Basic Concepts of GIS

Some Concepts That Underpin GIS

OVERVIEW

IN WHICH you begin to understand the rather large and complex body of ideas and techniques that allow people to use computers to comprehend and design the physical environment. And in which you use Esri's ArcCatalog to explore geographic data.[1]

You Ask: "What Is GIS About?"

A poem "The Blind Men and the Elephant" tells the story of six sightless men who approach an elephant, one by one, to satisfy their curiosity.

> It was six men of Indostan
> To learning much inclined
> Who went to see the Elephant
> (Though all of them were blind)
> That each by observation
> Might satisfy his mind.
>
> The First approached the Elephant,
> And happening to fall
> Against his broad and study side,
> At once began to bawl,
> "God bless em! But the elephant
> Is very like a WALL!"
>
> The Second, feeling of the tusk
> Cried: "Ho! what have we here
> So very round and smooth and sharp?
> This wonder of an Elephant
> Is very like a SPEAR!"
>
> The Third approached the animal,
> And, happening to take
> The squirming trunk within his hands,
> Thus boldly up and spake:
> "I see," quoth he, "the Elephant,
> Is very like a SNAKE!"

---

[1]Esri is the Environmental Systems Research Institute. Esri makes the software, ArcGIS, which you will use in this text to understand the concepts of geographic information systems.

The Fourth reached out an eager hand,
And felt about the knee
"What most this wondrous beast is like
Is mighty plain," quoth he:
"'Tis clear enough the Elephant
Is very like a TREE!"

The Fifth, who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can,
This marvel of an Elephant
Is very like a FAN!"

The Sixth no sooner had begun
About the beast to grope,
Than seizing on the swinging tail
That fell within his scope,
"I see," quoth he, "the Elephant
Is very like a ROPE!"

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong
Though each was partly in the right
And all were in the wrong!

[. . .]²

Excerpted from "The Blind Men and the Elephant"
(based on a famous Indian legend)
John Godfrey Saxe
American Poet (1816–1887)

And So You Ask *Again*: "What Is GIS About?"

Poet Saxe's lines could apply to geographic information systems (GIS) in that relating to the subject may well depend on your point of view. Asking what GIS³ is about is sort of like asking "What is a computer about?" The capabilities of GIS are so broad and its uses so pervasive in society, geography, urban and regional planning, and the technical world in general that a short, meaningful description is impossible. But for starters, here is a generic definition of GIS that you might find in a dictionary:

A geographic information system is an organized collection of computer hardware and software, people, money, and organizational infrastructure that makes possible the acquisition and storage of geographic

---

²[. . .] indicates an omission.
³In this text "GIS" stands for "a geographic information system," or for the plural "geographic information systems," depending on the context.

and related attribute data, for purposes of retrieval, analysis, synthesis, and display to promote understanding and assist decision making.

To better understand one facet of GIS, consider how you might use the technology for a particular application. Solve the following site selection problem:

Finding a Geographic Site by Manual Means

Wildcat Boat Company is planning to construct a small office building and testing facility to evaluate new designs. They've narrowed the proposed site to a farming area near a large lake and several small towns. The company now needs to select a specific site that meets the following requirements:

❏ The site should not have trees (to reduce costs of clearing land and prevent the unnecessary destruction of trees). A regional agricultural preservation plan prohibits conversion of farmland. The other categories (urban, barren, and wetlands) are also out. So, the land cover must be "brush land."

❏ The building must reside on soils suitable for construction.

❏ A local ordinance designed to prevent rampant development allows new construction only within 300 meters of existing sewer lines.

❏ Water-quality legislation requires that no construction occur within 20 meters of streams.

❏ The site must be at least 4000 square meters to provide space for building and grounds.

Figure 1-1 is a key to the following maps. It shows the symbols for land cover, soil suitability, streams, and sewers.[4]

Figure 1-2 is a map showing landcover in the area from which the site will be chosen. Different crosshatch symbols indicate different types of land cover; the white area in the northern part of the map is water. The land cover codes (LC Codes) and categories (LC Type) are as follows:

LC CODE	LC TYPE
100	Urban
200	Agriculture
300	Brush land
400	Forest
500	Water
600	Wetlands
700	Barren

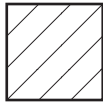[4]These maps are also available on the DVD that accompanies the text. They are the image files: Key_to_maps.jpg, Landcover.jpg, Soil_suitability.jpg, and Streams_&_Sewers.jpg located in the folder IGIS-Arc_AUX.
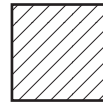
Keys to Maps

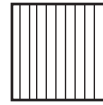Soil Suitability   Land Use
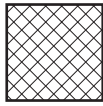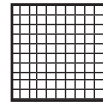
 Poor suitability   Urban

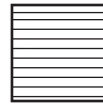 Suitable   Agriculture

 Brushland

 Forest
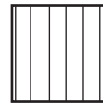
 Wetlands
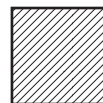
 Barren

Streams and Sewers

Light solid lines: Stream
Heavy solid lines: Sewers
Light dashed lines: Coastline (and border)

FIGURE 1-1   Key to maps of the Wildcat Boat Facility area

Figure 1-3 is a soil suitability map. Lines separate soils of different types. The different soils are categorized as suitable or unsuitable for building. Therefore, you will see the same symbol on both sides of a dividing line, indicating that, while such soil types may be different, their suitability is the same.

Figure 1-4 is a map that shows the streams (narrow lines) and sewers (broader lines).
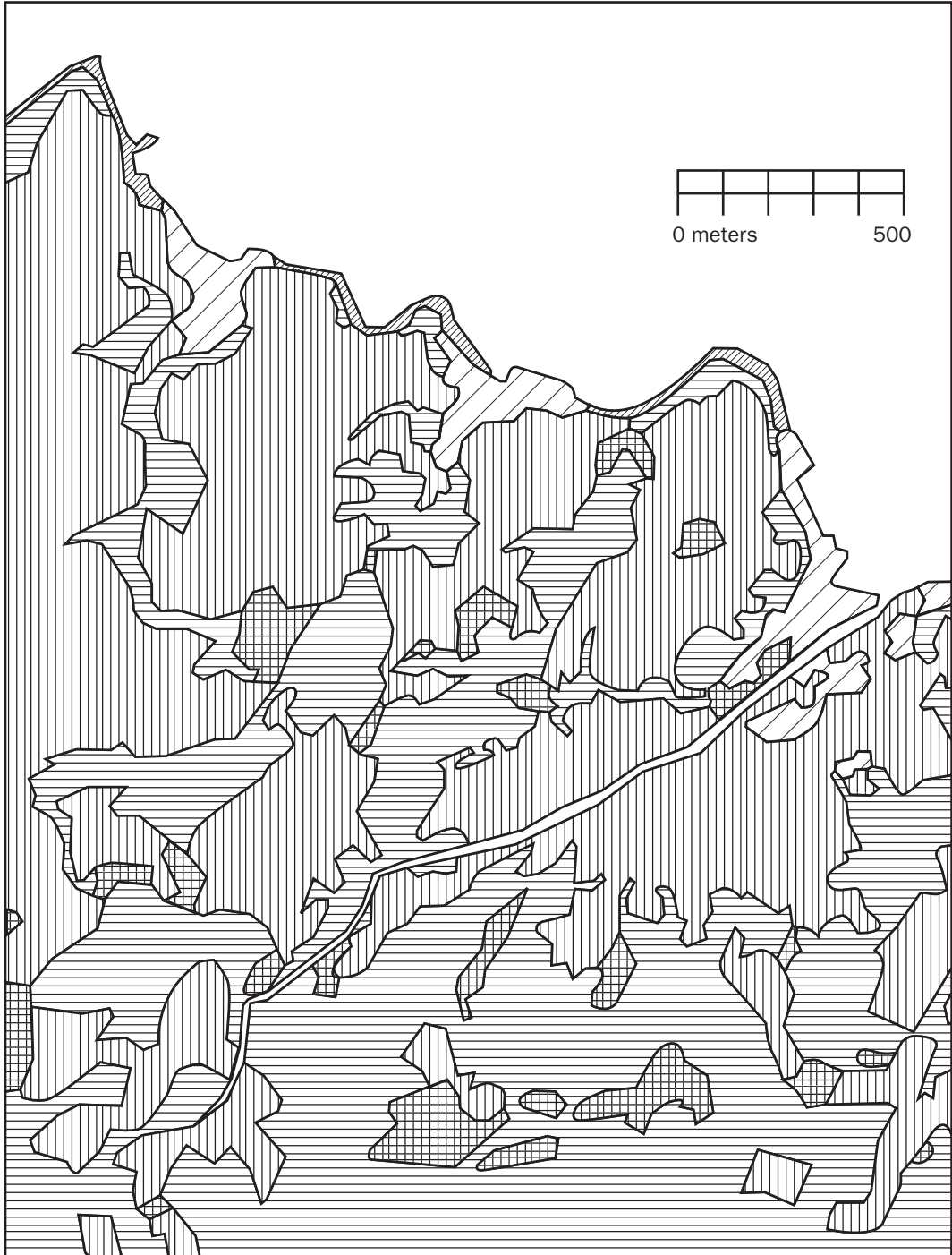
FIGURE 1-2   Land Cover
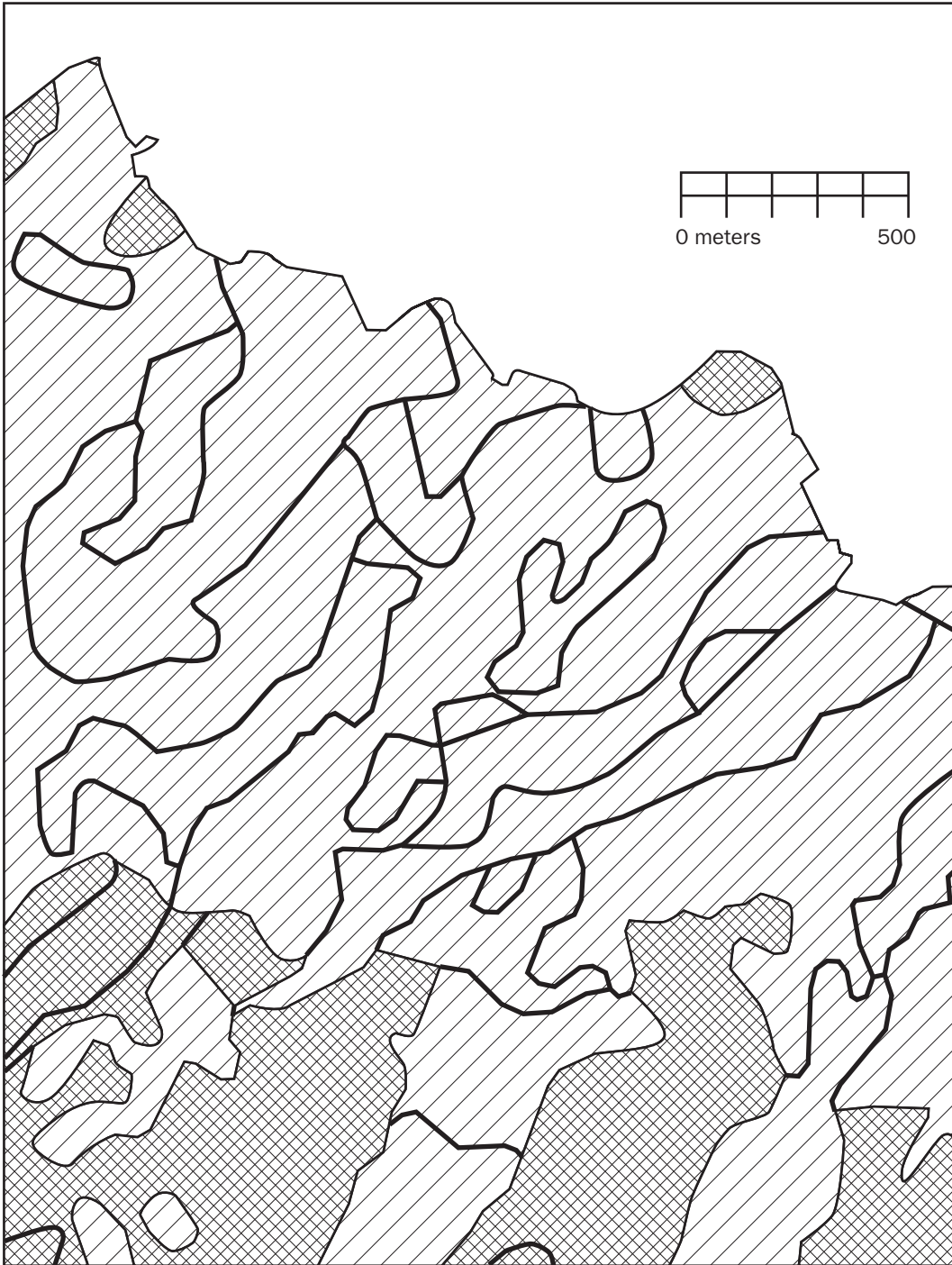
FIGURE 1-3   Soil Suitability
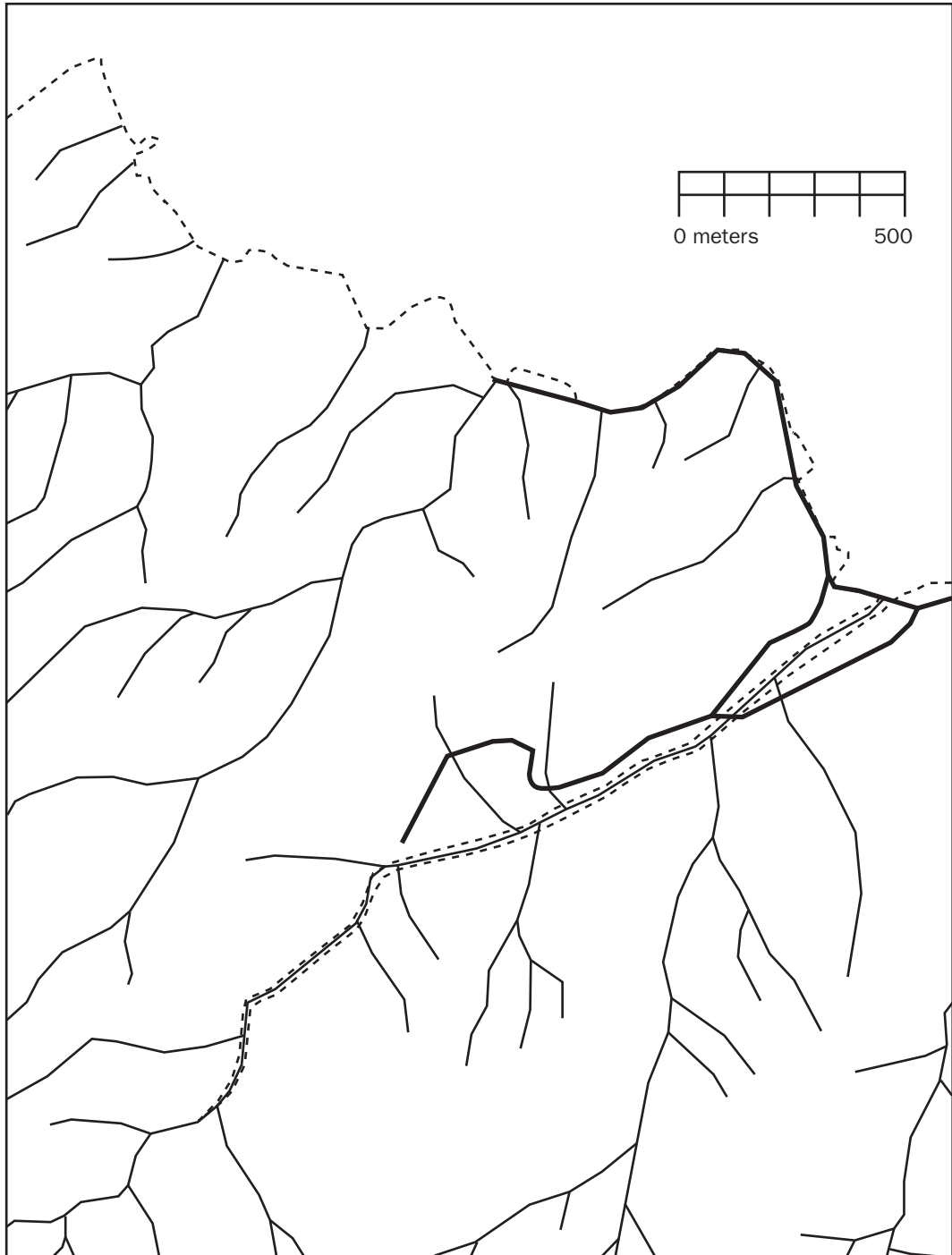
FIGURE 1-4   Streams and Sewers

You may use scissors, xerography, a computer-based drawing program, a light table, and any other tools to solve the problem.

You are asked to present a map that shows all, repeat ALL, the areas where the company could build, while meeting the requirements stated previously. Make your map the same scale and size as those maps provided on the DVD. Outline in red all the areas that meet the requirements. You don't need to produce a high-quality cartographic product. The main objective here—indeed the object of this textbook—is to analyze geographic data. While making maps is important, it is not the primary focus of this book.

Write a brief description (100 to 200 words) of the procedure you used to make the map.

The problem is much easier than it might otherwise be because the maps provided cover exactly the same area, have the same underlying assumptions regarding the shape and size of Earth, are at the same scale, and use the same projection of Earth's sphere onto the flat plane of the map. These benefits are often not available in the real world, where you frequently need a considerable amount of data preparation to solve such a problem. ArcGIS has many tools to aide in "lining up" geographic data. Despite these advantages, the process can be somewhat daunting.

More of What GIS Is About

Completing Exercise 1-1 showed how GIS can help you solve one kind of problem. There are many others. Computer-based GIS not only serves the purpose of traditional maps but also helps you perform activities that involve spatial analysis, even without maps. Understanding conditions that occur in the vicinity of Earth's surface are important in building structures, growing crops, preserving wildlife habitat, protecting ourselves from natural disasters, navigating from one point to the next, and a myriad of other activities.

Among the many uses of GIS are:

Land use—Helps determine land uses, zoning, environmental impact analysis, locational analysis, and site analysis.

Natural environment—Identifies, delineates, and manages areas of environmental concern, analyzes land-carrying capacity, and assists in developing environmental impact statements.

Energy—Examines costs of moving energy, determines remaining available energy reserves, investigates the efficiency of different allocation schemes, reduces waste, reduces heat pollution, identifies areas of danger to humans and animals, assesses environmental impacts, sites new distribution lines and facilities, and develops resource allocation schemes.

Human resources—Plans for mass transit, recreation areas, police unit allocation, and pupil assignment; analyzes migration patterns, population growth, crime patterns, and welfare needs. It also manages public and government services.

Areas of environmental concern—Facilitates identification of unique resources, manages designated areas, and determines the relative importance of various resources.

Water—Determines floodplains, availability of clean water, irrigation schemes, and potential and existing pollution.

Natural resources—Facilitates timber management, preservation of agricultural land, conservation of energy resources, wildlife management, market analysis, resource allocation, resource extraction, resource policy, recycling, and resource use.

Agriculture—Aids in crop management, protection of agricultural lands, conservation practices, and prime agricultural land policy and management.

Crime prevention, law enforcement, criminal justice—Facilitates selection of sites or premises for target-hardening attention, establishment of risk-rating procedures for particular locations, tactical patrol allocation, location selection for crime prevention analysis, crime pattern recognition, and selection of areas or schools for delinquency prevention attention.

Homeland security and civil defense—Assess alternative disaster relief plans, needs for stockpiling of foods and medical supplies, evacuation plans, and the proper designation of disaster relief areas.

Communications—Facilitates siting of transmission lines, location of cellular equipment, and education.

Transportation—Facilitates alternate transportation plans, locational analysis, mass transit, and energy conservation.

Next Steps: Seemingly Independent Things You Need to Know

Before we launch into the theory and application of GIS, let's look ahead at the remaining text in the Overview of this chapter. You may know some or all of this material already, depending on your background. To use GIS effectively, you should know something about several topics that may seem unrelated at first glance. The next few sections briefly review the relevant aspects of the following:

- ❏ Determining where something is: coordinate systems
- ❏ Determining where something is: latitude and longitude
- ❏ Geodesy, coordinate systems, geographic projections, and scale
- ❏ Projected coordinate systems
- ❏ Geographic vs. projected coordinates: which should you use?
- ❏ Two projected coordinate systems: UTM and state plane
- ❏ Physical dimensionality
- ❏ Global positioning systems
- ❏ Remote sensing
- ❏ Relational databases
- ❏ Another definition of GIS
- ❏ Computer software: in general
- ❏ Computer software: ArcGIS in particular

Determining Where Something Is: Coordinate Systems
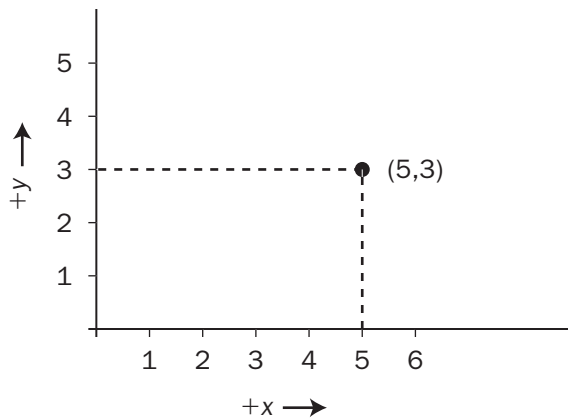
Cartesian Coordinate Systems

A coordinate system is a way of determining where points lie in space. We are interested in two-dimensional (2-D) space and three-dimensional (3-D) space. In general, it takes two numbers to assign a position to a 2-D space and three numbers in 3-D space.

Coordinates may be thought of as providing an index to the locations of points in space, and hence to the features that these points define.

To make a 2-D Cartesian[5] coordinate system, draw two axes (lines) that cross at right angles on a piece of paper. The point at which they cross is called the "origin." The sheet of paper is the x–y plane. Arrange the page on a horizontal table in front of you so that one line points left–right and the other toward and away from you. The part of the line from the origin to your right is called the positive x-axis. The line from the origin away from you is called the positive y-axis. Mark each axis in equal linear units (centimeters, say) starting at the origin, as shown in Figure 1-5. Now, a pair of numbers serves as a reference to any point on the plane. The position x = 5 and y = 3 [shorthand: (5,3)] is shown.

You can create a 3-D Cartesian coordinate system from the 2-D version: Imagine a vertical line passing through the origin; call it the z-axis; the positive direction is up. Now you can reference any point in three-dimensional space. The point x = 5, y = 3, and z = 4 [written more concisely as (5, 3, 4)] is shown in Figure 1-6.

This is called a right-hand coordinate system. The thumb, forefinger, and middle finger represent the positive axes x, y, and z, respectively. With your right hand outstretched, arrange those three digits so that they are roughly mutually orthogonal—that is, with 90º angles between each pair. Point your thumb to the right and your forefinger away from you. Now your middle finger will be pointing up.[6]

FIGURE 1-5   2-D Cartesian coordinate system

---

[5]Descartes, who lived from 1596 to 1650, made major contributions to both mathematics and philosophy. Descartes is credited with integrating algebra and geometry, by inventing the coordinate system that (almost) bears his name.
[6]It's best not to practice this exercise where other people can see you.
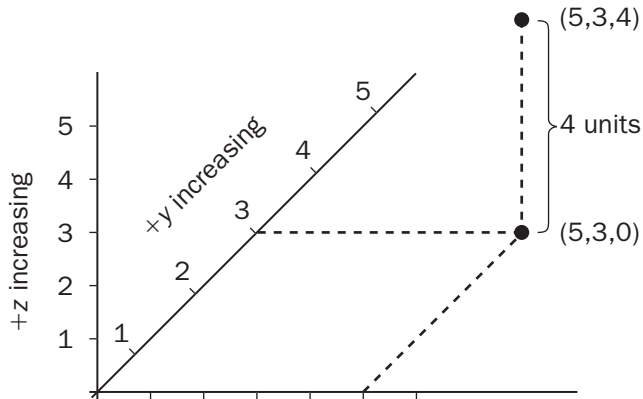
FIGURE 1-6   3-D Cartesian coordinate system

Spherical Coordinate Systems

A spherical coordinate system is another way to reference a point in 3-D space. It also requires three numbers. Two are angles, and the third is a distance. Consider a ray (a line) emanating from the origin. The angles determine the direction of the ray. See Figure 1-7.

The latitude–longitude graticule (a gridded reference network of lines encompassing the globe) is based on a spherical coordinate system. As often happens, different fields of endeavor use different descriptive approaches. Here, referencing navigation and Earth location issues requires a different system from
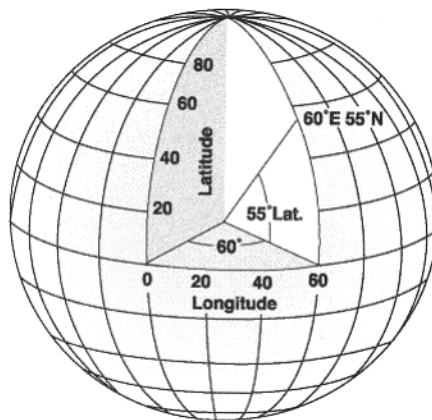
FIGURE 1-7   Spherical coordinate system showing latitude 55° North and longitude 60° East

the one mathematicians use in more abstract systems.[7] The origin is considered to be the center of the Earth. The equator serves as intersection of the x-y plane and the hypothetical sphere of the Earth. To determine the coordinates of a point, one angle (latitude) is measured from the x-y plane. The other angle (longitude) is contained in the x-y plane and is measured from the meridian that passes through Greenwich, England. The third number in a mathematical spherical coordinate system is the distance along the ray from the origin to the point. When added to the latitude-longitude system, altitude is usually defined instead to be the distance to the point along the ray from mean (average) sea level (MSL) or from a gravity-defined pseudo-ellipsoid used with the NAVSTAR Global Positioning System to be discussed shortly.

By using three numbers, you can determine and communicate the position of any point on Earth. Of course, an externally defined set of parameters must qualify these numbers. Any given point on the surface has probably been addressed by dozens of different sets of numbers, based on the parameters (e.g., units) of the coordinate system chosen. These parameters must match when you combine GIS data.

Determining Where Something Is: Latitude and Longitude

A fundamental principle underlies all geography and GIS: Most things on Earth don't move (or move very slowly) with respect to each other. Therefore, we can talk about the position of something embedded in or attached to the ground and know that its position won't change (much). It seems like a straightforward idea, but position confuses a lot of people when it is described as a set of numbers.[8]

Let's suppose that in 1955 somewhere in the United States you (or your parents, or their parents) drove a substantial metal stake or pin vertically into solid ground. Now consider that the object, unless disturbed by human beings or natural forces such as erosion or an earthquake, would not have moved with respect to the planet since then.[9] In other words, it is where it was, and it will stay there. Three numbers—latitude, longitude, and altitude—could identify the location of the object in1955. However, over the last half century, teams of mathematicians and scientists (skilled in geodesy) developed other sets of numbers to describe exactly the same spot where your object now resides. The actual position of the object didn't change, but additional descriptions of the where of the object have been created.

Ignoring the matter of altitude for the moment, suppose that the object was driven into the ground at latitude 38.0000000° (North) and longitude 84.5000000° (West), according to calculations done before 1955 that indicated the location of the center of the Earth, its shape, and the location if its poles. Most people and organizations in the United States in 1955 used the North American Datum of NAD27 (NAD27) to estimate the latitude-longitude graticules, based on parameters of the earth-approximating ellipsoid determined by Clarke in 1866.[10]

---

[7]For example, in the two dimensional Cartesian plane, a mathematician will measure angles starting from the positive x-axis (east) as zero and increasing counterclockwise, to 360 degrees (which is, again, zero). The navigator (think: compass) or cartographer will also use 360 degrees to represent a full circle, but measures clockwise from the positive y-axis (north).
[8]Descriptions of points aren't always just numbers. A possibly apocryphal "metes and bounds" description of a point in Kentucky a couple of centuries ago was "Two tomahawk throws from the double-oak in a northerly direction."
[9]Well, hasn't moved much. If it was on the island of Hawaii, it has moved northwest at about 4 inches per year. Also, if you are unfortunate enough to be in a place where there was an earthquake, it might well have moved and not returned to its original position.
[10]Based on a monument on Meades Ranch, in Kansas; the Clarke 1866 ellipsoid was meant to minimize the error between itself and the geoid in the United States.

The datum described as the World Geodetic System of 1984 (WGS84) offers the most recent, widely accepted view of where the center of the Earth is, its shape, and the location of its poles. The ellipsoid of WGS84 is virtually identical to the GRS80 ellipsoid.[11] In the coterminous states of the United States, this datum is virtually identical (within millimeters) to the North American Datum of 1983 (NAD83), although they result from different approaches and calculations.

According to the WGS84 latitude-longitude graticule, the object previously described would be at latitude 38.00007792º and longitude 84.49993831º. The difference might seem insignificant, but it amounts to about 10 meters on the Earth's surface. Or consider it this way: According to NAD83, a second object placed in the ground at 38º N and 84.5º W would be 10 meters away from the first one. Does that sound like a lot? People have exchanged gunfire in land disputes over smaller distances. Given a latitude and longitude, a GIS must know the datum that is the basis for the numbers. Hundreds of datums exist, and many countries have their own.

Geodesy, Coordinate Systems, Geographic Projections, and Scale

First, a disclaimer: This text does not pretend to cover in detail such issues as geodetic datums, projections, coordinate systems, and other terms from the fields of geodesy and surveying. Nor will the text rigorously define most of these terms. Simply knowing the definitions would mean little without a lot of study. Many textbooks and Web pages are available for your perusal. These fields, concepts, and principles may or may not be important in your use of GIS, depending on your projects. However, the datum, projection, coordinate system designations, and measurement units must be identical when you combine GIS or map information. If not, your GIS project may well produce inaccurate results.

How we apply the mathematically perfect latitude-longitude graticule to points on the ground depends partly on human's understanding of the shape of the Earth. This understanding changes the more we learn. Geodesy is the study of the shape of the Earth and the validity of the measurements human beings make on it. It deals with such issues as spheroid and datum. You don't have to know much about geodesy to use a GIS effectively, provided your data are all based on the same spheroid and datum (and projection and units, as you will see later). It is the application of geodesic knowledge that caused the differences in the coordinates of that hypothetical object I discussed earlier that was put into the ground six or seven decades ago. The object hasn't moved. We simply have a better idea of the location of the object relative to the latitude-longitude graticule.

Projected Coordinate Systems

For several reasons, it's often not convenient to use latitude and longitude to describe a set of points (perhaps connected by straight lines to make up a coastline or country's boundaries) on the Earth's surface. One is that doing calculations using latitude and longitude—for example, determining the distance between two points—can involve complex operations such as products involving sines and cosines. For a similar distance calculation, if the points are represented on the Cartesian x-y plane, the worst arithmetic hurdle is a square root.

---

[11]GRS80 is a global geocentric system based on the ellipsoid adopted by the International Union of Geodesy and Geophysics (IUGG) in 1979. GRS80 is the acronym for the Geodetic Reference System of 1980.

Latitude and longitude measures for many geographic applications do not work well for several aspects of mapmaking. Suppose you plot many points on the Earth's surface—say, along the coastline of a small island that is a considerable distance from the equator—on a piece of ordinary graph paper, using the longitude numbers for x-coordinates and latitude numbers for y-coordinates. The shape of the island would look strange on the map (it would appear horizontally stretched) compared to how it would appear from an airplane. You would not get useful numbers if you measured distances or angles or areas on the plot. This is due to a characteristic of the spherical coordinate system: The length of an arc of a degree of longitude does not equal the length of an arc of a degree of latitude. Those lengths are almost equal near the equator, but the difference grows as you go further north or south from the equator. At the equator, a degree of longitude translates to about 69.17 miles. Very near the North Pole a degree of longitude might be 69.7 inches. (A degree of latitude, in contrast, varies only between about 68.71 miles near the equator and 69.40 miles near the poles.)

For relatively small areas, mathematically projecting the spherically defined locations onto a plane provides a good solution to problems associated with calculations and plotting. Geographic projection might be thought of as imagining a process that places a light source inside a transparent globe that has features of the Earth inscribed on it. The light then falls on a piece of paper (or one that is curved in only one direction and may be unrolled to become flat).[12] The shadows of the features (say, lines, or areas) will appear on the paper. Applying a Cartesian coordinate system to the paper offers the advantages of easy calculation and more realistic plotting. However, distortions are inherent in any projection process; most of the points on the map will not correspond exactly to their counterparts on the ground. The degree of distortion is greater on maps that display more area. Accuracy suffers when you flatten a curved surface and thus convert a three-dimensional coordinate system to a two-dimensional one.

After constructing geographic data sets according to latitude and longitude, based on some agreed-upon spheroid (such as GRS80) and datum (such as WGS84), you decide how to represent them graphically for viewing. At one time, cartographers went directly from the latitude-longitude description of an area or feature to a graphical portrayal on paper. This usually involved "projecting" the data from a three-dimensional (3-D) spherical coordinate system to a two-dimensional (2-D) Cartesian one, and setting a scale: A certain number of units on a map represented that number of units on the ground. Using GIS changes this. As I commented earlier, latitude-longitude is the most fundamental and accurate way to represent spatial data. So large areas—those that are more subject to distortion by being projected—may best be left in latitude-longitude coordinates, and subareas projected to other coordinate systems when needed. Computers are really good at doing the complex computations necessary to project data.

The larger the area projected, the greater the tendency for things not to be where the projected coordinates say they are.

Four considerations for viewing and analysis are size, shape, distance, and direction. Myriad projections have been invented. Many distort size, shape, distance, or direction, and some preserve one or two of them.

Regarding linear units, GIS differs from cartography in that matters of scale may be left until the very end. The position defining numbers in the database should be real-world coordinates—not scaled coordinates. That is, GIS "maps" are stored in ground units rather than map units. Besides being a more

---

[12]This description is a sort of a cartoon to describe a map projection. A map projection is actually a mathematical transformation that "maps" points on the globe to points on a plane; the process may be quite complex depending on the projection; a single light source at the center of the globe does not suffice to explain it.

fundamental way to store data, this makes it possible to easily make maps of any desired scale on the computer monitor or on paper. Scale is only a consideration when you measure on a physical map. Computers take the worry out of determining scale accuracy. Because of the vast computational power of a computer, there is no difficulty in scale conversion. The days of: "Let's see, this distance is 5.3 inches on the map, and 1 inch is 12 miles, so the distance is about 64 miles" are over. Scale is a minor concept in GIS—one used only on final output.

Geographic vs. Projected Coordinates: A Comparison

❏ *Advantages of the spherical coordinate system*—You can represent any point on the Earth's surface as accurately as your measurement techniques allow. The system itself does not introduce errors.

❏ *Disadvantages of a spherical coordinate system*—You will encounter complex and time-consuming arithmetic calculations in determining the distance between two points or the area surrounded by a polygon determined by a set of points. Latitude-longitude numbers plotted directly on paper in a Cartesian coordinate system result in distorted—sometimes greatly distorted—figures.

❏ *Advantages of a projected coordinate system on the Cartesian plane*—Calculations of distances between points are trivial. Calculations of areas are relatively easy. Graphic representations are realistic, provided the area covered is not too large.

❏ *Disadvantages of a projected coordinate system on the Cartesian plane*—Almost every point is in the wrong place, although maybe not by much. All projections introduce errors. Depending on the projection, these errors are in distances, sizes, shapes, or directions.

Whether you use geographic or projected coordinates, ensuring that the parameters of geographic data match is of paramount importance in combining GIS data sets if you want the right answers![13]

Two Projected Coordinate Systems: UTM and State Plane

A coordinate system called Universal Transverse Mercator[14] was developed based on a series of 60 projections onto semi-cylinders that contact the Earth along meridians. (To consider, for example, one of these projections, imagine a sheet of paper curved so that it becomes a half cylinder whose radius is that of the Earth's. Then, with the axis of the cylinder oriented in an east–west direction—hence the term transverse—the paper is brought into contact with a globe along the meridian designating 3º longitude. Then, the surface of the Earth between 0º and 6º is projected onto the paper). This process is repeated for central meridians of 9º, 15º, 21º, and so forth up to 357º. The term "zone" is ambiguously used for this swath of territory. However, UTM projections are further subdivided into areas, also called zones, covering 6º of longitude and, for most zones, 8º of latitude. Further, ArcGIS divides a total zone into a

---

[13]The parameters of data sets may not always match but may be close enough so that any error introduced is trivial. In many instances and places, for example, NAD83 and WGS84 match within centimeters. However, you must research carefully to know when you can use data sets that don't match exactly. The Esri manual *Understanding Map Projections* can help.

[14]The idea of the Mercator projection was developed in 1568 by Gerhardus Mercator, a Flemish geographer, mathematician, and cartographer.

northern and a southern part. In any event, a coordinate system is imposed on the resulting projection such that the numbers in any given zone:

❏ Are always positive

❏ Always increase from left to right (west to east)

❏ Always increase from bottom to top (south to north)

The representation of our previously discussed object (at 38º N and 84.5º W) in the UTM coordinate system, when that system is based on WGS84, is a "northing" of 4,208764.4636 meters and an "easting" of 719,510.3358 meters. The northing is the distance to the point, in meters,[15] from the equator measured along the surface of an "Earth" that has no bumps. The easting is somewhat more complicated to explain because it depends on the zone and a coordinate system that excludes negative numbers. Consult a textbook on geodesy or cartography, or review the thousands of Web pages that come up when you type the words

```
UTM    "coordinate system"  "Transverse Mercator"
```

into an Internet search engine (e.g., www.google.com).

One version of the UTM coordinate system is based on NAD27. In this case, our object would have different coordinates: northing 4,208,550.0688 and easting 719,510.6393. This produces a difference of about 214 meters. If you combined WGS84 UTM data with NAD27 UTM data, the locations they depict might be in error by just enough not to be obvious at some scales, but great enough to cause trouble.[16]

Each state of the United States has one or more State Plane Coordinate Systems (SPCSs). They were developed, originally in the 1930s, by the U.S. Coast and Geodetic Survey. These systems are based on different projections (usually Transverse Mercator or Lambert Conformal Conic), depending on whether the state is mostly north-south (like California) or mostly east-west (Tennessee). The units of an SPCS may be international feet, survey feet, or meters, depending on decisions made by the state itself.[17] Zone boundaries frequently follow county boundaries. The coordinate system(s) used in one state are not applicable in neighboring states. Nor can you apply the SPCS of one zone to areas in the state only a short distance away in another zone. Furthermore, the difference between NAD27 and NAD83 (WGS84) can be startlingly large. In Kentucky, for example, 38.0000000º (North) and 84.5000000º (West) would translate into a northing of 1,568,376.1900 feet and an easting of 182,178.3166 feet when based on WGS84. However, when the basis in NAD27, the coordinates are 1,927.939.8692 and 182,145.9821, which makes a difference of some 68 miles!

Why the large differences in projected coordinate systems based on NAD27 and those based on WGS84? Because those responsible for the accuracy of the other coordinate systems took advantage of the

---

[15]In fact, the meter was originally defined as one tenth-millionth of the distance from the equator to the North Pole, along a meridian that passed through Paris.

[16]Comparing these coordinates with the WGS84 UTM coordinates, you see that virtually all of the difference is in the north-south direction. While true for this particular position, it is not true in general.

[17]The meter is the standard unit of length in most places in the world. Two different lengths of "foot" are defined in terms of the meter. A "U.S. Survey Foot" is one in which a meter is considered to be 39.37 inches, exactly; the other sort of foot is the international foot, where an inch is 0.0254 meters, exactly. The survey foot and the international foot are almost, but not exactly, the same length.

development of WGS84—a worldwide, Earth-centered, latitude-longitude system—to correct or improve or change the origin of those projected coordinate systems.

State plane coordinate systems are generally designed to have a scale error maximum of about 1 unit in 10,000. Suppose you calculated the Cartesian distance (using the Pythagorean theorem) between two points represented in a state plane coordinate system to be exactly 10,000 meters. Then, with a perfect tape measure, pulled tightly across an idealized planet, you would be assured that the measured result would differ by no more than 1 meter from the calculated one. The possible error with the UTM coordinate system may be larger: 1 in 2500.
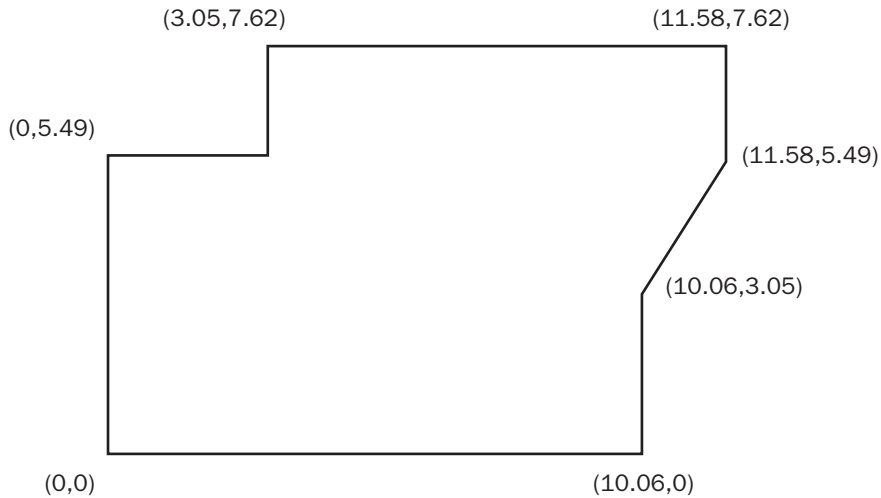
Coordinate Transformations

Coordinate transformation of a geographic data set is simply taking each coordinate pair of numbers in that data set and changing it to another pair of numbers that indicates exactly the same spot on the Earth's surface, but using a different system of assigning coordinates.

Let's make up a coordinate system for a garden delineated by lines between stakes. Suppose the origin (0,0) is at the southwest corner. Now drive several stakes in the ground so that if you passed a string around them it would outline a polygon. Suppose you use a surveyor's tape calibrated in survey feet to measure the Cartesian coordinate for each stake. Stake 0 is at the origin (0.0). Stake 1 is 33 feet east and 0 feet north (i.e., 33,0). Stake 2 is at (33,10), and so on. See Figure 1-8.

Suppose the garden grows and then becomes overgrown. Next year someone locates the origin and wants to find the original stakes. He also has a surveyors tape, but it is calibrated in meters instead of feet. He asks you to provide the data on where the stakes are in meters. In this case, the transformation of coordinate systems is easy: You simply multiply each number by a conversion factor (the number of meters per survey foot, which is 0.3048). So Stake 1's coordinates are 10.06, 0). Those of Stake 2 are (10.06, 3.05), and so on. See Figure 1-9.
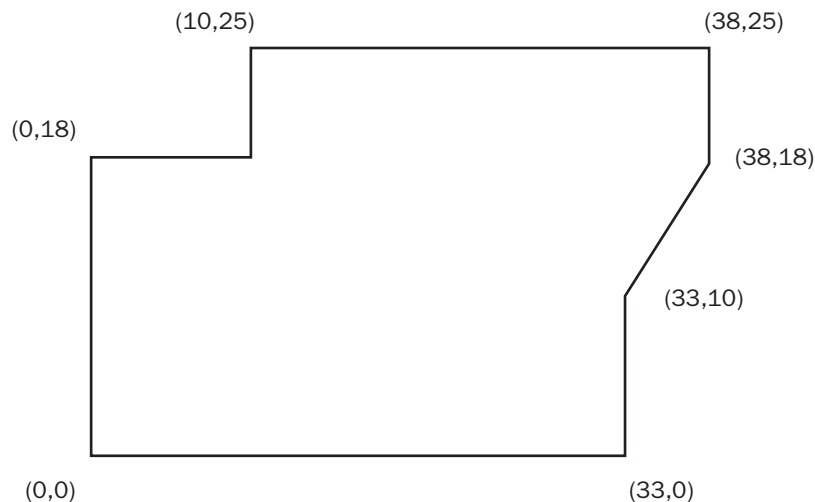
FIGURE 1-8   Locations of garden stakes

FIGURE 1-9   Locations of garden stakes

The stakes are in the same place that they were before. Their positions are simply referred to by a different set of numbers. Geographic coordinate or datum transformation is nothing more or less than this, except that the mathematical operations applied to each number are usually more involved.

Physical Dimensionality

All matter exists in four dimensions—roughly characterized as left-right, toward-away, up-down, and time. The first three are conceptually clumped together and called "spatial dimensions." Time is treated as a dimension here because we want to talk about "position" in space and time.[18]

Any physical object occupies space and persists in time. This is true of the largest object in the universe and the smallest atom. However, when the measure of one or more dimensions of an object is small, or insignificant, with respect to the measure of others or is tiny with respect to its environment, it is useful to describe and represent objects by pretending that they occupy fewer than three spatial dimensions. For example, a sheet of paper can be considered a pseudo-two-dimensional object; it has thickness (up-down), but that dimension is miniscule compared with left-right and toward-away. A parking meter in a city could be considered a pseudo zero-dimensional object, because the measures of all its dimensions are insignificant with respect to its environment. (The parking meter would certainly not be considered a spatially zero-dimensional object by its designer, manufacturer, or the driver who uses it. Therefore, the pseudo-dimensionality of an object depends on its context.)

---

[18]In physics, time is treated as the fourth dimension because it is inexorably bound up with the other three according to Einstein's theories (special in 1905 and general in 1916) of relativity. This connection need not concern us except for some esoteric technical matters related to time on global positioning system satellites (for one, time on a satellite runs faster because it is farther from Earth, where there is less gravity).

It is important to consider the pseudo-dimensionality of both an object and the field (the space) in which it resides. A point on a line is a zero-dimensional object in one-dimensional space. A line segment on a plane is a one-dimensional object in two-dimensional space. A polygon is a two-dimensional object in a two-dimensional space. A point in a volume is a zero-dimensional object in three-dimensional space. The dimensionality of the object must be less than, or equal to, the dimensionality of the space in which it resides.

This issue of dimensionality comes into play here because great economies of computer storage are achievable if an object is considered to have fewer dimensions than it actually does. For example, describing a fire hydrant as a complex three-dimensional spatial object would be an arduous task and involve many numbers and much text. If it is considered a zero-dimensional object, its location can be described (very precisely, but still inexactly) with three numbers, perhaps representing the latitude, longitude, and altitude of some point on the hydrant or, perhaps, its center of mass.

The spatial dimensions of a spatial pseudo-zero-dimensional object are insignificant with respect to the context or environment in which it resides. A theoretical geometric point is the prototypical example. Examples of such zero-dimensional objects on maps or in a GIS could be streetlights, parking meters, oil wells, census tracts, or cities—depending on the spatial extent of other objects or features in the database.

A spatial pseudo-one-dimensional object or feature has no more than one significant dimension or is made up of essentially one-dimensional objects. A straight-line segment is the prototypical example. In terms of making a one-dimensional object up from component parts, consider the example of a telephone wire strung from one pole to the next to the next and so on. It is considered a one-dimensional object, even though the poles may not be in a straight line, so that the phone line zigzags over a two-dimensional field. (The line itself sags and persists in time—and therefore exists in a four dimensional context—but these facts notwithstanding, it can be simplified or generalized into a one-dimensional feature—saving us lots of computer storage, processing time, and conceptual complication.) Roads, school district boundaries, pipes for fluid transportation, and contour lines on a topographic map are all examples of pseudo one-dimensional entities.

You can disregard the up-down dimension in a spatial-pseudo-two dimensional object. Plain areas such as voting districts and soybean fields are examples. Such pseudo-areas lack vertical components—variability in altitude—that can be important. The true surface area of hilly terrain may be considerably underrepresented by the plane area within the borders of the plane figure that represents it.

You must consider all dimensions of a spatial three-dimensional object in order to represent it. Three-dimensional features are volumes—say, a coal seam or a building.

Relative to each other, positions of features that we record with a GIS do not move, or do not move quickly, with respect to each other. Sometimes we want to know how conditions change or object move. The representation of the objects may be zero-, one-, two-, or three-dimensional. At its most complex, GIS would involve all four dimensions. Most GIS operations and data sets are two-dimensional; the data are assumed to pertain to moments in time, or stable phenomena or conditions over a period of time. Three-dimensional GIS could involve either three spatial dimensions (such as representing the volume of a limestone quarry) or two spatial dimensions and varying time (showing historical changes in a landscape). Rarely will you use four-dimensional GIS.

Results from a GIS are always only an approximation of reality. One of the reasons, among others to be discussed later, is that we simplify objects by reducing their dimensionality.

Global Positioning Systems

A global positioning system (GPS)[19] is a satellite-based system that provides users with accurate and precise location and time information. Using NAVSTAR GPS, you can determine locations on Earth easily within a few meters, and, with more difficulty and expense, within a few centimeters or better. Timing within 40 billionths of a second (40 nanoseconds) is easily obtained. Timing within 10 nanoseconds is possible.

The U.S. Department of Defense operates NAVSTAR GPS in cooperation with the U.S. Department of Transportation. The acronym NAVSTAR stands for NAVigation System Timing And Ranging Global Positioning System." Informally, it is the Navigation Star.

The Russian GLONASS (Global Navigation Satellite System) operates similarly. Concerns about U.S. control over NAVSTAR led Europe to begin development of its own independent Galileo system in 2002, but it is just now becoming operational. China is developing the BeiDou (Compass) GNSS.

A GPS receiver, which "remembers where it has been," is becoming a primary method of providing data for GIS. For example, if you drive a van with a GPS antenna on its roof along a highway, recording data every, say, 50 feet, you will develop an accurate and precise map of the location of the highway. The NAVSTAR GPS is discussed in Chapter 5.[20]

Remote Sensing

Remote sensing probably stated with photographs taken from balloons in the 1840s. The first automated system (not requiring human beings to be with the sensors) may have been in the 1890s in Europe when cameras programmed to take pictures at timed intervals were strapped to pigeons!

Evelyn Pruitt probably introduced the modern use of the term "remote sensing" in the mid-1950s when she worked as a geographer/oceanographer for the U.S. Office of Naval Research (ONR). Remote sensing uses instruments or sensors to capture the spectral characteristics and spatial relations of objects and materials observable at a distance, typically from above them. Using that definition, everything we observe is remotely sensed. More practically, something is sensed remotely when it is not possible or convenient to get closer.

We can categorize remote sensing for GIS many ways. Data can be taken from aircraft or satellite "platforms." The energy that the sensor "sees" can come from the objects or areas being examined as a result of radiation emanating from them (caused by the sun or other heat or light sources) or from radiation bounced off them by an energy source associated with the sensor (e.g., radar or lidar). The images produced may be developed on film or produced by digital sensors. Satellites in geosynchronous orbits can hang in a single spot over the equator. Those in near polar orbits can see different areas of the Earth as it turns. Chapter 2 offers examples of remotely sensed data.

---

[19]Also referred to as a positioning, navigation, and timing (PNT) system, or a global navigation satellite system (GNSS).
[20]See also Prof. Kennedy's textbook *The Global Positioning System and ArcGIS*, Third Edition, 2010, CRC Press (Taylor & Francis Group), 301 pages.

Relational Databases

GIS relies heavily on databases of text and numbers. The relationship between such information and geography is discussed later in the chapter. For now, you simply need to know how text and number data sets are stored.

For this discussion, a database is a collection of discrete symbols (numbers, letters, and special characters) located on some physical medium with at least one principal underlying organization or structure. An old-fashioned library card catalog is an example of a database with a single underlying structure: an alphabetical list of authors; the medium is 3 x 5 index cards, and the data are the symbols on the cards describing books and their locations in the library. Most libraries have substituted computer-based catalogs with the advantage that a user can search and find not only authors but titles and subjects as well, so a number of organizing themes may underpin a database.

Another example is a "hard disk" that has recorded the most common type of soil found in a specific acreage in a county. The disk is the physical medium, the codes assigned to a soil type constitute that data, and the location of each acre—as understood from the position of each datum on the disk—could be the underlying structure.

Existing general purpose databases usually:

❏ Result from some sort of project; some individual or team constructs it—frequently going to considerable effort.

❏ Need to be *updated* (modified and corrected as time progresses) if they are to continue to be of value.

❏ Contain errors *regardless* of size, care of construction, simplicity of data, or quality or physical medium used.

❏ Serve a function when allied with some process. The function may be as simple as supplying a telephone number from a physical phone directory (structure—alphabetical by name; medium—cheap bound paper; data—phone numbers, in very small type; process—looking up a name, finding the adjacent number). The function served by the database might also be quite sophisticated—supporting far-reaching land use decisions, for example.

Numerous databases are used to solve problems at all levels of government. Access to these databases is achieved through the use of referencing schemes. Following are some examples:

Referencing Scheme	*Examples of Data Contained*
Names of people	Salary, Social Security number, medical history, criminal record
Auto license plate number	Car color, owner, serial number
Street address	House value, lot size
Job title	Person employed, duties, salary
Transaction number	Money received, paid, transferred, invested
Events	Schedules, orders, crimes, accidents

Many schemes exist for presenting or storing data. Suppose parents want to find a name for a newborn child. A list of potential names in random order might be provided. It would probably be more useful if the list were divided into girl's names and boy's names. Another improvement could be to alphabetize the list, or present it in terms of current popularity of names.

As a second example, consider how you might store telephone numbers associated with names and addresses. You could order the list alphabetically, so you could easily find a phone number, knowing the name. Or you could order the list numerically by telephone number, so you could find a name, given a number.

A different sort of data structure is "hierarchical." You could use this approach to store the names and positions of people in a corporation or a military organization based on who reports to whom. Likewise you could store voting districts within counties within states. In the computer world, this "tree-structured directory" approach is used to organize the folders and files contained on a hard disk drive on a computer. Any folder on the disk may store other folders and files.

Many schemes exist to store information in the memory of a computer or on its secondary devices, such as disk drives or tapes. The primary method used to store large amounts of information is called the "relational database," or RDB, developed by E. F. Codd.[21] The software is described as a RDBMS (relational database management system). The idea is simple: use a set of two-dimensional tables; for a given table there is a prescribed format in which the rows relate to entities (objects, people, things in general), while the columns relate to attributes (characteristics, properties) of entities. The intersection of a given row and a given column is a cell, containing a value, which defines the particular attribute of the particular entity. See Figure 1-10.

You can use a database table to store names, occupations, and pay schedules of employees, as in Figure 1-11.

Another example shows part of a database of automobiles registered in a state. Each row would represent one car; each column would represent one property of cars. See Figure 1-12.

Here is some terminology: The structure that contains the entity, the row, is also called a *tuple* or *record*. The structure that contains the attributes, the column, is also called a *field* or an *item*.

Relational Database Nomenclature
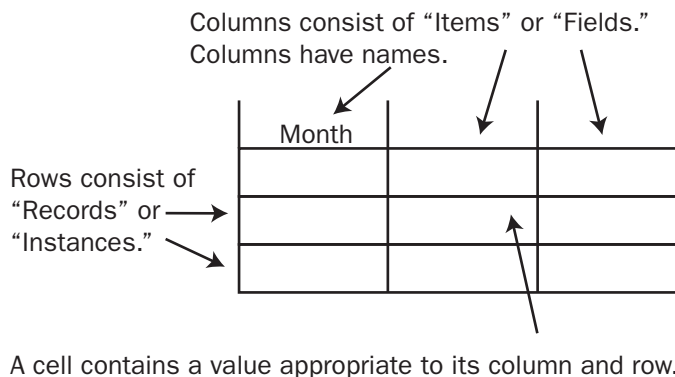
FIGURE 1-10   Components of a relational database table

---

[21]Edgar F. Codd was the originator of the relational approach to database management, which he introduced in 1970. For this work, he was the recipient of the Association for Computer Machinery (ACM) Turing award in 1981.

Garden-Variety Database

Name	Occup	Pay_Sched
Smith	Welder	Hourly
Jones	Accountant	Bi-Monthly
Adams	Attorney	Monthly

FIGURE 1-11   Trivial example of a relational database table

Mfgr	Num_Doors	VIN	Color	Weight
Porsche	2	123XXX	Silver	2300
Porsche	2	887ABC	White	2100
Toyota	4	9880123	Grey	2350
Honda	4	456789	Blue	2999

FIGURE 1-12   An RDB of automobiles

All the cells of a given column must contain the same sort of value. Some common ones are as follows:

❏ *Character*—Any valid ASCII character in a string of almost any length.

❏ *Short integer*—Can range from slightly less than −32,000 to slightly more than +32,000.

❏ *Long integer*—Can range from slightly less than negative two billion to slightly more than positive two billion.

❏ *Floating-point number*—Can have an exponent as small as 10^{-38} and as large as 10^{38}, and you can be assured of six significant digits of precision.

❏ *Double-precision floating-point number*—Can have an exponent as small as 10^{-308} and as large as 10^{308}, and you can count on 16 significant digits.

❏ *Boolean*[22]—A value that is either true or false.

---

[22]Named after George Boole, mathematician (which is why it is not spelled Boolian).

Getting Information from a Relational Database: Queries

Relational databases are designed to give you information. You can obtain the information by selecting a subset of records from the total set by writing an expression that is a mixture of attribute names, arithmetic and logical operators, and values.[23] For a trivial example, suppose that research has found that gray cars that weigh less than 2500 pounds put their occupants in greater risk than average. You want to select those records from the statewide automobile database. You might first get all the records of the cars that are gray.

```
SELECT: COLOR = 'Gray'
```

Given that subset, you might then write

```
SELECT: WEIGHT < 2500
```

Given this sub-subset of records, you could perhaps write letters to the owners of those cars, making them aware of the danger they face.

Languages to get subsets of records can provide flexibility and efficiency. For example, to do both of the preceding operations with one expression, you might write

```
SELECT: COLOR = 'Gray' AND WEIGHT < 2500
```

Suppose also that the research study showed increased danger to those occupying cars that were built in 1985 or before, regardless of color or weight. You might add to the preceding selection by saying

```
ADDSELECT: YEAR <= 1985
```

(ADDSELECT means add to the current set of selected records.)

Or you may use a single query to select all the records you want at once:

```
SELECT: (COLOR = 'Gray' AND WEIGHT < 2500) OR YEAR <=1985
```

Note the use of parentheses to indicate the order in which operations are done.

Economies in Relational Databases

In theory, the information in a relational database system could reside in a single table. This isn't the best policy and in reality may not even be possible. A relational database usually consists of a set of tables that relate to one another—thus, the word "relational."

Each relational database table must contain a column that has a unique identifier for each record in the table. This is known as the *key field*. If the relational database references people, the key might be Social

---

[23]There are many relational database software products and several different languages used to query them. What is shown in the following is not a specific query language but rather a generic representation of such a language. ArcGIS may be used with several relational database products, each with a somewhat different query language.

Security number, employee number, or student number. In the case of automobiles, the key code might be the vehicle identification number (VIN).

To illustrate why multiple tables might be used, suppose that, in our relational database, it is possible for one person to own more than one automobile. If the entire database is all in one table, then the names and addresses of each multicar owner must be repeated, which increases the amount of storage required. If a multicar owner gets a new street address, several records in the database table would have to be changed. Further, some of the data about owners might be located in another database and might be private. The elegant answer to these problems is to have several tables that contain the information, divided to minimize the repetitions. In the automobile example, it might be that only the owner's identification number is stored in the record of the car itself. This number could be the key field in the database table containing information about owners. If some of the owner information is private, it could be stored in a separate table as well, with the same key field. The database could be set up so that this table could not be accessed by the automobile table.

Relational database tables that meet certain requirements of efficiency are said to be in first normal form (1NF), second normal form (2NF), and so on. The higher the number, the more efficient the database.

In summary, it is useful to partition the information in a relational database into a number of tables. Such partitioning can

❑ Reduce the amount of redundant data stored

❑ Aid in updating the database

❑ Reduce the chances of inconsistency and instability in the database

❑ Aid in protecting private or sensitive data

Languages have been developed for retrieving information from a relational database. One is the Structured Query Language (SQL), developed by the American National Standards Institute (ANSI). You use it to execute queries of a database. Vendors of relational database management systems also have developed proprietary languages.[24] Esri products interface with a number of RDBMS from various sources.

Databases—What's Meant by "Relational"

What you have seen so far is one aspect of the RDB: a two-dimensional table in which you store entities as rows and attributes as columns. What is it that makes the relational database such a powerful approach to storing information? One answer lies in the fact that a RDB can be much more than a single table. Usually it is a number of tables that are related to one another, as previously discussed, that provide for efficiency, flexibility, and ease of updating.

Table 1-1 illustrates a database that might be created by a motor vehicles licensing department. Parts of the database connect to other databases formed by other government departments. Shown in Table 1-1 are nine relational database tables; the RDB table name is on the top line; the attribute names are on the second line. None of the tens of thousands of cells is shown.

---

[24]Some names of past and present relational database management system software products are Access, dBASE, Informix, Ingres, Oracle, SQL Server, and Sybase.

Here, mainly for purposes of illustration, the MASTER table is miniscule, consisting only of a license plate number and an owner identification. The key field, shown in bold font, consists of a unique character string (no two license plates are the same) and an owner identification number. The contents of this second column may not be unique, since one person my own several vehicles.

TABLE 1-1

MASTER		
	Plate_#	Owner_ID

OWNER			
	Owner_ID	Name	Address	SSN

VEHICLE					
	Plate_#	Color	Model	Year	Type_code	VIN

VEH_TYPE					
	Type_Code	Weight	Length	Width	H_Power	Fuel_Cnsup

VOTERS			
	SSN	Name	Precinct_Code	Regis_Party

PRECINCT			
	Prec_Name	Location	Supervisor	**Precinct_Code**

ARRESTED					
	Name	Alias	SSN	Offense	Convicted	**Court_Doc_#**

ACCOUNTING		
	Plate_#	Date_Paid	Insurance_Co_Code

INSURERS			
	Insurance_Co_Code	Co_Name	Location	Phone

Both fields in MASTER refer to—link to—the other eight tables, either directly or indirectly. Plate_# allows the user access to both VEHICLE and ACCOUNTING.

The VEHICLE table describes some of the attributes of the car or truck in question. Since some aspects of all vehicles of a certain year, model, and manufacturer are identical, it would be a waste of space and an updating nightmare to place this information in a record describing a particular vehicle. So each record carries a Type_Code that refers to a key field in the table VEH_TYPE. There you will find the vehicle weight, length, width, horsepower, and fuel consumption data.

The ACCOUNTING table indicates whether taxes and registration have been paid. Also, here is a reference to the INSURERS table, which carries information about the companies that insure vehicles in the state.

The other column in the MASTER table contains an Owner_ID code that matches up with the key in the table OWNER. That table could contain a host of information about the owner of the vehicle. Shown in the table are names and addresses. Also there is the owner's Social Security number (SSN). Perhaps there is a "motor-voter" effort to register all vehicle owners. With the SSN, which is the key field in the VOTERS table, a user could determine which owners were already registered to vote. Further, with the Precinct_Code, information about the location of the voting precinct could be obtained.

Finally, the SSN also allows a link to the table ARRESTED so that drivers with moving violations or driving while intoxicated could be identified. Both the VOTERS and ARRESTED tables probably would reside in some other department, so maintenance of those tables would not fall to the Department of Motor Vehicles, yet the DMV would have access to the information.

Most governments and organizations have extensive and perhaps sophisticated techniques or systems for storing and manipulating data that can be referenced by these and other schemes. Here, we are interested in the quite-useful referencing basis that has not been developed as extensively, however. It may be known as several names: geographic, land locational, geodetic, or spatial position.

Relational Databases and Spreadsheets

A relational database table may look a lot like an electronic spreadsheet, such as Microsoft Excel. There are some important differences:

❏ A relational database is structured, with rows and columns strictly defined. In a spreadsheet, you can put anything anywhere.

❏ In a relational database, the headings of columns are not stored in cells of the relational database. The column headings (attribute names) are known to the database software and are displayed, but are not part of the data. In a spreadsheet, column headings occupy cells.

❏ A relational database is a logical object about which conjectures can be made and theorems proved. An entire branch of computer science is devoted to work with relational databases.

Searching (and Indexing) in General

In the exercises for this chapter, you will be addressing the subject of searching for data. A considerable part of the discussion will relate to indexing names of data files. You may wonder why. The discussion below will explain the rudiments of indexing (which is basically sorting) and why it is important.

A computer can easily search a list of items for a particular item. It simply checks each item in the list against the "search string," and when it finds the two are equal, it declares success. For a trivial example, suppose we had the list of strings, each consisting of three characters and each associated with a number:

```
GHU   2343
UCK   7765
OPR   8828
PRO   1234
ZYX   7876
QWA   9500
ASD   3456
```

If we wanted to find the number associated with QWA, we could program the computer to compare the query (text string) QWA with the first string in the list: GHU. No match, so we go on to the next string (UCK). Again, no match. Finally, we would come to QWA in the list and get our answer: 9500. As you can probably guess, on average we would have to look through half the list to find a given string. In this case, it took six comparisons before the string was matched.

However, if we sort the list alphabetically (carrying along the associated number) we would have:

```
ASD 3456
GHU 2343
OPR 8828
PRO 1234
QWA 9500
UCK 7765
ZYX 7876
```

Now we can apply a clever strategy. We look first at the middle of the list: PRO. It doesn't match QUA, but we know that QUA is alphabetically "greater than" PRO, so we only need to be concerned with the bottom half of the list. We pick the middle of that for the next comparison: UCK. Now we know we've gone too far, so we go halfway back and find QWA. This process, called a binary search, took only three comparisons.

This example, which we've kept very small, only hints at the power we have. If our list to be searched consisted of two million strings, then the first searching method (sometimes called brute force) would, on average have to do a million comparisons. In the worst case, it would have to do two million comparisons. How many would a binary search have to do? In the worst case, about 22.

You may say that you realize the tremendous advantage of a binary search (in which you essentially "throw out half the remaining list each time and, thus, reap the rewards of reducing the list in an exponential fashion) but you point out that the list had to be alphabetized to begin with. Four considerations come in to play here. First, computers are very good at sorting—lots of very fast algorithms exist for the process. Second, the sorting can be done "in the background"—during times when you aren't sitting in front of the machine waiting for a search to succeed. Third, once the alphabetization has been done, innumerable sorts can be run—each highly efficient. Last, if a new element is to be added to the list, it can be inserted in its proper place in the list, without resorting the entire list.

Another Definition of GIS

Each record in a relational database references something: a person, object, idea, equation, feature, subject—some unique entity. In a GIS, the entity is usually a spatial three-dimensional feature such as a parking meter, lake, railroad, and so on.[25] These 3-D features are almost always reduced to abstract objects of fewer dimensions. The parking meter is a point, the railroad is a set of lines, and the lake is an area bounded by a sequence of lines.

Usually, a record in a relational database references a person or object without respect to its current location. A subject of a record in a relational database frequently moves around—like cars or people. However, when the subjects of a relational database are fixed in space, the position, or positions, of the feature may be included in the description of the feature along with the attributes. In one sense, the location of the object becomes one of its attributes.

---

[25]In raster-based GIS, the entity is a set of (usually) square areas that share common values. This approach will be discussed extensively in Chapter 8.

Points

Here, for example, is a map and part of a relational database that together describe fire hydrants in a town. (See Figure 1-13.) Each hydrant has a unique number and is, thus, the key field. Fire hydrants (each of which has a latitude and longitude specification) correspond one-to-one with the records in the database.

Let's consider another definition of GIS: From a computer software point of view, a GIS is the marriage of a (geo)graphical database and an attribute database (frequently a relational database).
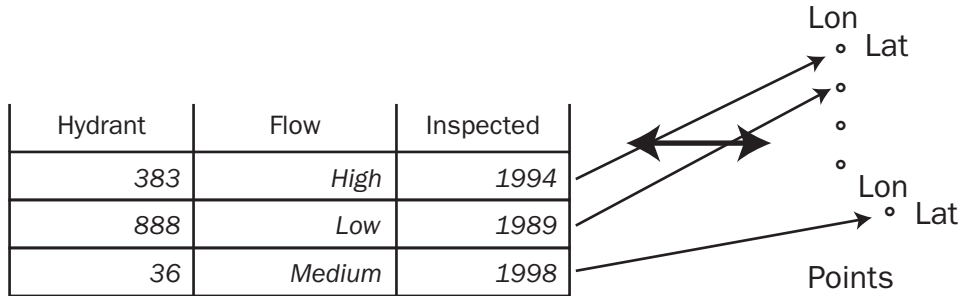
Hydrant	Flow	Inspected
383	High	1994
888	Low	1989
36	Medium	1998

FIGURE 1-13   Marriage of a point geographic database and a relational database table

Lines

As mentioned earlier, a GIS can store and analyze pseudo-one-dimensional entities such as roads. Here, each RDB table record would refer to segments of a roadway between intersections. The attributes might be number of lanes, highway number, street name, pavement type, length, and so on. See Figure 1-14.

Again: A GIS is the marriage of a (geo)graphical database and an attribute database.
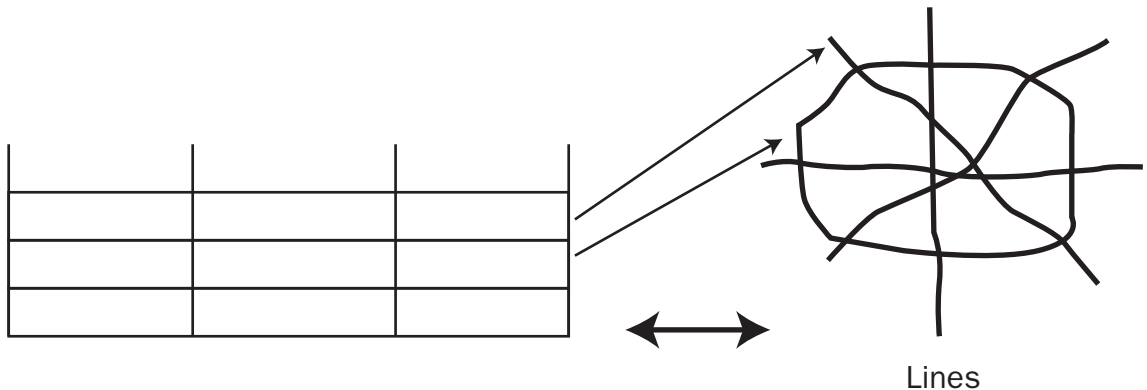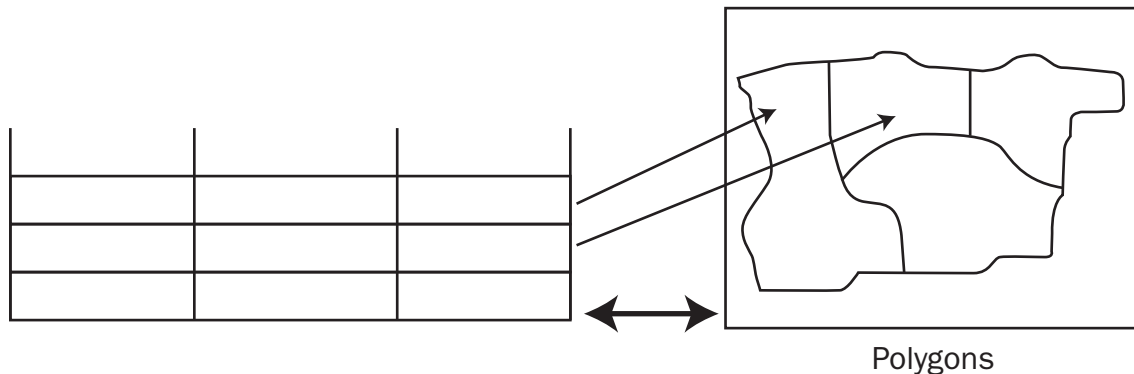
FIGURE 1-14   Marriage of a line geographic database and a relational database table

Areas

A GIS needs to be able to store information about areas, as well as points and lines. For example, such areas might be ownership parcels. Each parcel would be delineated by the lines determined by a land surveyor. Those lines define an area. A record would exist in the relational database that would correspond to the area. The attributes in the record might be owner's name, tax identification number, area of the parcel, and perimeter of the parcel. See Figure 1-15.

Did I mention? *A GIS is the marriage of a (geo)graphical database and an attribute database.*

Polygons

FIGURE 1-15   Marriage of a polygon geographic database and a relational database table

Computer Software: In General

Computers, considered at the most fundamental level, do only three things:[26] get input, manipulate data, and produce output; that is, read bits, stir bits, and write bits. What tells the computer what to do? In the early days of computers, more than half a century ago, each individual instruction to a computer came from outside, one at a time. This process was soon automated so that external media, such as perforated paper tape or magnetic tape, contained the instructions. Then, several scientists[27] got the idea of placing the instructions in the store of the computer itself. This major breakthrough in computer development allowed the computer to execute one group of instructions and then, on the basis of testing bits in its memory (e.g., Is number "a" larger than number "b"?), execute another group of instruction in a different place in that memory. This let the computer simulate "reasoning" and "decision making." This concept of a *stored program* revolutionized computer use.

The modern computer may hold an immense number of data sets and store a large number of programs. If a given program or data set not being used usually resides in "slow" disk (usually electromechanical) memory. Dealing with large numbers of diverse elements requires a management scheme. With computers this takes the form of an *operating system* that allows the user to execute specific programs

---

[26]Fundamentals that we'll discuss further in Chapter 6.
[27]John von Neumann, J. Presper Eckert, John Mauchly, Authur Burks, Maurice Wilkes, and others.

when desired, connects the computer to other computers (e.g., by way of the Internet or other network), makes copies of files, and performs many housekeeping activities.

Writing sets of instructions to be stored in the computer's memory is known as *programming*. Developing and marketing *programs* (software packages) is an immense business. The operating system that your computer uses is a software package. The word processor used to write this text is a software package. And the GIS you are about to learn, known as ArcGIS from Esri, is a software package, or rather a suite of software packages.

Computer Software: ArcGIS in Particular

ArcGIS is an integrated GIS that consists of different principal parts. We are working with *ArcGIS Desktop* software, which is an integrated suite of advanced GIS applications. Esri also has interfaces for managing geodatabases in a database management system (DBMS). Also there is an Internet-based GIS for distributing spatial data and services. Esri also has software extensions such as ArcPad, which provides GIS capability for palmtop or handheld computers.

ArcGIS is called a *scalable product* because it allows for deployment of GIS of many different "sizes." A user can choose to have only a system running on a single computer—a *personal GIS*—that allows viewing and simple selection of spatial data without capabilities of editing or extensively analyzing that data. At the other end of the scale, a *multiuser GIS* can serve an entire company or governmental agency. This *enterprise GIS* allows all the capabilities of the software. Options also include systems of intermediate functionality known as *workgroup* GIS and *department GIS*.

For many products, the price charged depends on the features and utilities that the product provides. When one manufacturer produces many different products, users must learn how to combine and package them to provide the correct degree of utility required (you do not need a school bus if a sedan will do). This is also true of Esri products.

Software Focus of This Textbook

In this introductory text, we confine ourselves largely to ArcGIS Desktop. The other products pertain primarily to managing GIS data for large applications or organizations and to distribute GIS capability and data to users over networks. Our emphasis is rather on the analysis, synthesis, and display of spatial and attribute information. You will see an exercise on publishing map data on the Internet.

ArcGIS Desktop itself is divided in a number of different ways. It has three main software packages:

❏ ArcMap

❏ ArcCatalog

❏ ArcToolbox

ArcMap lets you put together graphic and geographic elements to make sophisticated maps and interact with them. It allows you to obtain information from those maps and the attributes associated with their components, by using a variety of processing methods.

ArcCatalog is mainly an operating system for geographic data. ArcCatalog allows you to set up shortcuts to reach particular files of data quickly and easily and examines those files visually and textually. You can

use ArcCatalog to rename and copy spatial data sets. You can use a part of ArcCatalog, called the Catalog Tree, to place data into ArcMap.

You use ArcCatalog or ArcMap to access ArcToolbox, which provides immense geo-processing and analysis capability. It has tools to do 3-D analysis, spatial analysis, and spatial statistical analysis, to convert from one spatial data paradigm to another (vector to raster, and vice versa), and to convert spatial data into myriad geographic projections.

These three packages are fundamental to ArcGIS, but you have probably noticed many other Arc*Xxxxxxx* terms floating around. As a scalable product, ArcGIS desktop allows you to purchase what you need. The different levels of capability are bundled into four different packages. In terms of increasing utility (and cost) they are as follows:

ArcReader (free)

ArcView (as of version 10.1 of the software it is called "Basic")

ArcEditor ("Standard")

ArcInfo ("Advanced")

Knowing the capabilities of each is not important for our immediate purposes. You should know that some of these terms are recycled from earlier Esri product names. In particular, ArcView 10.x is a different animal from ArcView 3.x. ArcView 10.x is a level of capability of ArcGIS Desktop; ArcView 3.x is an older, but still supported and useful, independent software package.