computational impacts and they are significant. Greater detail and specificity means larger planning optimization problems and the need for more advanced and sophisticated solution techniques.