

GIS COMMONS: AN INTRODUCTORY TEXTBOOK ON GEOGRAPHIC INFORMATION SYSTEMS



Michael Schmandt
Sacramento State University

GIS Commons: An Introductory Textbook on Geographic Information Systems

Michael Schmandt

Sacramento State University

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (<https://LibreTexts.org>) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by [NICE CXone Expert](#) and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptations contact info@LibreTexts.org or visit our main website at <https://LibreTexts.org>.

This text was compiled on 05/19/2026

TABLE OF CONTENTS

Licensing

1: Introduction to GIS and Mapping

- 1.1: Introduction
- 1.2: Definitions Of Geographic Information Systems (GIS)
- 1.3: Manual Gis And Computer-Based GIS
- 1.4: GIS Education And Training
- 1.5: Concepts
- 1.6: Data Model

2: Data Input

- 2.1: Introduction
- 2.2: Phase 1- GIS Database Design
- 2.3: Phase 2- Data Acquisition
- 2.4: Phase 3- Data Capture

3: The Earth and Map Preprocessing

- 3.1: Introduction
- 3.2: Earth
- 3.3: Map Preprocessing
- 3.4: Map Verification and Editing

4: Attribute Tables and Data Preprocessing

- 4.1: Introduction
- 4.2: Attribute Data
- 4.3: Principles Of Database Management – Vector
- 4.4: Principles Of Database Management – Raster
- 4.5: Attribute Preprocessing And Editing
- 4.6: Attribute Verification

5: Analysis

- 5.1: Introduction
- 5.2: Selection and Measurement
- 5.3: Overlay Analysis
- 5.4: Neighborhood Operations
- 5.5: Connectivity Analyses

6: Output

- 6.1: Introduction
- 6.2: Map Communication
- 6.3: Map Types
- 6.4: Constructing Thematic Maps
- 6.5: Types Of Output

[Index](#)

[Glossary](#)

[Detailed Licensing](#)

Licensing

A detailed breakdown of this resource's licensing can be found in [Back Matter/Detailed Licensing](#).

CHAPTER OVERVIEW

1: Introduction to GIS and Mapping

[1.1: Introduction](#)

[1.2: Definitions Of Geographic Information Systems \(GIS\)](#)

[1.3: Manual Gis And Computer-Based GIS](#)

[1.4: GIS Education And Training](#)

[1.5: Concepts](#)

[1.6: Data Model](#)

This page titled [1: Introduction to GIS and Mapping](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.1: Introduction

Making good maps can be challenging, time consuming, and expensive, but recently, a new set of cheap and free mapping tools has enabled almost anyone with a computer to easily make a map—but good maps are not usually the result. They have the computer and software, but the new mapmakers lack the mapping concepts, principles, and methodologies. Their maps are often improperly designed and do not communicate easily nor effectively.

This e-text wants to change that by helping you create, analyze, and produce maps that communicate more effectively. By the use of symbols, colors, shades, and words, maps help us communicate with more impact; they make what we want to say attractive, compelling, convincing, and clear. This e-text focuses primarily on a technology called Geographic Information Systems (GIS), but most of the chapter's concepts are applicable to other geotechnologies including remote sensing, global positioning systems (GPS), Internet mapping, and virtual globes.

This page titled [1.1: Introduction](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.2: Definitions Of Geographic Information Systems (GIS)

Let's define GIS first.

Definition #1

Over the past 20 to 30 years, many authors (Dept of the Environment, 1987; Rhind, 1988; Parker, 1988; and Bolstad, 2002) have defined GIS, and most of their definitions are similar to one another. The definitions generally refer to a system of computer hardware, software, and people that support the capture, management, analysis, and display of spatial data.

It is a decent definition, but to understand GIS better, you should break it down into its four main subsystems the way Marble and Peuquet (1983) did in perhaps the first widely used GIS textbook. GIS have:



Figure 1.1: The four GIS subsystems.

Much of this e-text is divided into chapters that mirror these four GIS subsystems. Chapter 2 looks at data and the data input system. Chapters 3 and 4 focus on the graphic and database portions of GIS software. These two chapters finish by covering various “housecleaning” processes (preprocessing functions) that manipulate the data files to make them ready for analysis. Chapter 5 looks at how to analyze the datasets. Output is covered in Chapter 6.

Definition #2

An even shorter definition equates GIS to a “spatial database”, but you must focus on both of those words carefully. To do this, think of a computer screen displaying a simple parcel map.

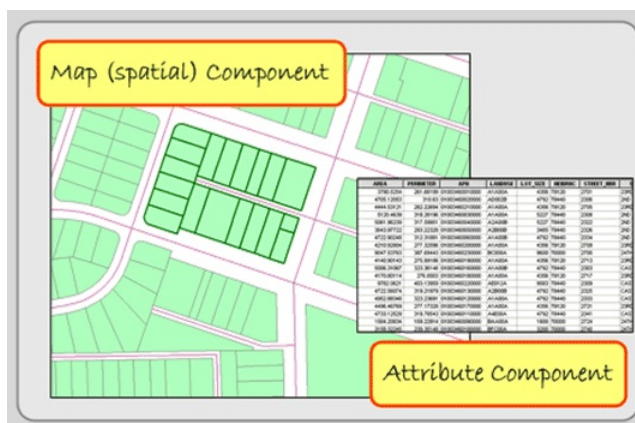


Figure 1.2: The two parts of a GIS.

Each parcel is a separate feature on the map, but they are more than just features on the screen, the computer stores many database characteristics about the feature like the parcel’s identification number and its owner’s name. In other words, there are two parts to a GIS: a map (or spatial) component and an attribute (or database) component. By making this link between the map and the stored attributes, GIS becomes a powerful tool for addressing and analyzing geographic data and environmental issues. This is its appeal.

GIS programs are capable of handling large and diverse geographic datasets, and we increasingly rely upon them for analyzing and making decisions.

The two definitions above emphasize GIS as a computer system; a simple definition for a computer system differs only in that GIS handles spatial data. Although the second definition gets closer, both definitions, do not give adequate weight to the geographic component of GIS.

Definition #3

It is no accident that GIS begins with G. A good understanding of GIS begins with geography and specifically with a geographic perspective, which is a way of organizing and thinking about portions of the Earth spatially. Again, look at the map above. You know that it and many other maps depict the locations of people and things (like roads, important buildings, parks, etc.). You also know that there are connections or relationships between the people and things drawn on the map. These relationships help explain the spatial patterns you see on the map. The process of finding, showing, explaining, and even predicting geographic patterns is at the heart of both geography and GIS.

In this context, GIS can be defined as a tool of exploration that helps us explore geographic (or spatial) patterns. At a minimum, it aids us in describing these patterns, but GIS can go beyond simply description to help us investigate and understand why these patterns (sometimes called distributions) exist, the impacts these patterns have on our life and land, and to discover potential future geographic patterns.

You can address five types of geographic questions with GIS and other geotechnologies (ESRI, 1992, pp):

1. What is at...? This basic question looks at what is at a particular location. An example might include, what is at the corner of Main Street and 12th Avenue?
2. Where is it? This question could simply inquire as to the location of something specific (like the nearest market or bookstore), or it can be a more challenging question that explores what locations meet a specific condition. For example, a city might identify all of the parcels that are larger than 5 acres, vacant, zoned commercial, and within $\frac{1}{2}$ mile of a freeway on-ramp.
3. What has changed since? Over a portion of the Earth's surface, how have conditions changed over time? An example is a county that identifies wetland areas that have decreased in size over the past 10 years.
4. What spatial patterns exist? This question describes and compares spatial patterns at different locations. It attempts to find spatial patterns—perhaps the concentration of phenomena. The process of finding, showing, and explaining geographic patterns is frequently termed spatial analysis. Stewart Fotheringham defines spatial analysis as manipulating spatial data to extract additional meaning. In a GIS context, spatial analysis asks two questions:
 - a. What is the relationship between two or more datasets that occupy the same location? For instance, you might see a direct relationship between a region's varying elevation and the amount of rainfall that falls across it.
 - b. What geographic variations exist over space? All geographic phenomena vary in their intensity over space. Consider fertility rates across the United States. Some areas are high and others low. To answer this question fully, you need to describe and explain these patterns.
5. What if...? "What if?" questions involve scenarios that differ when you change the model's parameters. An example includes looking at what happens to an area's population when a freeway is built through the area under different constraints.

These questions examine relationships among various geographic phenomena, and you use GIS and other geotechnologies to explore and help you answer these questions. As the geographer Ron Abler states, "GIS technology is to geographical analysis what the microscope, the telescope, and computers have been to other sciences..." (1988, 40). No other technology looks so closely at the spatial relationship of phenomena, and it aids geographers and others that use spatial data to describe, analyze, and predict spatial relationships and patterns. As the geographer Michael DeMers states, "We can now see deeper and farther than we could before, allowing us to map more of what is present on the landscape and to ask questions that could not have been imagined" (2003, 3).

This page titled [1.2: Definitions Of Geographic Information Systems \(GIS\)](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.3: Manual Gis And Computer-Based GIS

Abler's and DeMers' quotes may give you the impression that GIS is a new concept. Although the term GIS has been around for more than 30 years, the concepts surrounding GIS are old, and even the practice of doing GIS began before computers. The difference today is that GIS is computerized, but there is nothing a computer can do that cannot be done, at least theoretically, by hand if you had enough time, money, and energy. Computers process numbers and mathematical equations far quicker and more accurately than people can by hand or with the use of a calculator. Yet, before the concepts behind GIS were transferred to computers, people were doing manual GIS. They just combined spatial and attribute data on various types of media including hard-copy maps, hard-copy overlays (acetate or vellum), aerial photographs, written reports, field notebooks, and—of course—their eyes and minds.

With manual GIS, a large base map was often placed on a tabletop, and a series of transparent overlay maps, drawn at the same scale, were placed on top of the base map. One would then look for relationships among the base map and the features on the transparent overlays. Frequently, spatial data were copied from one map (or aerial photograph) to another. This took time, and because of it, many great ideas about the relationships of the Earth's features (both physical and human) were not analyzed. These ideas were constrained by the amount of time it took to do the analysis. Still, some impressive manual GIS projects did occur.

The much-repeated example of Dr. John Snow's Cholera map is a great example of manual GIS. In the 1840s, a cholera outbreak killed several hundred residents in London's Soho section. Snow, a physician, located the address of each fatality on a hand-drawn base map and soon a cluster of cases was visible (Figure 1.3).

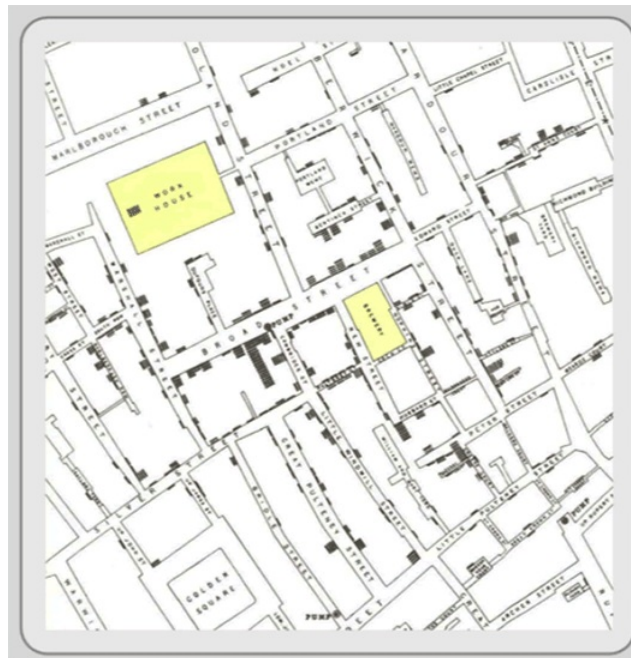


Figure 1.3: Dr. John Snow's Cholera Map of London's Soho.

Then on the base map, over the streets and fatalities, he drew the locations of water wells. Familiar with the idea of distance decay, he knew that people might go a far distance to purchase a product that was cheaper, but they would go to the nearest well because water was free and heavy to carry. Snow could see that the fatalities clustered largely among those who lived near the Broad Street water well. He and his students took the handle off the water pump, and new cholera cases dropped rapidly. By disabling the pump, Snow demonstrated the spatial relationship between cholera fatalities and the Broad Street water well, and, more importantly, he established the relationship between cholera and drinking water.

Even with the advent of computers, it took decades for GIS applications to get to where they are today. The largest and most powerful computers were mainframes that were available to some academics and government officials, but not to many researchers. In the 1980s, most GIS applications ran on workstation computers tied to mainframe computers because the early microcomputers (IBM, Apple, etc.) did not have enough memory, storage capacity, or processing ability. Today's personal

computers, however, are fast, capable of storing and processing large datasets, and can process multiple tasks simultaneously. This enables many academics, government agencies (from local to federal), organizations, and small and large businesses to use GIS. Computer-based GIS has its advantages, but requires educated users.

This page titled [1.3: Manual Gis And Computer-Based GIS](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.4: GIS Education And Training

GIS, virtual globes, GPS, and remote sensing are revolutionizing and revitalizing the discipline of geography. This has greatly improved the outlook for geography students and others who use spatial data in securing jobs. Along with nanotechnology and biotechnology, the U.S. Department of Labor identified geotechnology as one of the three most important emerging and evolving fields of study (Gewin, 2004).

Despite the good news, students often run high on anxiety when it comes to securing future employment, and they often ask for advice on what they should study to secure a GIS position. They ask about software, course selection, whether they should get a certificate, and even what should be their major or specialization. GIS professionals, administrators, and academics all have their own ideas and biases, but it usually comes down to this question: What is more useful, GIS training or GIS education? Students can usually answer this question if they can answer another related question: Do you want to apply GIS or focus on the technology itself?

GIS Education

If you want a position using GIS, you need to obtain conceptually both a geographic perspective and an idea of what one can do with a GIS. Both of these are conceptual topics that primarily stay away from the practical details of any particular GIS program. The focus is on applying GIS to answer real-world or discipline specific questions. It involves knowing about your data sources, the real-world processes of your discipline, and how one can use GIS to answer specific questions. GIS software training is a by-product, not the focus.

Specifically, a GIS education should provide students with:

1. A geographic perspective. This is the process of finding, showing, explaining, and predicting geographic patterns (discussed briefly under the third GIS definition above).
2. GIS concepts and cartographic principles. GIS concepts like buffering and overlay have their origin outside of GIS, but a number of procedures like these are closely tied to the technology. After becoming familiar with these processes, you need to determine which specific GIS processes are needed—and the order of these processes—to complete your projects. As Adena Schutzberg suggests, your objective is to “...get your head around how GIS ‘thinks’” (Schutzberg, 2003).

GIS Training

Who would argue that training is not important? No doubt, businesses and agencies benefit when new hires already have the training to run their specific software programs. That said, it is easier to train someone in the use of a specific program if they understand the underlying concepts.

The problem with most GIS training is that the trainees are learning how to perform a specific set of routines. They are not primarily learning the spatial concepts relating to those routines although they may be secondarily obtained with work experience. Yes, training is important, but you will not “get your head around how GIS thinks” with training alone. Many of the jobs associated with training alone are data entry jobs that many might find tedious over time and frequently one’s advancement is limited. To avoid this, but stick to the training route, one can develop extensive database and programming skills. With these skills, you can usually rise up to better paying positions.

GIS Experience

Perhaps the best piece of advice I could give is that both education and training are important, but so is on-the-job experience. For those just starting out, GIS experience comes in a couple forms: internships and research assistantships.

An internship is a valuable way to obtain on-the-job experience. A good internship allows you to be part of the agency or organization’s day-to-day activities. They are usually commonly available, especially with government agencies.

To find these opportunities, talk with your academic advisor, go to your campus’s career services office, network at GIS seminars and meetings, or go directly to government agencies, organizations, and local companies that interest you. If you go directly to agencies, organizations, and businesses, you should provide the name of a faculty member as a reference, an updated resume, and you may want to volunteer your services.

Be sure that you have the appropriate skills for the internship and that you stick with the internship for a specified period. If you leave it abruptly, it hurts the chances of future students because many businesses and agencies have a tenuous relationship with internships in general, and they might decide that interns are too much trouble. Think of internships as necessary work experience.

Internships are important because you are more employable after this experience, it provides you with a list of contacts for future job opportunities, and it gives you a sense of the type of job you might want after college. Additionally, you may be offered a position at the agency where you interned.

Research assistantships are student positions where you assist a faculty member with the faculty member's research. Some of these assistantships may involve GIS or other geotechnologies. Assistantships are available in many university academic departments, especially those that have graduate programs. Interested students should contact the department directly. Usually the assistantships are supported by external grant funding, so they are temporary positions but valuable experiences.

This page titled [1.4: GIS Education And Training](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.5: Concepts

Maps as a Model of Reality

The real world is too complex and unmanageable for direct analysis and understanding because of its countless variability and diversity. It would be an impossible task to describe and locate each city, building, tree, blade of grass, and grain of sand. How do we reduce the complexity of the Earth and its inhabitants, so we can portray them in a GIS database and on a map? We do it by selecting the most relevant features (ignoring those we do not think are necessary for our specific research or project) and then generalizing the features we have selected. Chapter 6, as well as later portions of this chapter, covers the selection and generalization process in more detail. For now, let's focus on features.

Features

As described in Definition #2 (and Figure 1.2), conceptually, there are two parts of a GIS: a spatial or map component and an attribute or database component. Features have these two components as well. They are represented spatially on the map and their attributes, describing the features, are found in a data file. These two parts are linked. In other words, each map feature is linked to a record in a data file that describes the feature. If you delete the feature's attributes in the data file, the feature disappears on the map. Conversely, if you delete the feature from the map, its attributes will disappear too.

Features are individual objects and events that are located (present, past or future) in space. In Figure 1.2, a single parcel is an example of a feature. Within the GIS industry, features have many synonyms including objects, events, activities, forms, observations, entities, and facilities. Combined with other features of the same type (like all of the parcels in Figure 1.2), they are arranged in data files often called layers, coverages, or themes. In this text, we use the terms feature and layer.

In Figure 1.4 below, three features—parcels, buildings, and street centerlines—of a typical city block are visible. Every feature has a spatial location and a set of attributes. Its spatial location describes not only its location but its extent. While “location” may be simple to grasp, it is difficult to locate features accurately and precisely. Accuracy and precision are examined in Chapter 2, but, in brief, precision deals with the exactness of the measurement. For example, some input devices, like GPS, have a certain error. They may be precise within a certain accuracy range if used correctly. Accuracy is the degree of correspondence between the data and the real world.



Figure 1.4: Each feature in the layers above has a spatial location and attribute data, which describes the individual feature.

Besides location, each feature usually has a set of descriptive attributes, which characterize the individual feature. Each attribute takes the form of numbers or text (characters), and these values can be qualitative (i.e. low, medium, or high income) or quantitative (actual measurements). Sometimes, features may also have a temporal dimension; a period in which the feature's spatial or attribute data may change.

As an example of a feature, think of a streetlight. Now imagine a map with the locations of all the streetlights in your neighborhood. In Figure 1.5, streetlights most are depicted as small circles. Now think of all of the different characteristics that you could collect relating to each streetlight. It could be a long list. Streetlight attributes could include height, material, basement

material, presence of a light globe, globe material, color of pole, style, wattage and lumens of bulb, bulb type, bulb color, date of installation, maintenance report, and many others.



Figure 1.5: Location of street lights, represented with a red circle, and their attributes.

The necessary streetlight attributes depends on how you intend to use them. For example, if you are solely interested in knowing the location of streetlights for personal safety reasons, you need to know location, pole heights, and bulb strength. On the other hand, if you are interested in historic preservation, you are concerned with the streetlight's location, style, and color.

Now continue thinking about feature attributes, by imagining the trees planted around your campus or office. What attributes would a gardener want versus a botanist? There would be differences because they have different needs. You determine your study's features and the attributes that define the features.

Points, Lines and Polygons

Now think of the feature's shape on a map. Single or multiple paired coordinates (x, y) locate individual features in space and define their unique shape. The x and y values of each coordinate pair are associated with real world coordinate systems, which are discussed in Chapter 3. For now, let's focus on the shape of features, which take the generalized form of points, lines, and polygons (see Figure 1.6).

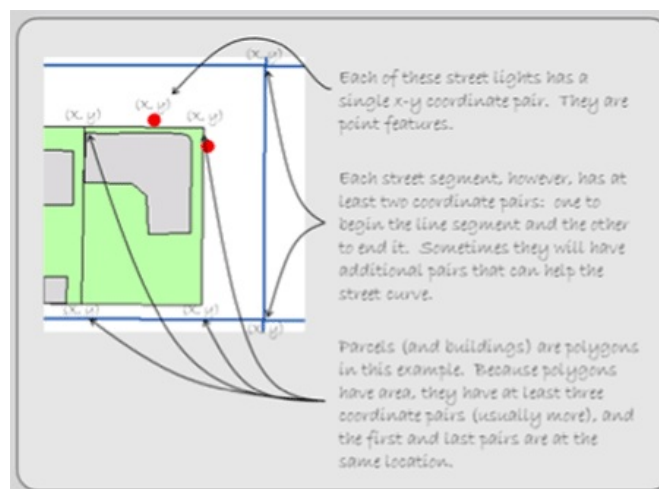


Figure 1.6: Each feature has a spatial position.

Points

Points are zero dimensional features (meaning that they possess only one x, y coordinate set) whose location is depicted by a small symbol. What you represent as a point depends on your study. Examples include streetlights, individual trees, wells, car accidents, crimes, telephone polls, earthquake epicenters, and even, depending on scale, buildings and cities.

Lines

Lines are formed from a sequence of at least two paired coordinates. The first pair starts the line and the last ends it. Two coordinate pairs form a straight line. Additional paired coordinates can form vertices between the starting and ending points that allow the line to bend and curve. Having length (which can be measured) but no width, a line feature is one-dimensional. Again, what is represented as a line depends on your study, but street centerlines, utility lines, canals, railroad tracks, rivers, flight paths, and elevation contour lines usually form lines.

Polygons

Polygons are features that have boundaries. Formed by a sequence of paired coordinates, polygons differ from lines in that the starting point is also its ending point. This provides polygons with both length and width, so these two-dimensional features can calculate the area contained within the feature. What is represented as a polygon differs from study to study, but examples include lakes, forest stands, buildings, counties, countries, states, and census districts.

Topology

One of the most important concepts associated with GIS and other geotechnologies is topology. As features are added to a GIS, they form spatial relationships—called topology—with each other (both with features within the same layer and with features in different layers). You might find topology a confusing term partly because it has both spatial and mathematical properties. For our purposes, you can define it as the spatial relationships among features. It deals with where features are in relation to one another and how they are related to one another. These relationships take the form of simple distance calculations from one feature to another, but also include the more complicated issues of adjacency and connectivity.



Figure 1.7: Fire hydrants are located along streets (so fire trucks can connect) and adjacent to structures that can burn.

1. **Distances between features.** The geographer Waldo Tobler created what some call the “first law of geography”, which states, “Everything is related to everything else, but near things are more related than distant things.” (1970, 236). This type of topology looks at the spatial relationships of where features are located. Consider the spatial locations of streets, bike lanes, sidewalks, and streetlights. They are positioned to work together. This is a type of topology; a relationship exists. Notice the relationship between the fire hydrant, building, and street in Figure 1.7.
2. **Adjacency.** Adjacency focuses on a single type of feature (like streets or buildings) and whether parts of two or more individual features are shared (or contained). Think of an individual street segment, and how it is most likely physically connected to at least one additional street segment at one or both of its ends. These adjacent street segments are in turn connected to additional segments, which in turn are connected to streets, forming a network. When a single point or line (like a boundary between two parcels) is shared by at least two features, the spatial data file stores only a single point or a single line to prevent duplication that could lead to errors. This topological relationship describes how features are related.
3. **Connectivity.** Also focusing on how features are related, connectivity specifies the way features are linked in a network. Even though a couple street segments may be physically connected in space, that does not mean that traffic can go in both directions. These are topological relationships that you can specify. Differing from adjacency, connectivity can include multiple feature types. For instance, you can determine the flow of water through connected pipe and valve features.

This page titled [1.5: Concepts](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

1.6: Data Model

Current GIS programs represent points, lines, and polygons differently. There are two fundamental models: raster and vector. Each model has its advantages and disadvantages, and neither is superior to the other in every situation. One data model may fit certain types of data and applications better than the other.

Raster

A matrix of rows and columns, the raster data model covers sections of the Earth's surface and represents features with cells or pixels. Pixels are the building blocks of the raster data model, and they are usually uniformly square and of consistent size within each layer. Each pixel represents a precise chunk of the Earth's surface; the geographic position of any cell can be determined. A specific attribute value, representing the condition of that specific portion of the Earth's surface (see figure 1.8), is associated with the pixel. If you need more than one attribute to describe the area contained within the pixel (and most likely you will), you need a second layer. The second raster layer gives you a second attribute. A third gives you a third attribute, and so on.

Individual cells and groups of cells represent the features of the real world (Figure 1.8). A point feature usually fills one cell while lines and polygons are constructed as a string or contiguous group of cells. Raster layers fill space; they describe what occurs everywhere in the study area. There are no blank spaces across the layer. "Empty" areas simply get a "0" value, but every pixel gets a value.

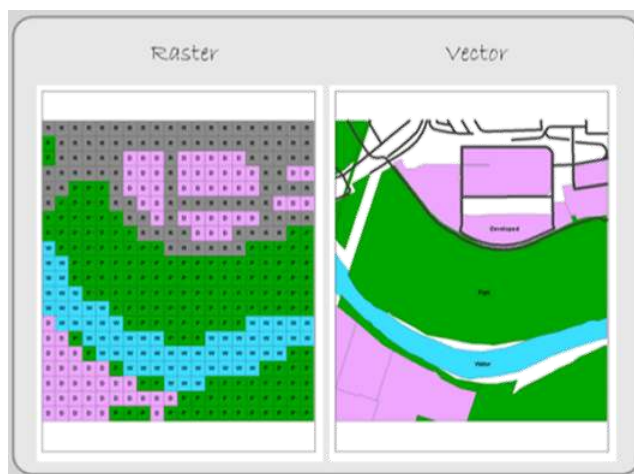


Figure 1.8: The raster and vector data models. Each stores features in a different way.

Conceptually, the raster model is simple. You take a portion of the Earth's surface, divide it into cells, and give each cell an attribute that represents that area. In the figure above, you might give each cell either a D (developed), P (park), or W (water). For those cells with both park and water, you can give these cells either another code PW (for park and water) or make a judgment as to what covers the majority of the cell. Another way to code these cells is with the percentage of the cell that is water. If 40 percent of the cell is covered by water, the cell gets a value of 40.

Vector

The vector data model uses discrete point and line segments to identify the locations of the Earth's features. Vector objects usually do not fill space like raster layers do; they depict where features occur and the space around those features is empty. Notice that there are white spaces in the vector model of Figure 1.8. No white spaces exist in the raster model; it covers the entire area.

Vector features are located with x, y coordinates. As described above, points are easy; they have one node (sometimes called a vertex). A node is a location in space that helps define the shape of point, line, and polygon features. As mentioned above, points have one coordinate pair that locates the feature in space. Lines have at least two nodes (their end points). Polygons have a minimum of three nodes to form an area. Lines and polygons usually have many more nodes that help define the course of the line or the polygon's area.

Contrasting with raster systems that record one attribute per layer, the vector data model can handle many attributes for each feature type. Different software programs have varying ways of organizing vector digital files, but usually they have at least two

files: one that stores spatial data and another that stores attributes.

The link between the spatial and attribute data files is made with a unique identifier. Each feature on the map and its corresponding attributes has a unique identifier that links the map feature to its database attributes. A type of unique identifier, a “key”, which links attribute files, is discussed in Chapter 3.

Raster versus Vector

Which is better? Although GIS users have their own personal favorite data model, the question of which is “better” is an incomplete question. There are advantages and disadvantages to both data models, so a better question is which is better for particular applications or datasets. Some in the GIS industry use the slogan “Raster is faster, but vector is corrector.” While this is a good starting point, it conceals the details. Yes, your computer can process raster data quicker, but today computer processors are so fast the difference may be negligible. Yes, vector output looks more accurate, but you can increase pixel resolution to something resembling vector resolution (this, however, greatly increases the database size). The following are some of the advantages and disadvantages of the data models:

Raster advantages:

1. **Easy to understand.** Conceptually, the raster data model is easy to understand. It arranges data into columns and rows. Each pixel represents a piece of territory.
2. **Processing speed.** Raster’s simple data structure and its uncomplicated math produce quick results. For example, to calculate a polygon’s area, the computer takes the area contained within a single cell (which remains consistent throughout the layer) and multiplies it by the number of cells making up the polygon. Likewise, the speed of many analysis processes, like overlay and buffering, are faster than vector systems that must use geometric equations.
3. **Data form.** Remote sensing imagery is easily handled by raster-based systems because the imagery is provided in a raster format.
4. **Some analysis functions** (surface analysis and neighborhood functions) are only feasible in raster systems. In addition, many new analysis functions appear in raster systems before migrating to vector systems because the math is simpler.

Raster disadvantages:

1. **Appearance.** Cells “seem” to sacrifice too much detail (Figure 1.9). This disadvantage is largely aesthetic and can be remedied by increasing the layer’s resolution.

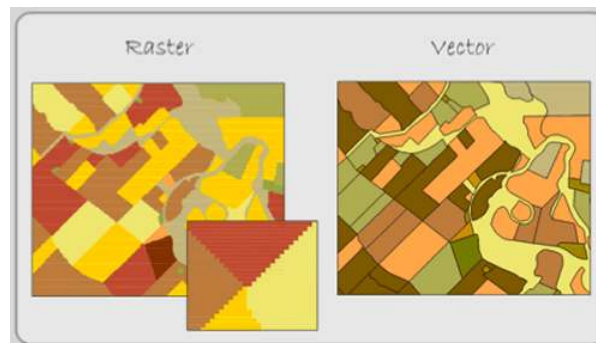


Figure 1.9: Comparison of raster and vector data models. Raster layers often appear pixilated and thus less accurate.

2. **Accuracy.** Sometimes accuracy is a problem due to the pixel resolution. Imagine if you had a raster layer with a 30 by 30 meter resolution, and you wanted to locate traffic stop signs in that layer. The entire 30 by 30 meter pixel would represent the single stop sign. If you converted this raster layer to vector, it might place the stop sign at what was the pixel’s center. Sometimes problems of accuracy (and appearance) can be resolved by selecting a smaller pixel resolution, but this has database consequences.
3. **Large database.** As just described, accuracy and appearance can be enhanced by reducing pixel size (the area of the Earth’s surface covered by each cell), but this increases your layer’s file size. By making the resolution 50 percent better (say from 30 to 15 meters), your layer grows four times. Improve the resolution again by halving the pixel size (to 7.5 meters) and your layer

will again increase by four times (16 times larger than the original 30-meter layer). The layer quadruples because the resolution increases in both the x and y direction.

Vector advantages:

1. **Intuitive.** In our minds, we picture features discreetly rather than made up of contiguous square cells.
2. **Resolution.** If the locations of features are precise and accurate, you can maintain that spatial accuracy. The features will not float somewhere within a cell.
3. **Topology.** Although the raster data model preserves where features are located in relation to one another, they do not represent how they are related to one another. This complex form of topology can be constructed in most vector systems, so you can track the connections in a municipal water network between pipe and valve features and thus track the direction and flow of water.
4. **Storage.** Vector points, lines, and simple polygons use little disk space in comparison to raster systems. This was once a major consideration when hard-disk storage was limited and expensive.

Vector disadvantages:

1. **Geometry is complex.** The geometrical algorithms needed for polygon overlay and the calculation of distances, depending on the projection/coordinate system used, require experienced programmers. This is not usually a problem for most GIS users since most functions are directly coded in the software.
2. **Slow response times.** The vector data model can be slow to process complex datasets especially on low-end computers.
3. **Less innovation.** Since the math is more complex, new analysis functions may not surface on vector systems for a couple of years after they have debuted on raster system.

This page titled [1.6: Data Model](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

2: Data Input

[2.1: Introduction](#)

[2.2: Phase 1- GIS Database Design](#)

[2.3: Phase 2- Data Acquisition](#)

[2.4: Phase 3- Data Capture](#)

This page titled [2: Data Input](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.1: Introduction

As mentioned in Chapter 1, GIS must have capabilities for inputting data; various housekeeping functions to edit, store, and reshape data; analytical operations that manipulate the data into usable information, and output functions. In this chapter, we focus on data input.

There are three distinct phases to the data input process (Figure 2.1): In the first phase, the design phase, you identify and “conceptually” code all the needed features and attributes for your project. Next is the data acquisition phase, which involves acquiring the needed data from various agencies, storehouses, and organizations, and getting it into a format that your GIS program reads. Finally, in the data capture phase, you digitize hard-copy maps and data directly into your GIS and transform existing digital data into a format your GIS reads.

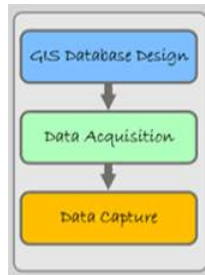


Figure 2.1: Three distinct phases to the data input process

This page titled [2.1: Introduction](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.2: Phase 1- GIS Database Design

What is your goal or your research question? How should you proceed? You need to define your objective at the very beginning. Having a well-defined research question, goal, or even multiple goals is the key to a successful GIS project because it guides the project's input, analysis, and output stages. Spend time and thought on the design of your GIS because good planning results in successful projects.

Start by thinking about the people, land, and the issues in your study. This has a direct bearing on what datasets (features and attributes) are needed. Next, think about how you will analyze the data. This could affect your choice of GIS software and your data model (vector or raster). Even envision how you would like to present your results. In other words, think through the entire project. If you are working for someone else, consider your employer's purpose and understand how the company, agency, or organization functions.

It seems like everyone wants to use GIS for their projects today. It is both trendy and powerful, but you should ask yourself whether GIS is the appropriate tool. It may not be. To determine if GIS can help achieve your objectives, think again about the relevant features (land, people, and issues) of your study and whether these can be displayed geographically. Almost all features have geographic locations, but you must further decide if the feature's "geography" is important. One might be interested in population and economic statistics like income, population density, age, and ethnicity, but are you interested in how these variables differ across neighborhoods, across cities and towns, or across space in general? If location does not play a role in your study, then the data are "aspatial," and you should close this e-text.

If you can conceptualize your variables as features on a map or depicted within geographic boundaries like census tracts, ZIP Codes, or agricultural fields, GIS can aid your project. As an employee or a researcher, you may already have your goal or research topic chosen for you by your superiors. Your goal or research question might be something like "I want to know the best places to plant sorghum in Mauritania when the new dam goes in." Alternatively, if you are in a large planning department, you might have multiple goals, and then you must try to satisfy all the tasks, which might include updating and producing high quality maps, generating lists of residents that need to be contacted when a zoning amendment or a liquor license is requested, or determining which roads need resurfacing. It all starts with your goals. If your goals are fuzzy in your own mind, stop and sharpen them. Nothing in the design phase is more important.

Although all components of a GIS project need to be planned, including what software and hardware you will use and what procedures and people will guide your operation, this chapter focuses on data.

In this first phase of the input process, you determine what features and attributes are needed and how they should be coded. This starts with identifying each type of feature and their related attributes, but it goes beyond identification to include several important planning decisions. Figure 2.2 features a list of planning tasks. A discussion of each item follows.

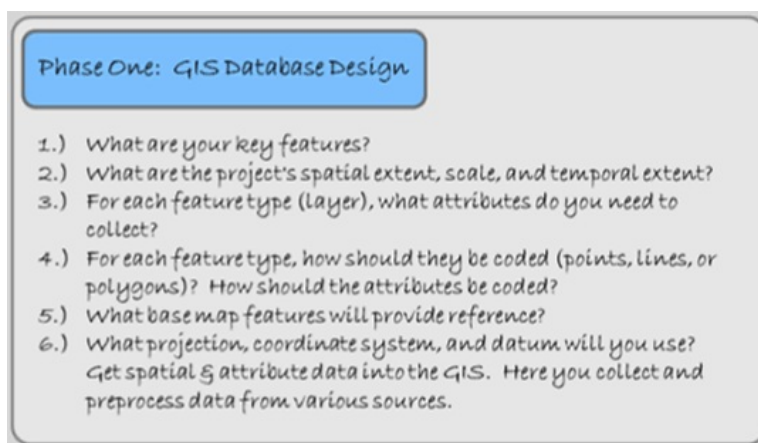


Figure 2.2: Key questions to ask yourself in planning a GIS database.

1. Determine Your Features

What features are necessary? Think back to your project's goals. If you want to analyze a particular plant species' distribution, it

may be necessary to have a feature devoted to the specific plant type. Equally important, however, are the other features—nearby plant species, soil types, climate conditions, land tenure practices, and landform conditions like slope and aspect. These other features, along with many others, play a role in the distribution of your plant. If you are developing a GIS database for a city’s planning department, you will want layers for many features including streets, parcels, parks, water, sewer, electricity, and buildings. In this case, the features are obvious.

The question of which features to use, however, may seem simple, but frequently it can be complicated. For example, you might want to interview hundreds of people in Modesto, California about their family income and quality of life. Is it appropriate to construct a layer with a point location at the home of each respondent, or would it be better to aggregate the responses of the individuals into neighborhood or census tract boundaries? In the first scenario, your feature layer might be called “respondents”, and each point feature would be located at the home of an individual respondent. Attributes would be stored within each individual feature. In the second scenario, you would have a feature called “census tracts” and the aggregated responses of all the individual respondents would be contained within the appropriate census tract’s data table. In other words, in the second scenario you would not have a respondent layer; you would have a census tract layer with aggregated responses. While the appropriateness question is hypothetical, you can see the ethical issue. You may not want to produce a map or provide the data to others that has the exact locations of your respondents if it includes sensitive or personal data.

Yet as a general rule, it is better to input data in its most precise and detailed form. In the income and quality of life example, you could create the more precise respondent layer and later aggregate (see Chapter 5) the individual respondents to census tracts for output purposes. The public will not see the detailed data. Benefits of having precise data come in verifying your census tract results, and in changing the resolution of your study. For example, what if—down the line—during the analysis portion of your project, you discover that the census tract boundaries are too coarse. It would be difficult, and perhaps impossible, to change or disaggregate the aggregated data into smaller, more detailed boundaries. If you have the respondent layer, however, one can quickly re-aggregate the data to a different and finer boundary feature like census blocks or a neighborhood unit.

2. Determine the Project’s Spatial Extent, Scale, and Temporal Extent

You must determine the area and the period in which your project focuses. Sometimes it is obvious. If you are working for a county planning department, chances are your extent is the county boundary. Cities, however, need to go beyond their boundaries to sphere of influence borders that may lay beyond the city limits because these regions have a direct impact on the city and this region may someday become part of the city. Other boundaries, especially those in research projects, are more difficult to establish.

Along with the project’s spatial extent, you should think about an appropriate scale. There is a relationship between scale and detail (see Figure 2.3). Small-scale maps depict large territories, but they usually are less precise and may require that some reference layers be left out. Large-scale maps show smaller areas but comparatively include more detail. Although GIS allows one to zoom in at increasingly larger scales, data captured at a “small scale” become inherently inaccurate when zoomed in on. The desired scale affects both the amount and precision of the data to be collected and the scale at which the geography can be cartographically represented. There is more on this topic later in this chapter under Precision Errors.



Figure 2.3: Map scale. Small-scale maps depict larger territories, but large-scale maps present more detail.

Similarly, you may want to define a temporal extent. Is time an important variable in your study? Most GIS projects focus on the contemporary scene and ignore the past. With contemporary databases, the assumption is that the GIS database is up-to-date. That

may be a major assumption since some of the layers might be quite dated. If, however, you want to determine how much an area has changed, you need to define a period for your project. Does your project focus on neighborhood change or the shrinking of Central Asia's Aral Sea? Then you need to establish a temporal extent. Determining the temporal period helps you determine your project's needed attributes. It also makes you aware that you may have to look for features that are no longer present on the landscape.

3. Determine the Attributes for Each Feature Type

As described in Chapter 1, attributes are the characteristics of features. You need to identify the required attributes for each feature type. The more you can do this before you collect your data, the less you will retrace your steps and collect additional attributes later. Again, look to the project's goals for some clues to the necessary attributes. Also consider how you will analyze the features. You cannot use some analytical processes (like many statistical tests) if the attribute values that you collect are in an improper form to be used in a particular analytical process (see levels of measurement below).

One other thing to consider at this point is that some attributes (like a polygon's area, a line's length, and even the number of point features falling within polygon features) can be generated automatically by the software. Additional attributes can be created by multiplying, dividing, adding, subtracting, truncating, or concatenating attributes with other attributes, numbers, or characters.

4. Determine How the Features and Their Attributes should be Coded

Once you have decided on the features and their attributes, determine how they will be coded in the GIS database. Remember from Chapter 1, there is not just one way to code features. Although roads are usually coded as lines, they do not have to be.

Decide whether to code each feature type as a point, line, or polygon. Then define the format and storage requirements for each of the feature's attributes. For instance, is the attribute going to be in characters (string) or numbers? If they are going to be numbers, are they byte, integer, or real numbers? You will have to establish these database parameters before you enter data into the GIS. Look at the example below (Figure 2.4). Listed are some attributes (under Field Name) relating to the feature "streets". Notice that street "LENGTH" has a data type called double (a type of real numbers), and in this case, the database will store up to 18 numbers including 5 decimal places for the length of each individual street.

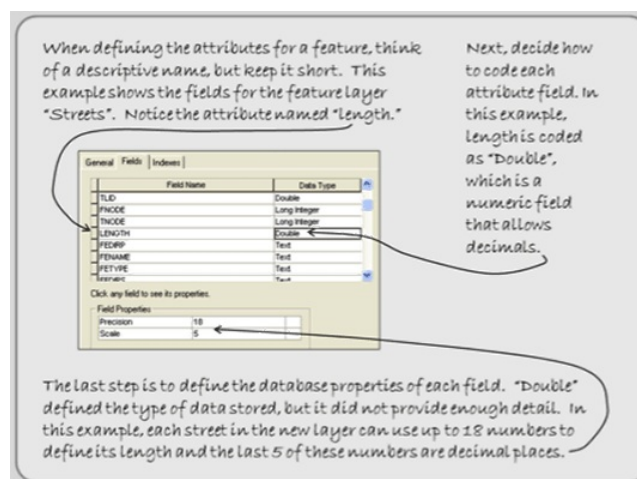


Figure 2.4: Each feature's attribute needs to be coded.

It is critical that you think about the value of your attributes before you code. Obviously, if one street segment needed room for 9 numbers to report its length, than 8 is not enough and the correct value could not be entered without modifying the field's length.

Also, while thinking about your attribute values, consider where it fits on the "levels of measurement" scale with its four different data values: nominal, ordinal, interval, and ratio. Stanley S. Stevens, an American psychologist, developed these categories in 1946. Although Steven's classification is widely used, it is not universally accepted. Some researchers have problems with the categories, and others with how categorization affects research. For our purposes, it is a useful way to conceptualize how data values differ, and it is an important reminder that only some types of variables can be used for certain mathematical operations and statistical tests, including many GIS functions. The different "levels" are depicted in Figure 2.5 and demonstrated using an example of a marathon race:

Nominal	Ordinal	Interval	Ratio
Runner ID	Order Finished	Time of Day Finished	Race Time
238	1	10:10	2:30
143	2	10:20	2:40
14	3	10:25	2:45
.	.	.	.
.	.	.	.
.	.	.	.
301	450	20:20	12:40

Figure 2.5: Levels of Measurement.

- **Nominal** data use characters or numbers to establish identity or categories within a series. In a marathon race, the numbers pinned to the runners' jerseys are nominal numbers (first column in the figure above). They identify runners, but the numbers do not indicate the order or even a predicted race outcome. Besides races, telephone numbers are a good example. It signifies the unique identity of a telephone. The phone number 961-8224 is not more than 961-8049. Place names (and those of people) are nominal too. You may prefer the sound of one name, but they serve only to distinguish themselves from each other. Nominal characters and numbers do not suggest a rank order or relative value; they identify and categorize. Nominal data are usually coded as character (string) data in a GIS database.
- **Ordinal** datasets establish rank order. In the race, the order they finished (i.e. 1st, 2nd, and 3rd place) are measured on an ordinal scale (second column in Figure 2.5). While order is known, how much better one runner is than the other is not. The ranks 'high', 'medium', and 'low' are also ordinal. So while we know the rank order, we do not know the interval. Usually both numeric and character ordinal data are coded with characters because ordinal data cannot be added, subtracted, multiplied, or divided in a meaningful way. The middle value, the "median", in a string of ordinal values, however, is a good substitute for a mean (average) value.
- The **Interval** scale, like we will discuss with ratio data, pertains only to numbers; there is no use of character data. With interval data the difference—the "interval"—between numbers is meaningful. Interval data, unlike ratio data, however, do not have a starting point at a true zero. Thus, while interval numbers can be added and subtracted, division and multiplication do not make mathematical sense. In the marathon race, the time of the day each runner finished is measured on an interval scale. If the runners finished at 10:10 a.m., 10:20 a.m. and 10:25 a.m., then the first runner finished 10 minutes before the second runner and the difference between the first two runners is twice that of the difference between the second and third place runners (see third column 3 Figure 2.5). The runner finishing at 10:10 a.m., however, did not finish twice as fast as the runner finishing at 20:20 (8:20 p.m.) did. A good non-race example is temperature. It makes sense to say that 20° C is 10° warmer than 10° C. Celsius temperatures (like Fahrenheit) are measured as interval data, but 20° C is not twice as warm as 10° C because 0° C is not the lack of temperature, it is an arbitrary point that conveys when water freezes. Returning to phone numbers, it does not make sense to say that 968-0244 is 62195 more than 961-8049, so they are not interval values.
- **Ratio** is similar to interval. The difference is that ratio values have an absolute or natural zero point. In our race, the first place runner finished in a time of 2 hours and 30 minutes, the second place runner in a time of 2 hours and 40 minutes, and the 450th place runner took 5 hours (see forth column in Figure 2.5). The 450th place finisher took over five times longer than the first place runner did ($12.667 \text{ hrs} / 2.5 \text{ hrs} = 5.0668$). With ratio data, it makes sense to say that a 100 lb woman weighs half as much as a 200 lb man, so weight in pounds is ratio. The zero point of weight is absolute. Addition, subtraction, multiplication, and division of ratio values make statistical sense.

5. Determine the Base Map Reference Features

What features are helpful to include? Add reference features that help people orient themselves within your study area even if you are not going to analyze these features. Major roads, rivers, and principal buildings are good examples of features that help orient map readers. These secondary features are often the easiest features to find on the Internet, and sometimes they are bundled with GIS software. In short, having these base-map features may not be important for analysis, but they are important for clarity.

6. Determine your Project's Projection, Coordinate System, and Datum

Before you collect or look for data, you should decide on which projection, coordinate system, and datum to use. These three terms, collectively termed “projection parameters”, are discussed in Chapter 3, but it is important that these parameters remain consistent throughout your layers. Consistency enables you to properly overlay your feature layers to produce maps and analyze feature relationships. Figure 2.6 shows an example of how a parcel layer is not properly overlaid upon street centerline and building layers due to small differences in parcel layer's projection parameters.

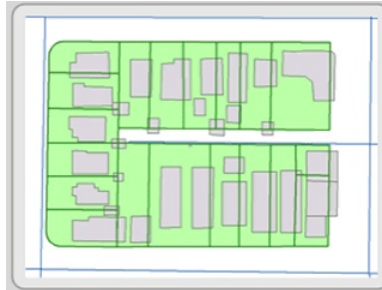


Figure 2.6: Due to differences in the projection parameters of the parcel layer, it is not lining up properly with the street centerline and buildings layers.

Deciding on the projection parameters upfront helps you select among nearly identical data files, if available, on the Internet because one file might already be in the desired projection, coordinate system, and datum. It is, however, rare to find everything the way you want it. Usually, you need to convert your layers, and GIS programs come with algorithms that convert between standard projections, coordinate systems, and datums. This topic is covered in Chapter 3.

Before choosing your projection parameters, do some investigating. If you are using mostly state GIS databases, it makes sense to use their parameters if they fit your needs. For instance, California has created its own Albers projection, and the vast majority of their GIS layers are in this projection. Local agencies, however, might use State Plane Coordinates or Universal Transverse Mercator. If substantial amounts of data are shared with a local jurisdiction, you may decide to use their projection parameters. You can change your projection parameters at any time in your project's life, but it takes time and organization.

The decision of which projection parameters to use can also be based on many other considerations. The project's scale, the desired projection properties (like the preservation of area, which is discussed in Chapter 3), the needed precision of positional accuracy, the attributes to be derived internally, and the project's location are among those questions that need to be evaluated before selecting a projection. Know something about projection parameters before you choose them for your project. For instance, if you desire to calculate area from the shape of your polygons, you need to use an equal area projection. If you do not use an equal area projection, the area measurements will be inaccurate.

Before actively collecting and inputting data, consider the six items listed on page 15. Next, find out what GIS datasets already exist, and—for those datasets that do not exist—the different ways to create the needed data. The next two sections of this chapter deal with these topics.

One last suggestion, when planning a GIS project, talk with GIS technicians and analysts at government agencies or researchers that have put together similar GIS projects. Their suggestions will save you time, and they could become a great contact for troubleshooting problems and providing you with needed GIS datasets.

This page titled [2.2: Phase 1- GIS Database Design](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.3: Phase 2- Data Acquisition

In the data acquisition phase, you obtain the data for your GIS. Getting all the data together (and in a suitable format) is the most costly and time-consuming task for any GIS project. Most estimates suggest that between 75 to 80 percent of your time is spent collecting, entering, cleaning, and converting data (phases 2 and 3 of this chapter). Before proceeding to data sources, we define the term “data” and look at various data issues including accuracy, precision, and metadata. Figure 2.7 features a list of data acquisition tasks.



Figure 2.7: Key tasks of the data acquisition phase.

Data

Data are frequently called facts. This definition suggests that they are pure, but all data are selected for a particular purpose and are shaped by that purpose. The term “raw data” implies an even greater purity—an objective truth—but even the most objective scientist has beliefs and knowledge that underlie data collection. Data obtained for any project are preconceived.

You can classify data as either primary or secondary and observable or non-observable. The geographer Frank Aldrich devised this simple data model (Figure 2.8) to depict the different data categories. Primary data (the light colored inner circle) are measurements that you or your team collects. They are usually derived from experiments or from fieldwork. Secondary data (the doughnut shaped darker ring) are datasets that someone else collects. These datasets, collected from experiments or fieldwork, were collected for a purpose other than your own. Most researchers prefer primary data because they have not been previously conceived and shaped. Still, secondary datasets are tremendously valuable if you determine how and why they were collected and if your project can accept those preconceptions.

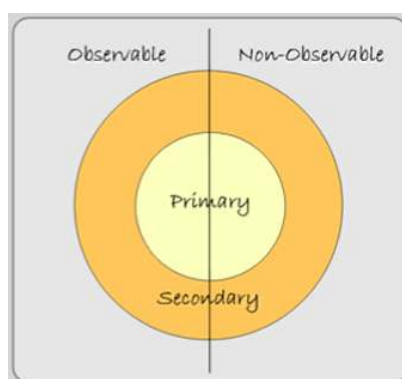


Figure 2.8: Aldrich's data model.

You can also classify data as either observable or non-observable. Notice in the figure above that a vertical line bisects the primary and secondary circles. Observable data (to the left of the bisecting axis) are when someone or something observes the characteristic or the behavior. Non-Observable data (to the right of the bisecting axis) are when respondents are asked questions in an interview or on a questionnaire, but the data gatherer does not physically observe the characteristic or behavior.

Combinations of these two types of characteristics are made. Primary Observable data are datasets you collect and observe. Secondary Observable data are datasets someone else collects and observes. Satellite imagery is an example. Primary Non-Observable data are datasets you collect, but the characteristic or behavior is not observed. Surveys and interviews that you administer are examples. Secondary Non-Observable data are datasets that someone else collects and does not observe. Census data falls into this category.

Data Evaluation

What is the quality of your data? Good datasets are both accurate and precise. There could be both spatial location problems and attribute errors. It is critical that you evaluate your data, especially your secondary datasets, because errors (if not corrected) can make a project's results worthless. Do not be lulled into a false sense of security when you receive or construct a dataset. Instead, evaluate your data. To better understand the types of errors that occur in GIS datasets, we turn to the term "error" and two parts of its definition.

Errors are obviously mistakes, and in the context of database errors, they are the result of two things: inaccuracies and data imprecision. Accuracy is the quality to which the data matches true and accepted values. In a layer, there can be inaccuracies in the location of features (and in their attributes). For example, features may be located in the wrong locations, not present when they should be, or a fictitious feature may be entered where one does not exist.

Precision is the exactness of a measurement or value, and it refers to the degree the value can be reproduced by similar data collection techniques. In other words, precision is a measure of exactness and repeatability. The precision of a feature's location could be accurate to a few inches or perhaps it has an acceptable precision of 10 meters. 10 meters may not seem precise, but it is usually good enough for many projects.

Many types of errors affect a GIS dataset's quality. Some are obvious, but others are difficult to uncover. Learn to recognize error and decide on an acceptable degree of precision for your project. The following discussion looks at the types of errors one might uncover. Chapters 3 and 4 address how to identify and fix these errors.

Accuracy Errors:

1. How old is the dataset? This is one of the most important questions to have answered because some datasets may be too old to be useful. If it exists, a metadata file, a file that describes a feature layer, is a good place to start. Metadata files are discussed later in this chapter.
2. The dataset may have locational inaccuracies. As described above, features may be sloppily positioned, omitted, or placed where they do not exist. This is usually the result of faulty fieldwork and observations, input error, or problems associated with converting data. On this last point, the accurate conversion of hard-copy maps to digital form is a major challenge because processing errors occur during digitizing and scanning (defined later in this chapter).
3. The attributes may be faulty due to keyboard error, faulty observation techniques, defective or non-calibrated instruments, or researcher bias.
4. If you are using secondary data, discover how the digital dataset was created and whether it underwent any digital conversion. Conversion problems occur when processing data from different formats, projections, data models, and resolutions. There can be inconsistencies (even slight) during the translation that alters geographic position and (in the case of raster resampling) the cell values.

Precision Errors:

1. The dataset may have positional imprecision due to the scale of the original maps. A map with a scale of 1:24,000 illustrates finer details than a smaller scale map of 1:500,000. With a vector-based GIS, as you zoom into your map, the feature locations may look precise, but its precision is based on the scale of the original map. It might not be as precise as it looks. Using a raster GIS, you can generate imprecision by starting with a low-resolution layer (large pixel size) and increasing its resolution. The pixels are capable of holding a higher resolution of data, but the precision will not improve.
2. Pixel resolution can also contribute to a dataset's positional imprecision. As discussed in Chapter 1, if a layer has a large pixel size, it may poorly represent the positional precision of its features. Small and narrow features will swim within large pixels. Their precise placement is unknown.

3. Positional imprecision can also occur if feature position is difficult to determine. Some features like streets and parcels are fairly easy to place, but there are features like soils, vegetation, and climate regimes that have fuzzy or less discrete borders. Some lines across a map are value judgments. The data may be accurate; they are just not precise.

Metadata

Given these possible errors, you can understand the danger and uncertainty of using undocumented data. Metadata is a data quality document, and its frequently repeated definition is “data about data” (although it is perhaps more accurate to define it as “information about data”). It describes the attributes and the location of the features in the layer. It gives you an impression about the dataset’s accuracy and precision. Metadata includes basic information about the dataset, including a description and if there are any use rules. Good metadata files should provide the answers to the following questions:

- What is the dataset’s age?
- What is the area covered by the dataset?
- Who created it?
- How was it constructed (digitizing, scanning, overlay, etc.)?
- What projection, coordinate system, and datum are used?
- What was the original map’s scale (if applicable)?
- How accurate and precise are the locations and attributes?
- What data model (vector or raster) does the layer use?
- How were the data checked (both location and attributes)?
- Why were the data compiled? What was their need or motivation?

After looking over the metadata file, you should ask yourself and your colleagues an additional question: is the data provider reliable? The presence and condition of the metadata might help answer this question.

Also within the metadata are data dictionaries that describe all of the features’ attributes. As you saw in Figure 2.4, attributes frequently have odd, short field names, like Zblack_00, that are difficult to decode. Metadata describes the attributes more fully. In addition, the values that go into these attribute fields are often coded with short values (or abbreviations) instead of using longer words, which are more subject to keyboard error. Data dictionaries decipher these codes and abbreviations. The metadata file in Figure 2.9 is an example. It documents a California weather station GIS layer from the California Spatial Information Library (CaSIL). It has been modified (shortened) for the purposes of this e-text.

Metadata is important. Years ago when GIS was in its infancy, there were no standards relating to data documentation and very little metadata existed. GIS layers were created with only a particular project’s specifications in mind, and many of the details went missing when the personnel that created the dataset moved on. As GIS datasets became numerous and agencies began sharing data, a common set of specifications arose to describe the GIS layers. A metadata file is generally attached or closely associated with a GIS layer.

LIBRARY: COUNTY, CA

LAYER NAME: WEATHERSTATIONS

COVERAGE NAME: WNSTNA

COVERAGE DESCRIPTION: The 'WEATHERSTATIONS' layer represents U.S. Weather Service surface weather observing stations. Many of the stations are located at airports. Others are located in or near USWS offices. This coverage was created from the file STATIONS.ZIP available from wmsd.cso.suic.edu (anonymous ftp).

VITAL STATISTICS:

Datum: NAD 27

Projection: Albers

Units: Meters

1st Std. Parallel: 34 00 00 (34.0 degrees N)

2nd Std. Parallel: 40 30 00 (40.5 degrees N)

Longitude of Origin: -120 00 00 (120.0 degrees W)

Latitude of Origin: 00 00 00 (0.0 degrees)

False Easting (Y shift): 0

False Northing (Y shift): -4,000,000

Source: wmsd.cso.suic.edu (University of Illinois)

Source Media: ftp file

Source Units: degrees/minutes

Source Scale: unknown

Capture Method: Converted to ARC/INFO using generate

Conversion Software: ARC/INFO Rev 5.1.1

Data Structure: Vector

ARC/INFO Cov. Type: Point

ARC/INFO Precision: Single

ARC/INFO Tolerances: default

Number of Features: 125

Layer Size: 22 MB

Data Updated: January 1994

DATA DICTIONARY:

DATAFILE NAME: WNSTNA.PAT, RECORD LENGTH: 61

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N DEC
1	AREA	4	12	F	3
5	PERIMETER	4	12	F	3
9	WNSTNA#	4	5	B	-
13	WNSTNA-ID	4	5	B	-
17	X-COORD	4	12	F	3
21	Y-COORD	4	12	F	3
25	NAME	16	16	C	-
49	ABB	3	3	C	-

AREA: The area of the polygon in square coverage units.

PERIMETER: The length of the polygon perimeter of the polygon in coverage units.

WNSTNA# : The software-assigned unique integer identification number.

WNSTNA-ID : A user-assigned identification number.

X-COORD : The Arc/INFO supplied x value for longitude in lat/long.

Y-COORD : The Arc/INFO supplied y value for latitude in lat/long.

NAME : The NWS-supplied station name.

ABB : The 3-character NWS station identifier.

DATA QUALITY ASSESSMENT:

The following are subjective comments regarding this data. Feature completeness is unknown. Feature accuracy is only to 1/100th of a degree of latitude or longitude. Attribute accuracy and completeness are also unknown.

DATA CONTACT:

Contact Name: ABCD, Contact's Phone: 916-###-####

DOCUMENTATION DATES: 10/27/####

Figure 2.9: A typical metadata file that provides information on both spatial and attribute data.

Despite its importance, metadata is still neglected. It is time consuming, but it is essential in an environment where data sets are shared and the background of the data must be known.

On-Line GIS Data

The Internet is a great place to start looking for data. If you find existing GIS datasets that serve your purpose and passes your specifications, it saves you time and money. A search may reveal multiple copies of what seems to be the same data, but check the details—examine the metadata—because minor differences might make one dataset better than the other. Much base map data (countries, states, counties, major roads, rivers, township and range) exists on the Internet.

It would be convenient to retrieve all of your GIS datasets from the Internet, and although more and more data are available, the Internet will not provide you with everything you need. When searching the Internet, don't just look for GIS files. Search for data that might be GIS compatible. Perhaps a spreadsheet can be modified and linked to an existing geographic layer. CAD files, aerial photographs, and satellite imagery might also be relevant. Finding good digital datasets frees you from collecting and entering the data yourself.

Other Sources of GIS Data

Most likely, you will contact GIS personnel at various government agencies, commercial data sources, and organizations about their data sources. Ask questions about the datasets accuracy and completeness. If you think their datasets might be helpful to you, ask them for permission to obtain and use the data. Frequently, a short agreement will be written and signed that allows one to use the data under certain conditions.

Many data companies modify "public" data to create a "value-added" product that you can purchase and load directly into your GIS. "Value-added" datasets usually originate from a government agency or an organization that creates the basic GIS dataset, but a commercial company obtains the data and "improves" it by adding attributes or improving its spatial precision. The commercial company can then sell the "value-added" portion of the data. Many of these datasets can also be obtained over the Internet.

Converting from one GIS format to another.

When obtaining GIS data from the Internet or from other sources, it requires extensive preprocessing to make it work with your other GIS datasets. Initially, the newly acquired dataset may require “extracting” (unzipping). They are often compressed to make them smaller for storage on a CD or to make their downloading quicker. If the acquired file’s type is ZIP, TAR or GZ, it needs to be extracted, which can be done with free or cheap software (like WinZip). At times, extracting files can be more complicated if the compressed files are nested within a compressed file. This requires extracting the first compressed file and then extracting the nested files.

GIS datasets are typically stored in one of the leading GIS software formats (frequently as a shapefile) or in a format specified by the U.S. federal government. Fortunately, most GIS packages read or convert many of the most common GIS formats. At times, however, you may need access to “third party” software to read the data and export it into a format that your GIS program reads.

Even after the datasets are extracted and converted into your GIS format, they will require further processing. Many types of data preprocessing, manipulation, and conversion are discussed in Chapters 3 and 4.

This page titled [2.3: Phase 2- Data Acquisition](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

2.4: Phase 3- Data Capture

Introduction

When you have exhausted your contacts and the Internet, it is time to capture the data yourself. In this phase, you create new GIS datasets from both digital data that are not currently in a GIS format and from non-digital, hard-copy data sources. Examples of digital and non-digital data sources include maps (hard-copy and digital), aerial photographs (hard-copy and digital), questionnaires, field observations, digital satellite imagery, survey data, and Global Positioning System (GPS) coordinates.

The data capturing phase is often tedious, laborious, and frustrating, but necessary. Figure 2.10 breaks the stage into three main steps, which are described in the text below:



Figure 2.10: The key steps in data capture phase. Here you digitize hard-copy maps and data directly into your GIS or transform existing digital data into a format your GIS reads.

Converting Digital Data

This section looks at digital datasets that are currently not in a GIS format, but that are often manipulated to create GIS layers. These sources include automated surveying, photogrammetry, GPS, Light Detection and Ranging (LIDAR), non-GIS mapping programs like Computer Aided Design (CAD), and remote sensing imagery. This chapter emphasizes the conversion of GPS data and remote sensing imagery. Importing spreadsheet data, a frequent way of importing attribute data, into a GIS is covered in Chapter 4.

Automated surveying uses electronic data capturing instruments like theodolites, electronic distance measurement (EDM) systems, and total stations to capture spatial and attribute data. The most sophisticated of these instruments is the total station that combines the theodolite's angle-measuring capabilities with the EDM's distance calculations. Surveyors download the distance and direction data from their instruments directly into many vector-based GIS programs. The data, however, usually requires pre-processing before it can be used to make a map.



Figure 2.11: Total station, a surveying instrument, measures distance, direction, and location. Image comes from Illinois State University, Geography Dept.

Photogrammetry obtains accurate measurements from aerial photographs. Photogrammetric techniques determine ground distances and directions, heights of features, and terrain elevations. Photogrammetry creates GIS data through 3-D stereo digitizing and by producing spatially rectified aerial photographs that can be entered into the GIS as a layer.

GPS (Global Positioning System) is a radio-based navigation system that uses GPS receivers to compute accurate locations on the Earth's surface from a series of orbiting satellites. With a small, inexpensive, hand-held GPS receiver (Figure 2.12), you can determine your location usually within about three meters. Using two GPS receivers simultaneously (called Differential GPS or DGPS) or using a Wide Area Augmentation System (WAAS)-enabled GPS receiver, which uses satellites and ground stations that provide GPS signal corrections, you can get sub-meter accuracy.



Figure 2.12: Typical low-end GPS device.

The U.S. GPS network, called NAVSTAR (Navigation with Satellite Timing and Ranging), has at least 24 satellites that orbit in six planes around the Earth (see Figure 2.13). The network's configuration secures at least four satellites—the minimum number of satellites needed to capture location data—above the horizon for every point on Earth.

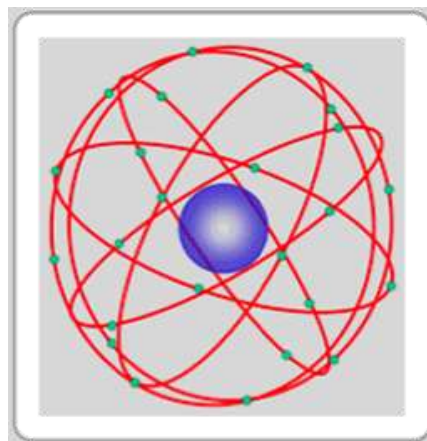


Figure 2.13: NAVSTAR GPS constellation of satellites.

The U.S. Department of Defense (DoD) developed and controls NAVSTAR, and DoD can turn the whole system off, as they briefly did immediately following the terrorist attacks of September 11, 2001. DoD monitors and tracks the satellites (which are equipped with radio transmitter/receivers and a set of atomic clocks) from five stations across the globe where they compute precise satellite orbital and clock corrections. These corrections are transmitted from the master control station at Schriever Air Force Base in

Colorado to the satellites, which make the adjustments. All of the satellite's locations are precisely known, and by knowing their exact locations, we can determine the location of every point on Earth with a GPS receiver.

This is possible because each satellite transmits a unique radio signal, which can be received by GPS receivers. Using this signal, your GPS receiver calculates the distance to each of the four satellites it is tracking by the amount of time it takes for the signals to travel from the satellites to your receiver. This is a high-tech version of triangulation (see Figure 2.14), called trilateration. The first satellite locates you somewhere on a sphere (top left of Figure 2.14). The second satellite narrows your location to a circle created by the intersection of the two satellite spheres (top right). The third satellite reduces the choice to two possible points (bottom left). Finally, the fourth satellite helps calculate a timing and location correction and selects one of the remaining two points as your position (bottom right).

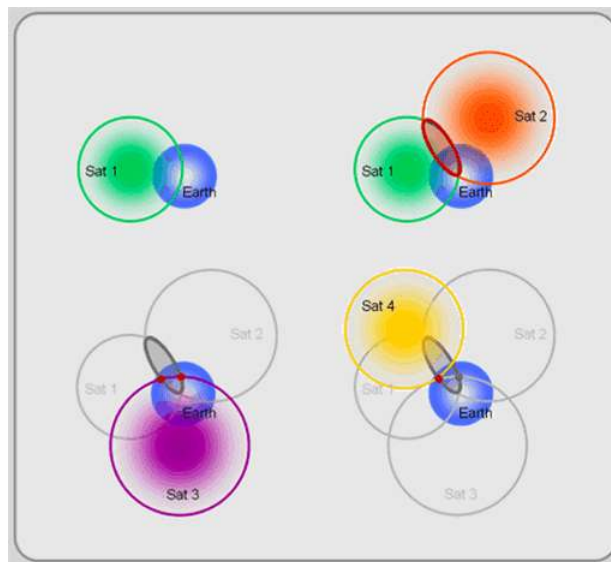


Figure 2.14: GPS satellite trilateration.

When dealing with satellites thousands of miles from the Earth's surface, GPS's accuracy is amazing. Still, the technique is not without error. Small errors in the receiver's clock, variations in the satellite's orbit, atmospheric conditions that slow radio waves, and radio signals that bounce (called "multipath" or "ghosting") off buildings and cliffs are some possible distortions. In addition, GPS has difficulty penetrating thick forests and urban canyons created by tall buildings.

Another source of error is Geometric Dilution of Precision (GDOP). This is the spatial relationship between the GPS receiver and each of the potential satellites. In general, the fewer the satellites available, and the closer they are clustered, the less accurate your readings. GPS receivers try to avoid GDOP by selecting the set of satellites that provide the least error. It chooses satellites that are well above the horizon, minimizing atmospheric thickness and interference from buildings, yet not so high that they are clustered together.

GPS is a major data input tool for GIS. Most receivers, even inexpensive units, contain a hard-drive where you can log your positions. Each logged position is called a waypoint, and together, waypoints depict the location and extent of the features you enter in the field. They can be downloaded from GPS receivers (sometimes with the help of a separate software program) and imported into many GIS programs.

Some of the more expensive GPS units have "feature lists" that streamline the data entry process. Feature lists are databases that you define to contain a list of the possible features you will locate. These lists have associated attributes for each feature type, and drop down lists of common attribute values can also be predetermined for each attribute to save time in the field. Both the feature locations and their associated attributes can be downloaded into your GIS.

LIDAR (Light Detection and Ranging) is a remote sensing technology that uses laser light pulses to measure the distance to a surface. It is similar to other types of radar, but uses light instead of radio waves. Airborne LIDAR systems have resulted in topographic layers that depict the tops of ground-based features better than traditional remote sensing and radar methods, and this

results in topographic layers that portray the shape of our cities (including the widths and heights of buildings) and forest canopies more accurately. Figure 2.15 is an example of LIDAR image in an urban area.



Figure 2.15: LIDAR image depicting feature heights.

Both vector and raster data models use LIDAR data. Many GIS programs have automated routines that convert the x, y, and z (elevation or altitude) points first into vector point files, which can be converted into raster layers (rasterization is described in Chapter 3) where additional processing and analysis can take place. In vector systems, LIDAR points frequently generate topographic layers called Triangular Irregular Networks (TIN), which can be used to represent terrain in vector-based systems.

Conversion of Spatial Data from Non-GIS Programs is a common way to acquire data. Other than GIS, there are additional computer programs that display, create, and edit vector spatial data. A brief list includes Computer Aided Design (CAD) systems (like AutoCAD, MicroStation, and ArchiCAD) and vector drawing programs (like Adobe Illustrator, Corel Draw, and OpenOfficeDraw). The conversion process varies based on the format of the dataset and the import/export capabilities of both the host software and the GIS program. Capturing and importing CAD and drawing data into vector-based GIS often involves converting the digital data from the host format into the desired GIS format or into a format that can be easily read by your GIS. Sometimes the conversion process is that easy. Sometimes it's not; the data may need to be entered into a "third party" format that both programs read or export to. AutoCAD files are an example. They usually need to be saved in DXF format within the CAD software before many GIS programs can read them.

Remote Sensing imagery is an important source of raster data. This section briefly describes satellite remote sensing, digital image processing, and the conversion of remotely-sensed imagery into GIS.

The word "remote" means "from a distance." Sensing, in this case, means "to record." So a basic definition of remote sensing is to record from a distance.

To understand remote sensing, we need to understand some important core concepts. First, all features on the Earth's surface reflect and absorb the sun's radiant energy. They also emit and reflect back parts of that energy. The amount and type of radiation that is emitted and reflected from the Earth's features depends on the properties of those features (see Figure 2.16).

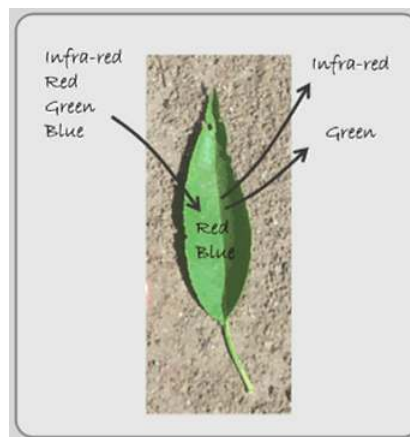


Figure 2.16: Bands of light including infra-red, red, green, and blue are emitted from the sun. The green leaf absorbs the red and blue bands, but largely reflects the infra-red and green bands thus appearing to our eyes green.

What our eyes see is the reflected radiation within a very small part of the electromagnetic spectrum (see Figure 2.17). We see the blue, green, and red bands of light, known as the visible spectrum, but other wavelengths are present that we do not sense. Satellites, however, have sensors that record not only the visible spectrum, but also infrared, near infrared, and thermal infrared bands. Satellite sensors, however, vary by maker, and so they record different wavelengths. The precise bands (wavelengths) a satellite captures is referred to as spectral resolution.

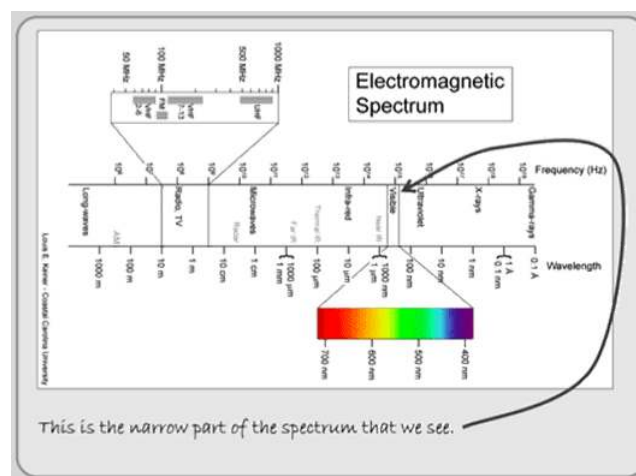


Figure 2.17: The electromagnetic spectrum with the visible waves highlighted. Primary image is by Louis Keiner, Coastal Carolina University. See licensing at <http://commons.wikimedia.org/wiki/File:Electromagnetic-Spectrum.png>.

Through telemetry, satellites transmit each of these individual wavelengths into remotely sensed data files, called images, which measure the reflected and emitted electromagnetic energy from the Earth's features. The images are raw meaning they are unprocessed, and they need to be digitally enhanced and combined to highlight particular features (land use, climate changes, agricultural productivity, and environmental properties).

Like any raster image, remote sensing images consist of pixels. Each pixel in a data file records a particular wavelength (band of light) of the electromagnetic spectrum for a specific chunk of the Earth's surface. The pixel size is its spatial resolution, and it differs from spectral resolution in that it involves pixel size rather than a particular wavelength or band of light recorded by the satellite. Like spectral resolution, different satellites acquire data at different spatial resolutions.

Different types of features emit and reflect back different intensities of each wavelength of light, so remote sensing images are combined and manipulated to highlight specific features. This procedure of combining images to isolate features is called digital image processing. Picking the proper images comes with experience and knowledge, but it is based on the feature's spectral

signature, which depicts the percent of each band of light that is emitted and reflected back off the features. Each spectral signature is kind of a feature's unique fingerprint.

Digital image processing is done in one of many specialized remote sensing software packages like Erdas Imagine or ER Mapper (but some raster-based GIS programs, like Idrisi, are capable of many image-processing functions). Once processed, the images can be entered into your GIS. Some GIS programs read some of the major remote sensing image formats, but you may need to export the image to a format that your GIS program reads. The processed images are usually used in raster-based GIS as a continuous data surface, such as climate data. The images can be used in vector-based systems simply as a reference background image or to trace vector features over (a process called head's up digitizing, see next section).

Converting Non-Digital Data

Existing, hard-copy maps and aerial photographs (physical paper documents) are a major source of spatial data for GIS. Different processes, including digitizing, scanning, and "heads up" digitizing, exist to input these hard-copy sources into GIS. For most projects, you will need to capture both the spatial location of the feature and some of its attributes.

Data input is usually the major bottleneck in the development of a GIS database, and converting hard-copy data from maps, aerial photographs, printed reports, and field notebooks is often the least desirable option because it is tedious and time consuming. Still, it is a way of making sure that you get a certain level of accuracy and precision for your project.

Scanning is a popular way to convert hard-copy maps and aerial photographs into digital images. The resultant scanned image is a raster file, arranged as an array of pixels in columns and rows. Scanners capture what is on the original document by assigning a color or grayscale value to each pixel in the array. The pixel value is based on the intensity of the color or gray shade on the original document.

Scanners come in a range of types, sizes, and degrees of sophistication. Scanner types include flat bed, sheet fed, drum, and video. Flat bed (or desktop) scanners are the most common and consist of a glass board where you lay the documents you want to scan. It works like a copier, but the output saves to an image. Usually, they scan areas no larger than a legal-sized document but a couple brands make flat beds up to 24" x 36".

For hard-copy documents larger than 24" x 36", you can choose a sheet fed (see Figure 2.18), drum or video scanner. Sheet fed scanners work like a fax machine. The document moves past a scanning head, and the image is captured. It scans only loose sheets and the image is sometimes distorted if the document is not handled properly. A few varieties can handle rigid maps on poster board. Drum scanners are a great alternative due to their size and accuracy, but they are expensive, which generally limits their use to large commercial firms and government agencies. You feed maps into them and its rotating drum systematically scans the data. Video scanners use a high-resolution camera system to scan—incrementally—portions of the map. They have become an alternative to drum scanners because they are cheaper and their accuracy is improving (still, there are geometrical distortions and uneven brightness values across their scanned images). They are, however, extremely fast; most scans from a video scanner take about a second.



Figure 2.18: Wide-format, sheet-fed scanner.

For many GIS applications, you may also need to consider the scanner's resolution and its bit depth. Resolution is the detail the scanners read. Scanners support resolutions from 72 to 1200 dpi (dots per inch), but most scanning for GIS applications is done at resolutions between 200 and 600 dpi. Higher resolutions create higher quality images, but they also create substantially larger file sizes. Also, consider that at the same resolution, color images are larger than grayscale images, which in turn are larger than black and white images.

Bit depth is the number of bits used to represent each pixel. The greater the scanner's bit depth, the more colors and shades of gray that can be represented in your digital image. For example, a 24-bit color scanner can represent 2 to the 24th power or 16.7 million colors. Greater bit depth, however, creates larger files.

Before scanning any document, remember to clean it—smooth out folds and tape any tears. This reduces spatial inaccuracies. Also, erase any marks you do not want to appear on the resultant image. Similarly, make wanted notes and highlighted features legibly on the hard-copy document if you want to see them on the scanned image.

Once scanned, remember that you have essentially a “dumb” image; its pixels do not have meaningful values (just a gray or color value taken from the scanned document). This limits its use because it is just a pretty picture. Although future scanning technologies focus on generating vector features directly from scanned images (called auto-vectorization), these technologies are not ready. Scanning, however, remains beneficial because of a second input process called “heads up” digitizing.

Heads Up Digitizing – After you create a scanned image, you georeference it (see Chapter 3) and use it as a background image within your vector system. Then with the image at its proper geographic location, trace the features that appear on the scanned image. This process, called “heads-up” digitizing (or on-screen digitizing), is like manual digitizing (described below) but without a physical digitizing board. Instead, you see on the screen a scanned image in its correct geographic position, and, with your mouse, you trace the position of features into new or existing point, line, and polygon layers. In the example below, building footprint polygons are traced over an image to create a vector polygon layer of buildings.



Figure 2.19: Head's up digitizing. By eye, trace the features on your georeferenced image.

After you trace each feature, enter the feature's key identifier into the layer's attribute table (see Figure 2.20). How you enter key identifiers depends on whether the feature's layer is brand new or whether you are updating an existing feature layer. If you are updating an existing layer, the field will already be present, and you simply need to fill in the feature's value in the appropriate field. If you are creating a new layer, you must define at least the key identifier's attribute field before digitizing. This is important because each feature needs to be identified uniquely. The rest of the attributes are usually created in a separate file (perhaps in Access or Excel) and “joined” to the feature's minimal attribute table after digitizing is complete. The joining process is covered in the second half of Chapter 4.

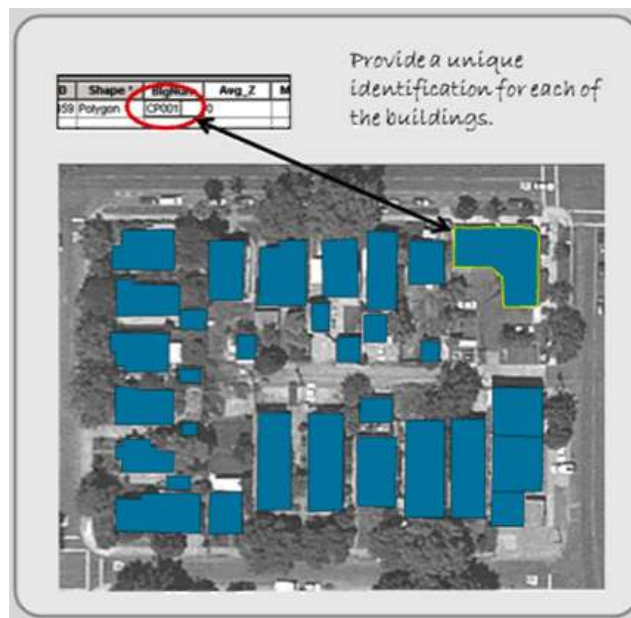


Figure 2.20: Entering key (unique) identifiers into the new layer's attribute table.

The procedure for digitizing features is similar to that used in “manual” digitizing (see below). Due to the increased popularity of heads-up digitizing, manual digitizing is less frequently used, but for certain applications it is an important way to input map data.

Digitizing involves tracing by hand the extent of features directly from a hard-copy map or photograph that is mounted onto a digitizer (see Figure 2.21), a large table or board with an imbedded electronic grid that senses the position of a pointer called a puck (a mouse like device).



Figure 2.21: Digitizer with puck.

All GIS packages have a specific procedure for manual digitizing. Generally, it involves three steps: mounting the map on the digitizer, establishing control points, and adding map features.

In the first step, you mount the map in the middle of the digitizer's surface with its corners held by masking tape. Stretch out any map creases, but try to obtain maps that are in good shape because spatial errors may result. If you are working with the same map over a series of days, you need to remount the map periodically because humidity can expand paper maps, and they can sag slightly.

Step two involves establishing your map's control points. Although software packages vary, establishing control points involves a routine where you identify at least four points that are common to both the mounted map on the digitizer and the map on your screen. To minimize error, pick control points that surround—and one that is within—the area to be digitized. Then point to the four control points on the digitizer's map with the puck, and in the same order, identify the same four locations on the screen's map with the mouse. This results in a spatial connection between the map on the digitizer and the screen's map.

Once this connection is established, you can trace the map features on the digitizer (Step 3). Take the puck and position it directly over one of the feature's vertices (corners). Various buttons on the puck allow you to create vertices, delete vertices, and close polygons. Like "heads up" digitizing, you will want to enter at least the feature's key identifier as you digitize

GPS digitizing involves using a GPS receiver to record feature data in the field. Using GPS for point locations (waypoints) was described above, but mapping-grade GPS units, like Trimble's GeoXT, are capable of recording the nodes and locations of points, lines and polygons by following the feature's extent and registering a waypoint at each of its vertices.

Pilot Project

A pilot project is a rehearsal. Here you collect a small subset of the GIS datasets you require for the larger project. Then you input the data into the GIS, preprocess the datasets, analyze them, and create some output. When something goes wrong, you tweak the project's parameters until the process works smoothly.

Pilot projects give you the opportunity to "ground truth" your secondary data. Remember, it is foolish to believe these datasets are without flaws. You need to ground truth your GIS data to ensure that the datasets are representative of what's on the ground. It is done by traveling to your study area and using your eyes to verify your datasets. Normally, you do not need to generate statistics nor check every feature. Checking an informal sample is good enough. Likewise, if any of your datasets are old or some features have oddly shaped features, this is a good opportunity to check them too. If ground truthing is difficult due to access or the aggregation of attributes (as in census data), cross validate your GIS data with other sources like aerial photography, satellite imagery, assessor data, or demographic figures.

This page titled [2.4: Phase 3- Data Capture](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

3: The Earth and Map Preprocessing

[3.1: Introduction](#)

[3.2: Earth](#)

[3.3: Map Preprocessing](#)

[3.4: Map Verification and Editing](#)

This page titled [3: The Earth and Map Preprocessing](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.1: Introduction

This chapter and the next focus on key concepts and the preprocessing of your data. Preprocessing is an important but unheralded subject. It shapes the datasets you place in the GIS; it prepares them for analysis. This chapter looks at preprocessing the map or spatial component of your features. The next chapter focuses on preprocessing your feature's attributes. These are the "housecleaning" tasks mentioned in the first two chapters.

This chapter begins with concepts that define the geographical referencing standards of the Earth. Topics include latitude and longitude, projections, coordinate systems, and datums. These concepts help you understand map preprocesses like changing projections, converting layers from vector to raster, and reclassifying or resampling layers. A large part of map preprocessing is to make your data usable by providing consistent projection parameters throughout all your data sets. The goal is to make your layers fit properly over each other.

This page titled [3.1: Introduction](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.2: Earth

Latitude and Longitude

Any feature can be referenced by its latitude and longitude, which are angles measured in degrees from the Earth's center to a point on the Earth's surface (see Figure 3.1). Across the spherical Earth, latitude lines stretch horizontally from east to west (left image in Figure 3.2), and they are parallel to each other, hence their alternative name, parallels. Longitude lines, also called meridians, stand vertically and stretch from the North Pole to the South Pole (center image in Figure 3.2). Together these “north to south” and “east to west” lines meet at perpendicular angles to form a graticule, a grid that encompasses the Earth (right image in Figure 3.2).

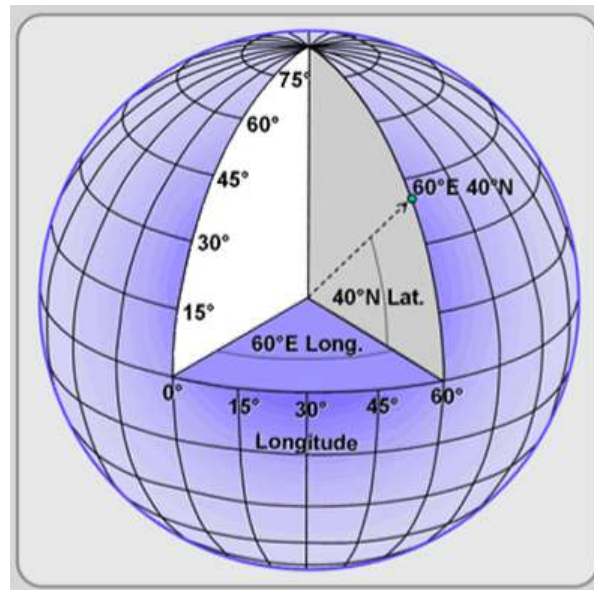


Figure 3.1: Latitude and longitude are angles measured in degrees from the Earth's center to a point on the Earth's surface.

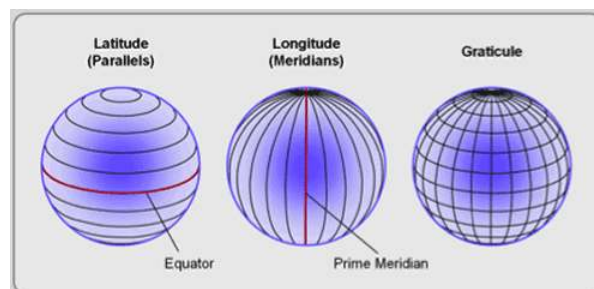


Figure 3.2: Latitude, longitude, and the Earth's graticule.

Midway between the poles, the equator stretches around the Earth, and it defines the line of zero degrees latitude (left image in Figure 3.2). Relative to the equator, latitude is measured from 90 degrees at the North Pole to -90 degrees at the South Pole. The Prime Meridian is the line of zero degrees longitude (center image in Figure 3.2), and in most coordinate systems, it passes through Greenwich, England. Longitude runs from -180 degrees west of the Prime Meridian to 180 degrees east of the same meridian. Because the globe is 360 degrees in circumference, -180 and 180 degrees is the same location.

Earth's Shape

If the geographic extent of your project area was small, like a neighborhood or a portion of a city, you could assume that the Earth is flat and use no projection. This is referred to as a planar surface or even a planar “projection,” but with the understanding that it does not use a projection. Planar representation does not significantly affect a map's accuracy when scales are larger than 1:10,000. In other words, small areas do not need a projection because the statistical differences between locations on a flat plane and a 3-dimensional surface are not significant.

For small-scale maps (those that encompass a large area, see Figure 2.3), you must consider the Earth's shape. Our assumption that the Earth is round or spherical does not accurately represent it. The Earth's constant spinning causes it to bulge slightly along the equator, ruining its perfect spherical shape. The slightly oval nature of the Earth's geometric surface makes the terms ellipsoid and spheroid more accurate in describing its shape, but they are not perfect terms either since differences in material weights (for instance iron is denser than sedimentary deposits) and the movement of tectonic plates makes the Earth dynamic and constantly changing. The Earth is a geoid with a slight pear shape; it is a little larger in the southern hemisphere and includes other bulges. The difference, however, between the ellipsoid and the geoid is minor enough that it does not affect most mapping. Until recently, projections based on geoids were rare because of the complexity and cost of collecting the necessary data to create the projection, but satellite imagery has helped with measurement and geoid projections are now more common.

Map Projections

Globes do not need projections, and even though they are the best way to depict the Earth's shape and to understand latitude and longitude, they are not practical for most applications that require maps. We need flat maps. This requires a reshaping of the Earth's 3-dimensions into a 2-dimensional surface. This reshaping cannot be done without introducing some error. To illustrate this point, imagine taking a cardboard globe, cutting it in half at the equator, and then cutting both the northern and southern hemispheres into four equal parts apiece. Resting on a table, the pieces are not flat; they arch in the center. Try flattening one of the pieces. If you succeed, part of the cardboard will be scrunched together and other parts will tear apart. By flattening it, you modify its geography.

Map projections enable the reshaping of the Earth by mathematically transforming spherical coordinates (x , y , and z) to 2-dimensional (x and y) space. They are the foundations we use to represent the Earth's surface or portions of it.

Projections are abstractions, and they introduce distortions to either the Earth's shape, area, distance, or direction (and sometimes to all of these properties). Different map projections cause different map distortions.

One way to classify map projections is to describe them by the characteristic they do not distort. Usually only one property is preserved in a projection. This chapter confines its focus to just two properties—area and shape—because the projections that preserve these properties—equal-area and conformal—are the most common.

Equal area (or equivalent) projections preserve the area (or the amount of space) within features. On a small-scale political map of the world, the areas within each country are preserved. In reality, the area of Mexico and Greenland is similar, and in the right-hand map in Figure 3.3, which is drawn in an equal area projection called Mollweide, the two territories are approximately the same size. Equal area projections, however, distort all the other properties. Shape, distance, and direction are not preserved.

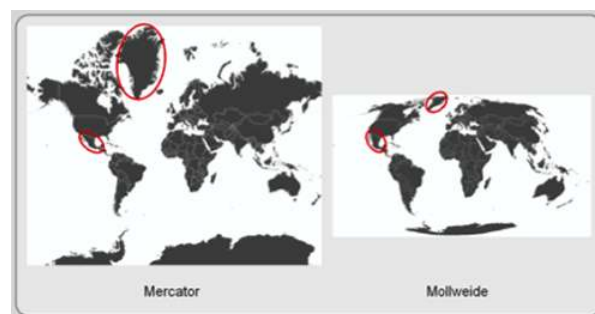


Figure 3.3: Projections transform the 3-D Earth into a 2-D plane. Some projections, like Mercator, attempt to preserve area while other projections like Mollweide preserve the area contained within the landmasses.

Conformal (also known as orthomorphic) maps preserve shape by preserving the angles of feature boundaries like countries and continents. Maintaining angles, however, distorts the area within the features (see left map in Figure 3.3). In the Mercator projection, Greenland looks like Greenland, but it is far larger than Mexico, its spatial equivalent. In addition, no conformal projection preserves the shapes of features that extend close to the poles (notice Antarctica).

Another way to classify map projections is by their projection surface. Imagine a translucent globe with country boundaries and latitude and longitude lines drawn upon it in black. In addition, imagine a light bulb positioned at the globe's center. If you placed large pieces of paper on or around the translucent globe and turned on the interior light, you would see the country boundaries and

the latitude and longitude lines projected onto the paper. If those projected lines were imprinted onto the paper, the paper could be removed from the globe, cut, and flattened to produce a 2-dimensional map.

What you just imagined is the way many projections were first conceived. Even today, in the age of computer modeling, most of the map projections we use are variations on three basic projection surfaces: planar, conic, and cylindrical (see Figure 3.4).

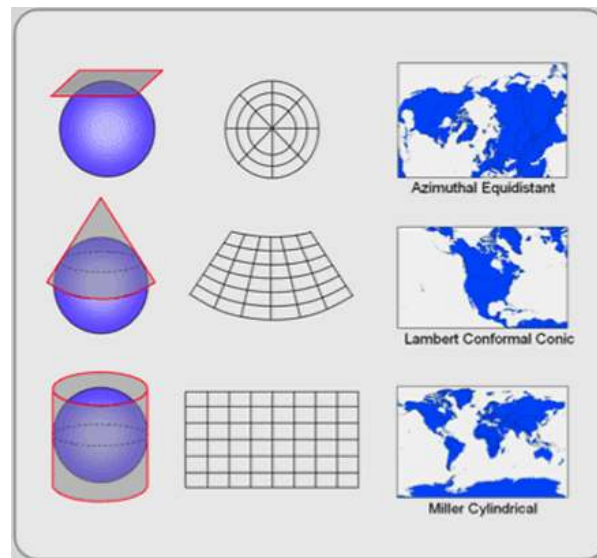


Figure 3.4: Projection surfaces.

Planar projections, the least common, can be conceptualized by placing a flat sheet in contact (at one point) with the translucent globe, usually at the North or South Pole, and the lines on the globe are projected onto the sheet. The projected map creates a circular graticule (see top row of Figure 3.4). Direction, one of the properties not described, is usually preserved from the center of the map outward. Some planar projections preserve area or distance. Consider using a planar projection if your research area is at one of the poles.

1. *Conic* projections, a common projection surface, are conceptualized by placing a paper cone on the globe, and the lines on the globe are projected to the cone. After unraveling the cone, the graticule appears fan shaped (middle row of Figure 3.4). Conic projections preserve different properties including area and shape, but never both in a single projection. Distortion across the map varies. No distortion exists along the parallel (latitude) where the cone touches the globe, but distortion increases in both directions away from this line of tangency. Consider using a projection with this surface if your study area is in the mid-latitudes including the U.S.
2. *Cylindrical* projections are developed by wrapping paper around the globe in the shape of a cylinder. The lines on the globe are projected to the cylinder (bottom row of Figure 3.4), and the resultant graticule is rectangular. There is no distortion along the equator (its point of tangency), but distortion increases toward the Earth's poles. This projection surface preserves different properties including area and shape (but again, both are not preserved in a single projection). Consider using this projection surface if your study area is worldwide or in the tropics.

Many variations can be made using these three projection surfaces. Instead of having the paper come to a simple point or line of tangency with the globe, you could cut the globe's surface (called secant), so that conic and cylindrical projections intersect the globe at two lines (latitude) and plane projections create a single circle. No distortion occurs anywhere the projection surface (the paper) intersects with the globe. Where the projection surface is outside the globe, features appear larger than they are in reality. Where the projection surface is inside the globe, features appear smaller.

Additional variations result when you move the position of the globe's interior light or combine multiple projection surfaces. In addition, with computers, mathematical projections not based on these projection surfaces exist, and some of these projections are very popular.

There are thousands of different projections, but only a few dozen projections are noteworthy and used. Examples include Albers Equal Area Conic, Lambert Conformal Conic, Mercator, Miller Cylindrical, and Robinson. Many of these projection names include

words like equal area, conformal, conic, and cylindrical; they provide clues to the projection's characteristics and projection surfaces.

As mentioned in Chapter 2, it is important to choose an appropriate projection for your GIS project to achieve accurate results. Are you interested in calculating the area of features? If so, you must use a projection that preserves area, or your calculations will be inaccurate. Improper projections distort attribute accuracy, positional accuracy, and thus the information in your final maps and reports. As Chapter 6 describes, choosing an unsuitable map projection is one way to lie with maps.

How accurate do your locations need to be? If you are making a world map with the locations of the largest ports, precise locations are probably not necessary. If you are drilling a train tunnel, however, positional accuracy is required. Some projections use spheres to model the Earth's shape. Remember (from the first part of this chapter) that a sphere is the most generalized shape of the world and the least accurate. Many projections are based on spheres, and these projections are suitable for world maps and large world regions that do not require a high degree of positional accuracy. Most projections today, however, are based on ellipsoids (and spheroids) that distort the uniformity of the sphere to bulge a bit at the equator. Statistically speaking, there is no significant difference between most ellipsoids and the true shape of the Earth for most mapping purposes. Still, for those projects that require even more precision, there are projections based on geoids. These projections were rare until recently because of the time it took to calculate the projections and the difficulty of the math and measurement. With increasing satellite imagery, however, they are more common.

Coordinate Systems

Projections and coordinate systems are two separate things. As described above, projections convert the Earth from 3-dimensional space to a 2-dimensional map. Coordinate systems are referencing systems used to describe specific locations and measure distances on maps. They provide x, y locations (sometimes designated as Eastings and Northings) for features, and, within GIS, they are used to spatially register layers of features that occupy the same area.

While coordinate systems are not projections, they usually use them. Latitude and longitude, the best known coordinate system, however, does not use a projection, but in most cases, coordinate systems incorporate a map projection, reference spheroid, datum, one or more standard parallels, a central meridian, and possible shifts in x or y (easting and northing) directions.

Like projections, there are many coordinate systems. Some of these coordinate systems focus locally and some globally. The most common at the world level is latitude and longitude, but because it is not a "projected" coordinate system, plotted points usually have a high degree of distance and shape distortion when plotted on a 2-dimensional flat map (and so it should not be used for making two-dimensional maps). Latitude and longitude uses the Prime Meridian and the Equator as reference planes, and it is best used when conceptualizing the Earth as a globe.

There are two ways of providing latitude and longitude coordinates. One method uses degrees, minutes and seconds. For instance, the CSUS Geography department is located at 38°N 33' 32" latitude and 121°W 25' 31" longitude. Another method is decimal degrees, and the same coordinates are represented as 38.55889 latitude and -121.42527 longitude.

The following is a description of a few of the most used coordinate systems in the United States: Universal Transverse Mercator, State Plane Coordinate System, and United States National Grid. Latitude and longitude, also widely used, was described earlier in this chapter.

Developed in the 1940s by the U.S. Army Corps of Engineers, Universal Transverse Mercator (UTM) is a coordinate system that largely covers the globe. The system reaches from 84 degrees north to 84 degrees south latitude, and it divides the Earth into 60 north-south oriented zones that are 6 degrees of longitude wide (see Figure 3.5). Each individual zone uses a defined transverse Mercator projection. The contiguous U.S. consists of 10 zones. In the Northern hemisphere, the equator is the zero baseline for Northings (Southern hemisphere uses a 10,000 km false Northing). Each zone has an arbitrary central meridian of 500 km west of each zone's central meridian (called a false Easting) to insure positive Easting values and a central bisecting meridian. In UTM, the CSUS Geography Department is located at 4,269,000 meters north; 637,200 meters east; zone 10, northern hemisphere.

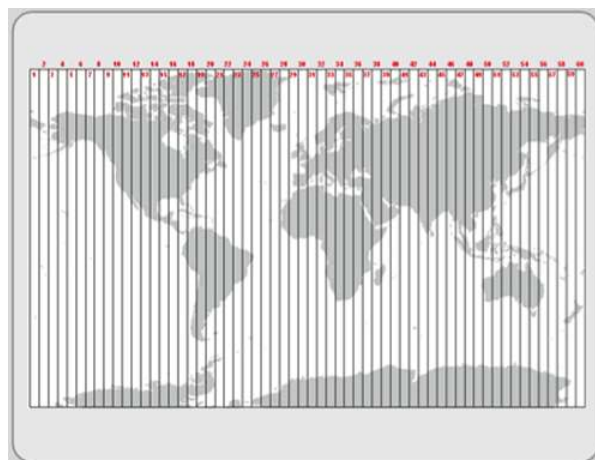


Figure 3.5: Universal Transverse Mercator zones.

The State Plane Coordinate System (SPCS) is a projected coordinate system that divides the U.S. and its possessions into over 120 zones (see Figure 3.6). Some smaller states use a single zone while larger states are divided into several zones. California has six zones (see map in lower-left corner of Figure 3.6). Each zone provides a local reference system that has its own parameters. Zones oriented east to west use the Lambert Conformal Conic projection while zones stretching more north to south use Transverse Mercator (not to be confused with UTM). Used principally by cities, many counties, and some states, they are popular projected coordinate systems. In SPCS, the CSUS Geography Department is located at 599200.796 feet; 2050091.975 feet; CA zone 2.



Figure 3.6: State Plane Coordinate System.

Many GIS projects cover more than one SPCS or UTM zone. In response to calls for a single coordinate system that covers the entire U.S., the United States National Grid (USNG), was created in 2001. After the attacks of September 11, 2001 and the increasing use of GPS-enabled and location tracking devices, a consistent grid takes on additional importance, and in 2005, the Department of Homeland Security (DHS) recommended that any DHS grant should reference their data to USNG. USNG incorporates the U.S. Military Grid Reference System's hierarchy (not described here), but the basic zones are identical to UTM.

Datums

A datum is a starting point for locating features on the Earth's surface; it is the origin point of a coordinate system. It defines the position of the ellipsoid (or spheroid) relative to the Earth's center. There are many different datums and hence many different starting positions. Like both projections and coordinate systems, international organizations and individual nations have established datums for their specific needs. The World Geodetic System 1984 (WGS84) is the most widely used datum internationally. In the U.S., the two most used datums are North American Datum 1927 (NAD27) and North American Datum 1983 (NAD83). NAD83 updates NAD27 by using a more accurate ellipsoid for North America, derived from better satellite imagery, and changing reference units from feet to meters.

This page titled [3.2: Earth](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.3: Map Preprocessing

Map preprocessing functions are housecleaning tasks that make the data you input into the GIS usable for data analysis. The objective is to get all of your GIS datasets into the same projection, and then to make each layer spatially in tune with each other. Many map-preprocessing tasks do just and include reprojection, georeferencing, resampling, reclassification, and edge matching. In addition, verifying, editing, and manipulating your map features are part of this chapter as well.

Reprojection: Changing Projections, Coordinate Systems, and Datums

All of your project's feature layers must be in the same projection and coordinate system if you intend to use them for analysis or map production. Both raster and vector GIS programs allow you to convert layers of features from one projection, coordinate system, and datum to another. In vector systems, it involves translating the x and y coordinates of all the features to new coordinates. In raster systems, it involves coordinate translation and resampling the pixels of one image into a new image (in some raster-based systems, changing projections is called resampling). For both vector and raster systems, the processes are not error free and datasets that are repeatedly translated back and forth compound errors.

All GIS programs have projection utilities that allow you to change your layer's projection, coordinate system, and datum. When reprojecting data, you need to know both the existing and the output projection parameters (parameters include projection, coordinate system, and datum). Existing projection information is found in the layer's metadata if it exists. If metadata does not exist, you need to speak with someone who created or at least uses the data set. As for the output projection parameters, presumably, you know these (this should be determined in the planning phase). When reprojecting, many programs give you the option to import your projection parameters by selecting an existing GIS layer that already uses them. If you choose this option, the GIS program takes the selected layer's parameters and establishes them in the reprojected layer. This saves time especially if you have multiple layers to reproject.

Georeferencing

Any scanned image can be entered into a GIS, but to be useful, the image needs to be placed in its proper geographic location. Georeferencing aligns images to their spatial location. This process is common due to the popularity of "heads up" digitizing (described in Chapter 2).

Georeferencing is typically done by aligning the image to existing projected feature layers that are in their correct position. Since any scanned image is fundamentally a matrix of pixels, georeferencing the raster layer involves moving and stretching this matrix so that it rests at its true location (see Figure 3.7). To do this, you need to load the unprojected image and the projected feature layers and—in order—select corresponding control points, which are locations you can distinguish on both the image and the feature layers (left map in Figure 3.7). For greater accuracy, select as many control points as possible and make sure they are scattered throughout the image. If they are clustered in a corner of the unprojected image, only that part of the image will be georeferenced properly.



Figure 3.7: Georeferencing an image to its real-world location. The left map displays both the image we want to move and the area's parcels. The red arrows represent the corresponding points where you want to move points in the image. When the georeferencing is complete, the image will be at its real-world location.

Georeferencing assigns coordinate information about where the image rests in relation to the Earth's surface. When you save your georeferenced image, a "world" file is created. This is an ASCII text file that has the exact name of your image file but with a

different yet related file type. For example, if you have a TIFF image called Mexelev.tif, the world file (called a GeoTiff file) will be Mexelev.tfw. The “w” at the end of the file type denotes its status as a world file. Most GIS software packages are able to interpret these files and display the images in their proper location as long as the file names are the same and the two files are located in the same directory.

The first column of Figure 3.8 is an example of a world file. It has six lines with locational values. The second column describes what the six locational values are, and it is not contained in the world file.

5.00	Pixel resolution in x direction. This describes the distance on the Earth's surface that a pixel covers.
0	Rotation term for row. This and the next line describe rotational components. If the mapsheet is unrotated, they are set to zero.
0	Rotation term for column. See above.
-5.00	Pixel resolution in y direction. Usually the same as the first line but with a negative if the pixel is square.
492169.690	This is the x coordinate of the center of the image's upper left pixel.
5426523.318	This is the y coordinate of the center of the image's upper left pixel.

Figure 3.8: Typical world file format.

Resampling

As briefly mentioned above, resampling changes raster layers from one projection to another, but it can also be used to transform the resolution of raster images. For example, resampling can convert each 2 by 2 array of pixels (4 pixels total) into a single but geographically larger pixel. To accomplish this, it changes the pixel's attribute values with mathematical formulas to best approximate the attribute values for the new layer. For instance, it might average the four numeric values and place the mean in the single resampled pixel that replaces them in the new image. In Figure 3.9 below, the image to the right is a generalized resampling of the image to the left. Resampling is important if you are working with multiple raster images with varying resolutions. You need to translate your images to a common resolution (much like a common projection) to analyze them.

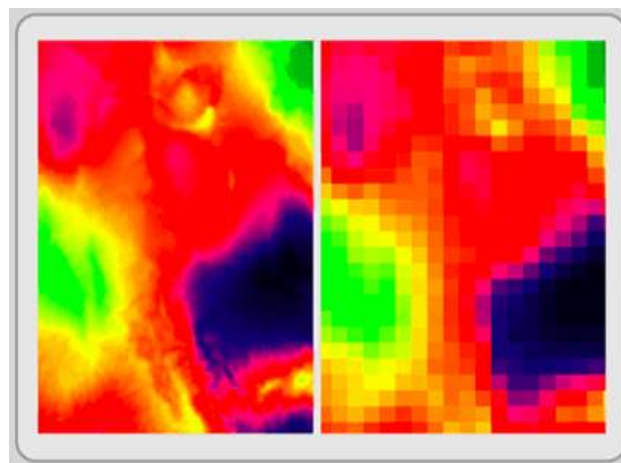


Figure 3.9: Resampling. The original image on the left had a pixel resolution of 250 meters. The resampled image on the right has a resolution of 2500 meters.

Reclassification

Reclassification generalizes values in a raster layer to highlight broader classes. This popular preprocessing technique re-assigns values in an input raster layer to create a new, more generalized, raster layer. Reclassification changes pixel values based on a criterion that you specify. In Figure 3.10, a raster image that denotes land covers is reclassified into two values. Reclassifying the database may reveal broader patterns by removing the layer's unique classes. Reclassification is also commonly used to convert interval and ratio attribute pixel data into ordinal data used in the overlay process.

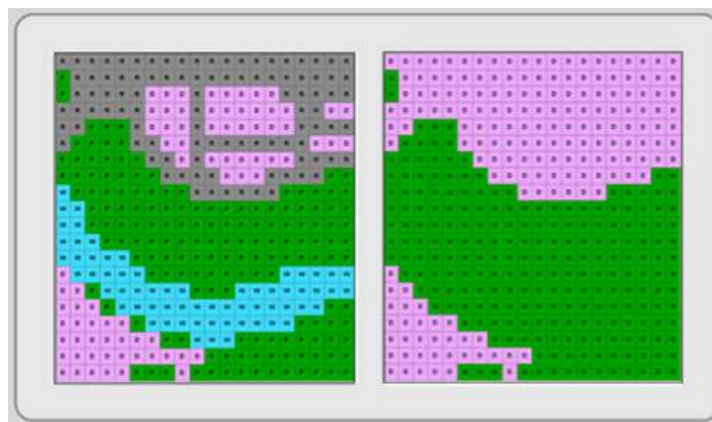


Figure 3.10: Reclassification. The image on the left depicts 4 different land covers. The image on the right aggregates land covers (D and R become D; P and W become U) into two classes. Image by Mike Tuck.

Edge Matching

When side-by-side map layers are retrieved and displayed, they might not line up well with each other (see Figure 3.11 below). Edge matching adjusts the location of features that extend across one map's boundaries into another.

Edge matching requires your input in matching together the common edge of the two maps. The features that you believe are positioned correctly are usually "anchored" down, and the remainder of the map is moved, stretched or contracted like a sheet of rubber to line up the features on the maps. The map features, except those that are anchored, are spatially adjusted.

Which map features should you anchor down and which should be stretched? It is not an easy question. The answer might be found in the layer's metadata. Perhaps one layer was entered at a coarser (less accurate) scale or with less precision. If, however, the properties and author of the two layers are identical, you should use a third layer (perhaps a georeferenced aerial photograph) that you have some confidence in to check the positional accuracy of the features within these two layers. When all else fails, you might have the features split the difference.

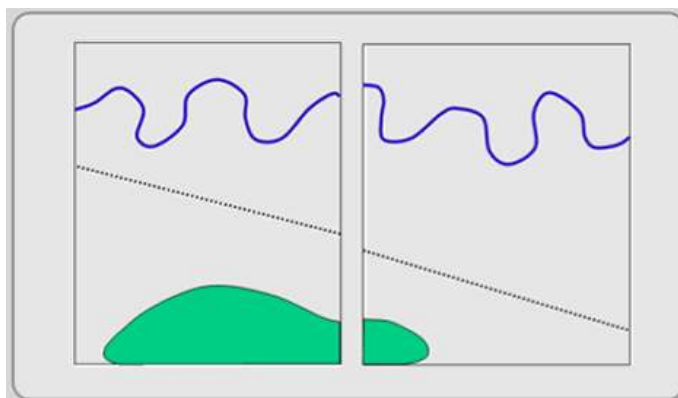


Figure 3.11: Edge matching. Sometimes two maps that should fit up against each other do not. Edge matching manipulates one of both of them until they do.

Conflation

Conflation is similar to edge matching with one difference: It does not rectify the placement of features across maps. Instead, it tries to rectify feature locations within a single raster image. For that reason it is also referred to as rubber sheeting. It is an interactive process where you tack down features that are positioned correctly and move the remainder to more accurate locations.

Tiling

Sometimes workspaces get large; geographically they can be vast and thematically numerous. Tiling involves breaking up your workspace into more manageable and logical geographic subunits. Tiling subdivides existing layers (both the geography and the

attributes) by geographic units. Figure 3.12, displays a portion of the U.S.G.S. topographic map grid across California. Tiling can be done by splitting existing layers with the larger geographic boundary or it can be planned from the outset of the GIS project. The GIS then maintains a library of all the tiles that represent the project area.

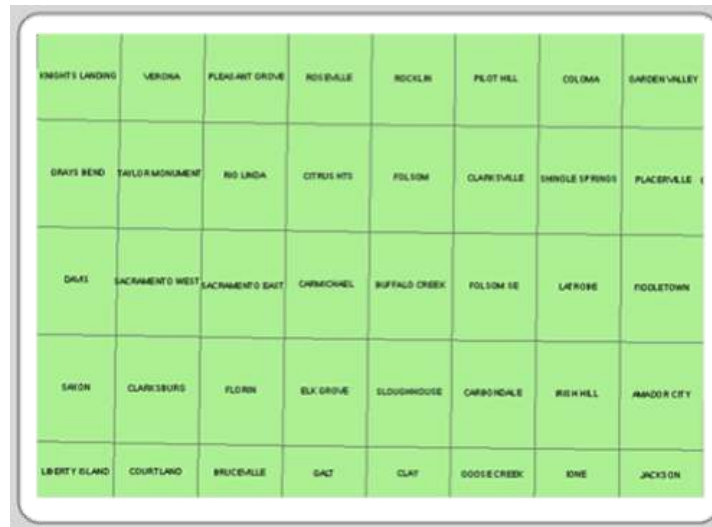


Figure 3.12: U.S.G.S. quadrangle maps for a portion of California.

Vectorization and Rasterization

These two common processes switch feature layers between vector and raster. For example, you might digitize data into a vector format but want to use it in a raster form. Vector layers are converted to raster by a process known as rasterization (see A in Figure 3.13). Alternatively, raster data can be converted to a vector layer through vectorization (see B in Figure 3.13).

Like any translation, it is not error free. Think, for example, of converting points from vector to raster. Each precise point location in the vector layer swims in the pixel of the new raster layer in which it now belongs. The precise spatial locations of the points are lost because the points now reside in much larger areas, which are determined by the pixel's resolution. Now convert the new raster point layer back to vector. The points in the resultant vector layer are located at the center of each pixel in which they were contained. Comparing the new and original vector layers, you would see that they resemble each other but do not line up exactly.

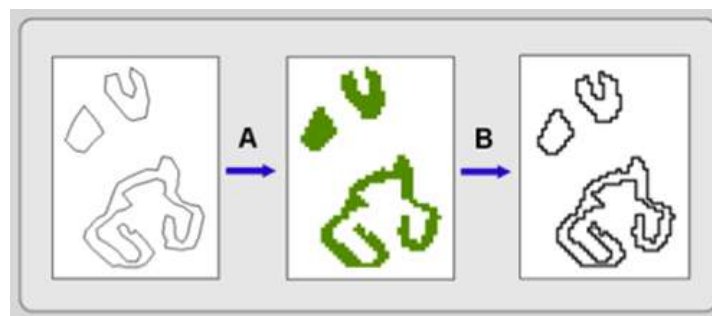


Figure 3.13: From vector to raster to vector again. Converting a vector file to raster is called rasterization (process A). Converting from raster to vector is vectorization (process B). Notice the differences between the first and third maps, both vector layers.

Coordinate Thinning

Coordinate thinning (also known as map generalization) generalizes or “smoothes” feature shapes by removing nodes (vertices) from line and polygon features. It reduces layer storage size, and it can be used to remove unwanted detail from map features. Sometimes detail held in a layer is not always appropriate for a small-scale map. For example, in the top map of Figure 3.14, notice how the detail of portions of the coastline resembles ink blobs because of the amount of coastal detail. The detail may take away from the map's purpose. If one were to enlarge the image, the detail might be welcomed. By thinning the vertices along the coast,

the map becomes simpler and clearer (bottom map in Figure 3.14). Perhaps this map is too generalized at this scale. Notice that a couple of islands have disappeared.

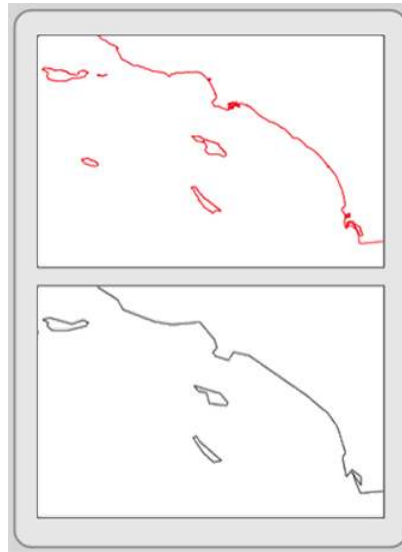


Figure 3.14: Coordinate thinning. Notice the differences in detail.

Topological Functions

Topology, the spatial relationships among features, focuses on where features are in relation to one another and how they are related to one another. Focusing on how features relate to each other, topological functions (a semi-automated process) help you clean up your layer's spatial errors and determine what parts of different features are shared, contained, or connected to other features. In other words, these functions build topology. Most vector-based systems provide routines that help you find the following common topological problems (see Figure 3.15):

1. Slivers, the most common topological problem, are small polygons that occur when either shared boundaries are entered separately for contiguous polygons or when the features of two layers are overlaid but do not match precisely. Topological functions can remove many of these slivers and reconcile common boundaries.
2. Overshoots and undershoots usually occur when features are entered without the aid of a snapping routine. The feature vertices extend beyond (overshoots) or just short of (undershoots) their intended location. Topological functions can clean up these errors when you define a distance tolerance. If within the distance, the overshoot or undershoot snaps the vertex of one feature to the vertex of another feature.
3. Redundancy occurs when two or more features in the same layer share the same node (vertex) or line but the layer duplicates these nodes (vertices) and lines. Layers should only store one node or line, which prevents duplication that could lead to errors. Most GIS programs have automatic elimination routines to eliminate duplicates.

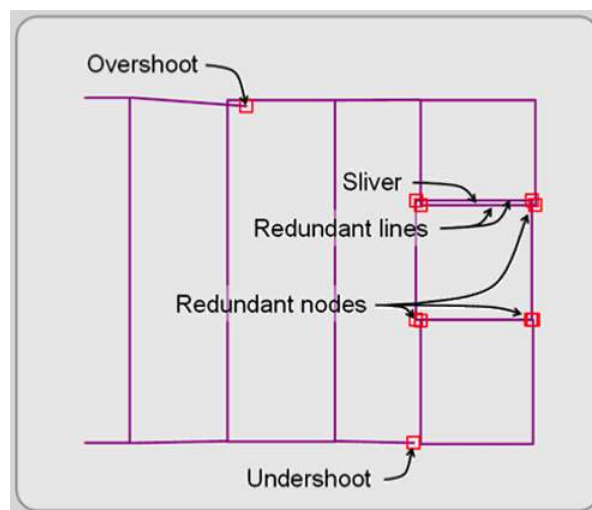


Figure 3.15: Typical topological errors.

This page titled [3.3: Map Preprocessing](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

3.4: Map Verification and Editing

Within each of your GIS layers, you need to identify and edit errors. This applies not only to your primary and secondary datasets, but also to every layer you preprocess and create as the result of analytical functions (those discussed in Chapter 5). Many spatial errors are the result of careless digitizing, inaccurate source maps, or a change in the area's geography (like new buildings, boundaries, and forest fires).

Map Verification

This section looks at identifying and editing spatial errors. The last section of the next chapter deals with database errors. Map verification includes the following three steps:

1. Visual examination. Use your eyes and your familiarity with the study area and the subject matter to check the spatial locations of features. In this step, verify that all of the features are present, in their correct location, and that they are the correct size and shape. Make sure that no extra features are present that do not exist (see Figure 3.16). Examine vector layers by loading a georeferenced and accurate reference image (an aerial photograph or a Digital Orthophoto Quadrangle (DOQ)) under the layer you are verifying. For raster images, compare the layer with other accurate raster layers and visually overlay vector layers. If you are familiar with the subject matter and the study area, you can pick up many errors during this step.



Figure 3.16: Map verification: 1 = layer is missing building. 2 = building's position needs to be altered. 3 = layer may be missing buildings (check with additional sources or field check). 4 = layer depicts building that does not exist.

2. Cleaning lines and vertices. This process is usually done by software first and interactive editing second. The software's cleaning applications may include coordinate thinning and topological functions (see above).
3. Comparison with source document. Although visual examinations pickup many errors, compare the layer's completeness, position, size, and shape directly with the original hard-copy document. To do this, you can plot the layer you are verifying at the same scale as the original document, and with the help of a light table, superimpose the layer that you are verifying over the original and note the discrepancies. Take your time with this step; systematically check off each feature. Again, like in step 1, using an additional, high-quality layer to independently verify the position of features is a good idea especially if you have any doubts about the quality of your original map.

Data Editing

Spatial data editing involves adding, deleting, moving, and changing the shape of features. Data editing fixes the errors you find during the verification process. Each GIS program has its own particular way of editing the location of features (or a portion of a feature). Besides the elimination of slivers, overshoots, undershoots, and redundant nodes and lines, you may need to merge features, split features, and simply move individual nodes (vertices).

To make clean edits, know how to use your program's snapping routine. When entering or editing contiguous or connected features, snapping moves your cursor slightly to align it with an existing node (vertex). This reduces overshoots, undershoots, and redundant points and lines.

While all GIS packages have entering and editing capabilities, many programs do not make editing (and entry) easy or intuitive. Because of this, many GIS users import their feature layers into CAD systems for detailed editing and data entry. CAD programs have specialized, easy-to-use tools for editing features. Why is CAD easier for entry and editing? They were created for precise drawing and drafting. GIS programs usually focus on compiling layers and on analyzing them. Many agencies and companies that use both CAD and GIS programs often refer to their CAD programs as legacy systems—an adjective that refers to their past prominence and infers that their organization is moving away from them. Many organizations, however, recognize CAD's superior data entry capability and expect to keep these systems at least for a while.

Geometric Conversion

Geometric conversion consists of routines that change feature types between points, lines, and polygons. Perhaps the most frequent geometric transformation is the conversion of lines to polygons. This happens a lot because many CAD programs often use line segments to build parcels (and other features), but within a GIS these features are best coded as polygons. Another routine changes points to contours (lines), which are used to portray surface relief as a set of lines that connect points of the same value. Conceptually, contours are “threaded” through the points (or pixels) along approximate lines of constant value. Points are also frequently produced from polygons. This routine creates a “centroid”, a central point, usually placed at the intersection of the north-south and east-west mid points.



Figure 3.17: Creating centroids (point features) from polygons. The process automatically places a point at the half way mark of both its west-east and north-south range. In some cases, this places the state's centroid outside of its polygon. Many of the centroids in the above map have been moved to more of a visual center.

This page titled [3.4: Map Verification and Editing](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

4: Attribute Tables and Data Preprocessing

[4.1: Introduction](#)

[4.2: Attribute Data](#)

[4.3: Principles Of Database Management – Vector](#)

[4.4: Principles Of Database Management – Raster](#)

[4.5: Attribute Preprocessing And Editing](#)

[4.6: Attribute Verification](#)

This page titled [4: Attribute Tables and Data Preprocessing](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

4.1: Introduction

GIS programs link attribute data files to digital maps. The previous Chapter focused on the map side of this equation. Let's focus now on the attribute data files. Like the previous chapter, this chapter examines several key concepts and covers the preprocessing of your GIS data, but it specifically focuses on attributes, data files, and the editing of your attribute data. The concepts focus on attribute data and principles of raster and vector database management. Understanding these concepts will help you to effectively edit and manage your attribute data. The bulk of the chapter focuses on various preprocessing routines including adding and deleting fields, deleting records, joining data files, selecting and sorting records, calculating attributes, and geocoding. The chapter ends with a short discussion regarding attribute verification.

This page titled [4.1: Introduction](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

4.2: Attribute Data

As described in the previous chapter, spatial data occupies geographic space. It has a specific location that is tied to one of the world's geographic referencing systems (like latitude and longitude). Besides spatial data, GIS files contain non-spatial attributes that describe the spatial features. This section focuses on these non-spatial attributes.

Related to the discussion of “measurements of scale” in Chapter 2, your attributes can be classified as either qualitative or quantitative and actual or derived. Quantitative data focus on numbers and frequencies rather than on subjectivity, meaning, and experience. They are easy to analyze statistically, and their values are often the result of field work and laboratory experiments. Maps exhibiting quantitative data depict differences in magnitude among features.

Qualitative data, by contrast, often provide deeper description and meaning. Maps displaying qualitative data show differences in kind or type. You might subjectively judge whether a quantity is low, medium, or high. You might also classify detailed land uses into broader categories of residential, commercial, and industrial. The statistical options are narrowed too due to the subjectivity of the data and the categorization of data into classes.

Data can also be defined by whether they represent some intrinsic characteristic of the feature being measured (absolute), or whether they are in a sense “created” (derived). Absolute data consists of both the quantitative and qualitative data just described, but it represents phenomena that are measured (like election data or the amount of water stored), the ranking and rating of attributes (even though this process can be subjective), and personal, subjective accounts gained from questionnaires and surveys.

Derived attributes either do not occur naturally, or they cannot be directly gathered; they are the result of statistical manipulation that produces the data. An example is average July temperatures, which is the calculated result of averaging many actual temperature values. Derived data may result from averaging actual values like these, or they represent the relationships between already gathered attribute data, which take three forms: ratio, proportion, and percentage.

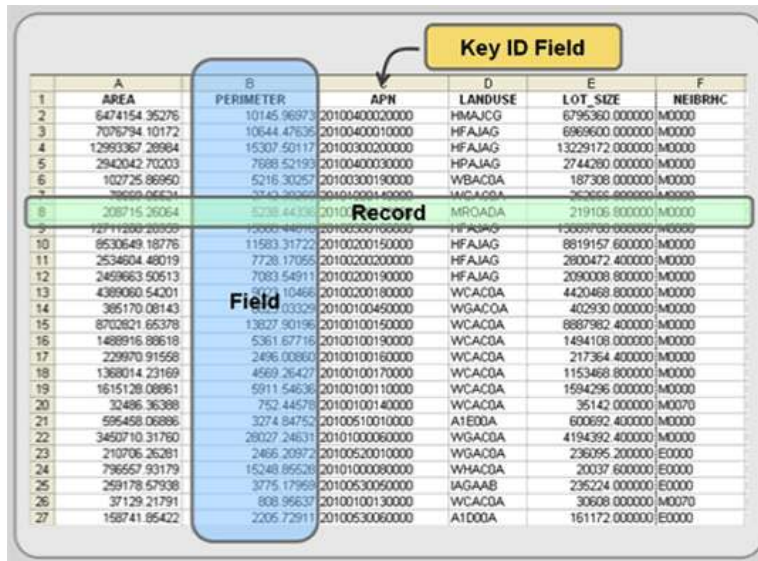
- *Ratio* attributes are derived when the value of one attribute is divided by the value of another. Population density is a good example. The total number of people within a particular region is divided by the region's area. Both the population and area attributes may be “actual” values, but the calculated population density attribute is derived.
- *Proportion* compares the value of one attribute to the total value of all related attributes. The proportion of all African-Americans to the total population is derived by dividing the number of African-Americans (actual data) by the total number of people (also actual).
- Many people think of proportions as *percentages*; they are similar, but percentages multiply proportions by one hundred.

This page titled [4.2: Attribute Data](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

4.3: Principles Of Database Management – Vector

Let's turn our discussion from characteristics of data to how these values are organized within a data file. Data files are the basic "database" for many programs including spreadsheets, statistic programs, and GIS. Within a GIS, there is a data file for each particular type of geographic feature (e.g. streets, street lights, buildings, and parcels of land). They are the database's version of your features. The data files are automatically created when feature layers are defined in your GIS. You place into them the attributes related to the features.

Data files, often called "tables," arrange attributes within a matrix of fields and records. Fields form the columns of a data file (see Figure 4.1), and they contain the values for each specific attribute you are collecting. For example, parcels might include attributes such as area, land use, and Assessor's Parcel Number (APN). In this example, you would have at least three fields: one called area, another titled land use, and one labeled APN.



	A	B	C	D	E	F
	AREA	PERIMETER	APN	LANDUSE	LOT_SIZE	NEIBRHC
1						
2	6474154.35276	10145.96973	20100400020000	HMAJCG	6795360.000000	M0000
3	7076794.10172	10644.47635	20100400010000	HFAJAG	6969600.000000	M0000
4	12993367.26984	15307.50117	20100300020000	HFAJAG	13229172.000000	M0000
5	2942042.70203	7688.52193	20100400030000	HFAJAG	2744280.000000	M0000
6	102725.86950	5216.30257	20100300190000	WBACDA	187308.000000	M0000
7	78959.05634	3343.30350	20101000140000	WCACDA	303665.000000	M0000
8	208715.26064	5238.44358	20100300180000	MROADA	219105.000000	M0000
9	12711200.20350	13000.44670	20100300160000	HFAJAG	13009700.000000	M0000
10	8530649.18776	11583.31722	20100200150000	HFAJAG	8819157.600000	M0000
11	2534604.48019	7728.17055	20100200200000	HFAJAG	2800472.400000	M0000
12	2459663.50513	7083.54911	20100200190000	HFAJAG	2090008.800000	M0000
13	4389060.54201	7501.10466	20100200180000	WCACDA	4420468.800000	M0000
14	385170.08143	33329.03329	20100100450000	WGACDA	402930.000000	M0000
15	8702821.65378	13827.90196	20100100150000	WCACDA	8887982.400000	M0000
16	1488916.88618	5361.67716	20100100190000	WCACDA	1494108.000000	M0000
17	229970.91558	2496.00860	20100100160000	WCACDA	217364.400000	M0000
18	1368014.23169	4569.26427	20100100170000	WCACDA	1153468.800000	M0000
19	1615128.08861	5911.54636	20100100110000	WCACDA	1594296.000000	M0000
20	32486.36388	752.44578	20100100140000	WCACDA	35142.000000	M0070
21	596458.06886	3274.84752	20100510010000	A1E00A	600692.400000	M0000
22	3450710.31760	28027.24631	20101000060000	WGACDA	4194392.400000	M0000
23	210706.26281	2466.20972	20100520010000	WGACDA	236095.200000	E0000
24	796557.93179	15248.85528	20101000080000	WHACDA	20037.600000	E0000
25	259178.57938	3775.17959	20100530050000	JAGAAB	235224.000000	E0000
26	37129.21791	808.95637	20100100130000	WCACDA	30608.000000	M0070
27	158741.85422	2205.72911	20100530060000	A1D00A	161172.000000	E0000

Figure 4.1: Key parts of a data file.

Remember from Chapter 2 that each of these fields has a specific "data format" that defines the type and length of the value that can be directly entered into the data file. Frequently attributes are coded as one of the following, but there are many data formats and the specific name of the data format often changes from one software program to another. Broad data format categories include:

Integer	Numeric values consisting of whole numbers. No decimals.
Real	Numbers consisting of integers with decimals.
Byte	Numeric values ranging from 0 to 255.
Character	Alphanumeric values.

Figure 4.2: Data format categories.

A single record, a row in the data file, represents the database's version of a single feature, including all of its specific attribute values (see Figure 4.1). A few of these attributes may be system variables that the GIS needs for data integrity reasons and to link the data file to the feature's spatial files. In addition, some GIS programs automatically generate length calculations for line features and both area and perimeter calculations for polygon features. Each data file should have a key identifier field that uniquely identifies each feature (i.e. each record). The remaining attributes are up to you and the purpose of your study.

Data files are a collection of related records. If you have 25 street lights within your GIS, you will have 25 street light records in its attribute file. As briefly described above, a largely empty data file is created when a new layer is defined within a GIS program. It

is your job to add fields and attribute values to the data file. These descriptive attributes can be entered by hand or imported from external sources. It is likely that you will enter some attributes by hand (and it can be time consuming and tedious), but many—if not most—of the attributes you seek will be imported or “joined” from separate, non-GIS data files. This is because many non-spatial data files predate your need for their incorporation into a GIS, but it is deeper than that. Data manipulation within GIS is clumsy, and since most GIS users are familiar with data management programs like Excel and Access, they prefer working with these programs and then exporting their data and “joining” the external data file to the GIS data file. The joining process is described later in this chapter.

These external data files are coded in one of many “file formats”. Some file formats are specific to a particular software program while others are somewhat universal. Even those using a program’s proprietary format can export the data file into one of many formats that most GIS programs can read. Some of the file formats that can be read by most GIS programs include:

dBase This industry standard format is read by just about every GIS program. Many GIS programs use this format internally rather than creating their own.

Excel and Access – Microsoft’s file formats for Excel and Access can be read by many GIS programs. If your GIS program does not read these formats, open the data file in Excel or Access and export it into a format that your system reads.

ASCII (American Standard Code for Information Interchange) – Since most computers use ASCII to represent text, it is possible to transfer data from one computer to another in this format. It is also read and written by most GIS programs, but it is rarely used as the primary GIS file format (with the exception of some raster-based GIS programs). Some government data sets are contained in this file format. Text files come in several different “delimited” forms, and all may include numeric or alphanumeric content (see “Joining Data Files” later in this chapter).

Data files contain a matrix of fields and records for each feature layer. A database is a collection of several related data files (like parcels, street lights, and buildings). In other words, databases contain data files for related layers. Accessing these data files are done through either the GIS software or increasingly from external database management systems (DBMS) that are linked to the GIS. DBMS are specialized programs that organize, manipulate, and report non-spatial data and help you store your data more efficiently. They are particularly valuable when working with large data sets because you can select a subset of your records and fields to work with. The entire attribute file does not have to be used. Examples of external DBMS programs include Access, Oracle, Ingres, SQLServer, INFORMIX, and to a lesser degree Excel, which can serve as an elementary database program. Regardless of whether you are accessing the data files within the GIS software or from an external DBMS, all databases have standard operations which include sorting and selecting records, deleting records and fields, and editing fields and attributes.

Different databases have different structures or ways to organize data. The hierarchical and network data models are two examples, but they are rarely used for GIS (and so will be skipped in this section). For vector systems, the relational database model is the most common data model arguably because they are more flexible, the table structure is easy to understand and program, and outside of GIS, data files are commonly held in relational databases.

Linking or joining data files is the relational database model’s strength. Key identifiers, found in multiple data files, are used to link records from one data file to another. In other words, you cross reference multiple data files using common attributes and attach (or join) these external data files to your internal GIS data file. This link takes the selected fields in the data file you wish to join and relates them to the appropriate records in the GIS data file. This requires that each data file have at least one common field to perform a join. There are different names for the key identifier including key and primary key. This process is highlighted later in this chapter.

Many, however, think that the relational database model does not adequately represent spatial data. For some, records in a relational data file are too discrete; they do not properly depict the continuous and multi-dimensional nature of the features they are representing. We use relational data models because they are simple and convenient, but we artificially bend geographic features to conform to existing database standards that were created for non-spatial data.

This has led to the development of object-oriented data structures, which are seen as a more sophisticated database model. The database discards many of the foundational concepts that we have applied throughout this book. Features are defined differently; object-oriented features blur the line between points, lines, and polygons. Also, instead of having multiple files for each GIS layer, the geography and attribute data are integrated into a single file. This allows for simultaneous geographic and attribute editing and

quicker processing. The more sophisticated model, however, is a more complex model, and that may have slowed its spread even though “object-oriented” databases were one of the hottest topics in GIS in the 1990s. It may still be the touted successor of the relational model, but it seems that the relational model, despite its drawbacks, has significant pluses—including its ease of use—that will help it dominate at least into the near future.

This page titled [4.3: Principles Of Database Management – Vector](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

4.4: Principles Of Database Management – Raster

As described in Chapter 1, the raster data model aligns the Earth's surface into a grid of columns and rows. Cells, or pixels, the building blocks of the raster data model, form at the intersection of the columns and rows, and each cell contains a single attribute value, representing the condition of a specific portion of the Earth's surface. That means that a single raster layer only contains the values for one specific attribute across space. That last point is important because raster layers fill space. Their attributes occur everywhere in the study area; there are no blank spaces. Empty areas get a "0" value, but every pixel gets a value. If you need more than one attribute, you construct multiple layers, each containing a single specific attribute for the same area. Conceptually, it is a simple model. As in Figure 4.3, your study area is divided into cells, and each cell of each layer has a single attribute that represents that area.

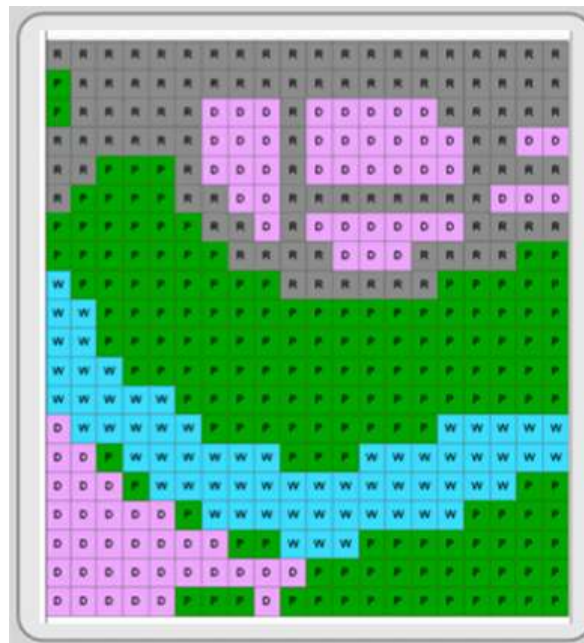


Figure 4.3: Raster image. Image by Mike Tuck.

There are many ways—some more complex than others—the raster data model may be stored. The two general categories are regular and irregular. The regular structure is conceptually simple, and includes two types: full raster encoding and run-length encoding. Full raster encoding creates a data file that records the attribute value for every pixel. It's as though you read an image's pixels like a book, starting in the upper left corner and reading from left to right and downward row by row. The data file looks a bit different. It records each pixel's attribute value on a separate line, so if you had an image with 640,000 pixels, your data file would have 640,000 lines, making it a very long data file. Figure 4.4 is a simplified example.

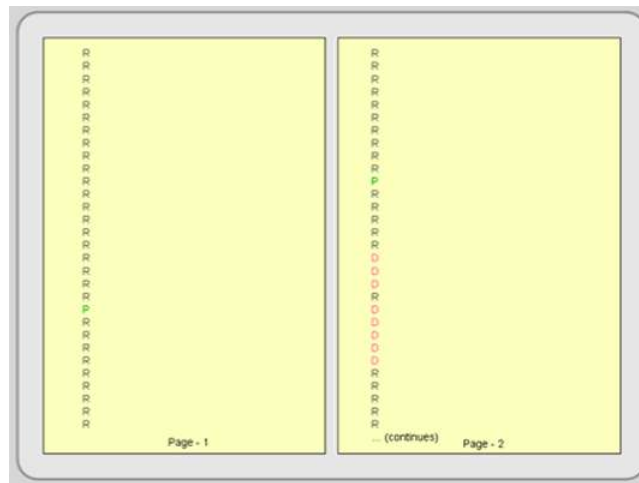


Figure 4.4: Full raster encoding. This figure is the beginning—just the first three rows—of the data file for the image in Figure 4.3. Color is added to highlight the different attribute values.

Run-length encoding is more efficient than full raster encoding. Since the same values often occur in runs across several cells, run-length encoding enters the attribute values as pairs: the first number is the run length and the second number is the cell's value. This substantially reduces file size especially if contiguous pixels have the same value. Contrast Figure 4.5 with Figure 4.4.

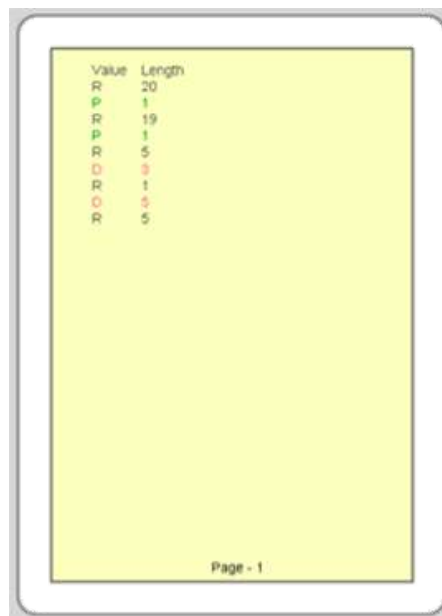


Figure 4.5: Run-length encoding. This figure also depicts the first three rows of Figure 4.3. Compare run-length encoding with full raster encoding (Figure 4.4). Color is added to highlight the different attribute values.

Irregular raster data structures, like quadtree and others, are more complex, proprietary, and beyond the scope of this e-text. They usually make file size smaller and provide ways to store raster data for quick retrieval.

This page titled [4.4: Principles Of Database Management – Raster](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

4.5: Attribute Preprocessing And Editing

When you add feature layers, containing both spatial and attribute data, to an active workspace, the attribute data file might not be immediately visible. Opening and editing the attribute files are easy processes, but they are specific to individual programs. Once the attribute table is open, you can enter data by typing attribute values directly into the data file or loading and joining external data files to it. Other processes like editing attributes, adding or deleting fields, deleting records, querying attributes (record selection), calculating fields, and geocoding are completed through the data file interface.

Adding and deleting fields

As described above, fields define feature attributes. Most GIS programs provide a way for you to add or delete fields from within your open data file. The GIS program will instruct you to define a new field. You will give it a name and select from options that determine the data format of the values that will be placed into the field. Deleting a field usually involves selecting the field and deleting it.

Deleting records

You can delete a single record or a group of records in a data file by first selecting them and then deleting them. Since records are the database representation of features, when you delete records in the attribute file, you are also permanently discarding their spatial representation. The entire feature, graphic and record, is deleted.

Generally, you can not add a record through the data file interface because it must also be represented spatially. See Chapter 3 for how to add a feature. Its record is automatically created when the graphic feature is added to the workspace.

Joining Data Files

Once a GIS layer is created, its attribute file can be linked (“joined”) to external data files. Joining is one of the most frequently performed data file processes because it brings together feature attributes that are contained in multiple digital data files. To perform a join, a unique matching field, the key identifier, must be observed in both data files. As stated in Chapter 3, the key identifier could be something like a social security number or an assessor parcel number. It is a field that gives the feature a unique identification. Once linked, the join can be temporary or made permanent.

The external files that you load into the GIS to perform a join are typically in file formats such as dBase, ASCII, Microsoft Excel, or Microsoft Access. The precise steps involved in joining together two files are software specific, but it usually involves:

1. loading the external file that you wish to join to the GIS attribute file,
2. selecting the external file and the GIS attribute file that you wish to join,
3. selecting the field (containing the key identifier) in each file, and when joined,
4. making sure that the join was successful.

In the example in Figure 4.6, the parcel layer exists, but it does not include assessed value. It does contain a field named APN (Assessor’s Parcel Number) whose values are unique to each record and which could be used to join other data files. A spreadsheet file, with assessed value, also exists, and it must be loaded into the GIS either in its native format (if accepted) or exported from the spreadsheet program to a format that the GIS can read. The spreadsheet has a field named APN_NUM, which, after a visual check, has the same values as those under APN in the parcel layer, and it can be used to perform the join.

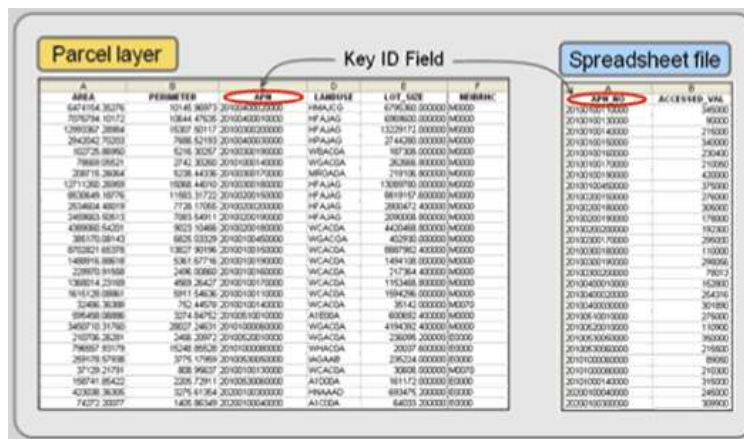


Figure 4.6: Joining two attribute files together requires that the two files each have a common key identifier.

Once the spreadsheet file is loaded, you begin the joining process by specifying the two files (the layer's table and the spreadsheet file) and the two field names that the join will be made on. APN and APN_NUM are the key identifiers of these two files (see Figure 4.7), and even though the field names are not identical, the GIS will be able to join these two files together provided that the values under the two field names match.

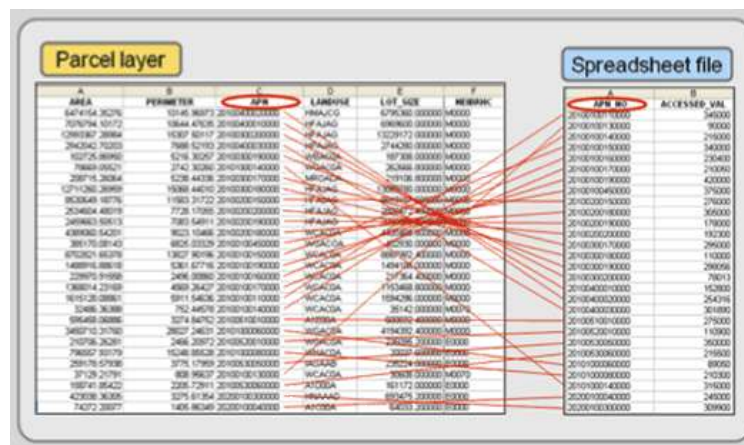


Figure 4.7: Matching key identifiers.

If the match is successful, your two files will be joined together into a single file (see Figure 4.8).

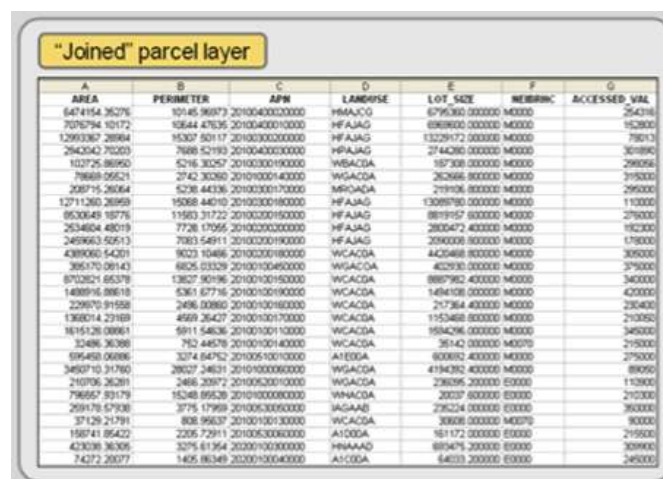


Figure 4.8: A joined file with accessed values a one of the attributes.

Perhaps the most time consuming tasks are the first and fourth steps. Loading an external data file should be easy—and frequently it is—but sometimes the imported data file may be misformatted or unreadable. If it is, return to the host program (your spreadsheet or DBMS programs) and save it in a different format. The probability of your GIS program being able to read the external data file usually improves as you go from more sophisticated file formats (like Excel and Access) to dBase to ASCII (basic formats). Many data files are coded in ASCII because of its almost universal compatibility with computers and software programs, but it does have its complications—it comes in several forms. Below are four of the most used variants of ASCII based on what delimits the file's fields.

Whitespace delimited ASCII files differentiates fields by the use of one or more spaces. Since spaces separate fields, fields that have no value must be represented by a non-blank code and character attributes cannot contain spaces between words (underscores can be used to separate words). You can open ASCII files in any word processor or text editor. A whitespace-delimited ASCII data file with five records might look something like the following:

```
M1 Betsy_Burns Yes 38.5 0.85
P1 Dan_Arreola No 45.7 0.99
M2 Frank_Aldrich Yes 32.8 0.55
P2 Fritz_Steiner No --
P3 Ruth_Yabes No 37.72 --
```

Spacequote delimited ASCII is a variant of whitespace delimitation, but the attributes containing multiple words are enclosed in double quotes, and consequently, they can contain embedded spaces between words. The spacequote delimited ASCII file may look like the following in a text editor:

```
M1 "Betsy Burns" Yes 38.5 0.85
P1 "Dan Arreola" No 45.7 0.99
M2 "Frank Aldrich" Yes 32.8 0.55
P2 "Fritz Steiner" No --
P3 "Ruth Yabes" No 37.72 --
```

Tab delimited files separate fields by the use of a single tab. Two tabs in a row signify a blank field. Values within an attribute field cannot contain embedded tabs. A tab delimited ASCII file would look like the following in a text editor.

```
M1 Betsy Burns Yes 38.5 0.85
P1 Dan Arreola No 45.7 0.99
M2 Frank Aldrich Yes 32.8 0.55
P2 Fritz Steiner No
P3 Ruth Yabes No 37.72
```

Comma delimited, also known as comma-quote delimited and CSV, separate fields by commas. Character fields may be enclosed in double quotes, and need to be if they contain an embedded comma. Two commas in a row signify that the field is blank. Usually whitespace is not allowed before or after fields (although this may be tolerated in the CSV form). The comma-delimited ASCII file might look like the following in a text editor:

```
M1,"Betsy Burns",Yes,38.5,0.85
P1,"Dan Arreola",No,45.7,0.99
M2,"Frank Aldrich",Yes,32.8,0.55
P2,"Fritz Steiner",No,,
P3,"Ruth Yabes",No,37.72,
```

Sorting records

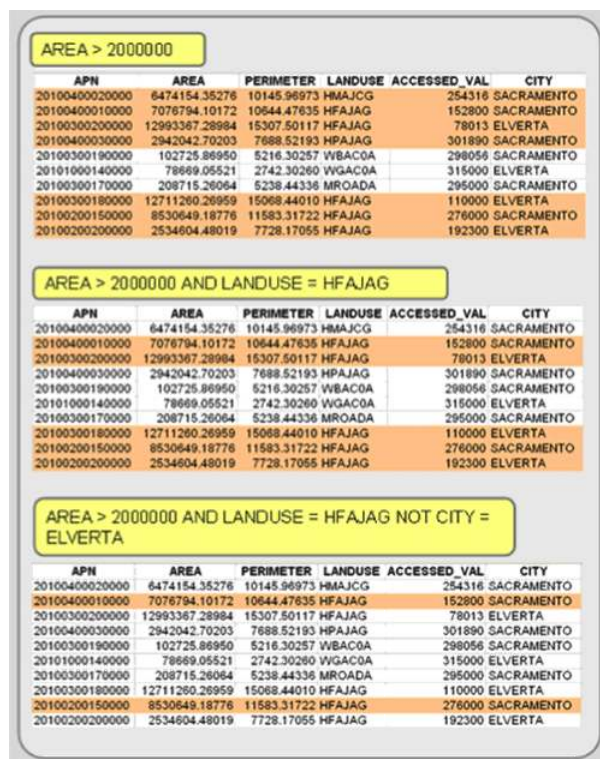
Sorting temporarily rearranges your data file records, so you can view, select, update, or print them in the new sorted sequence. Although the specifics vary by program, you generally choose the field (or fields) you want to sort by. The first sort field arranges, usually in ascending or descending order, the records based on the field's contents. For example, a class roster might be sorted alphabetically by last name. Some systems allow you to choose a second sort field (or more), which arranges records (in ascending

or descending order) when two or more records have the same first field value. In the example above, if your alphabetical list has four students with the last name Smith, those four records could be rearranged in alphabetical order based on their first name.

Record selection/Attribute Query (Boolean Selection)

Selecting specific records is one of the most common database functions. Often called attribute query, it consists of highlighting a subset of the records based on a specific criteria. In other words, you create an expression—a formula—that queries all the records in the data file and the GIS highlights—both in the data file and on the map display—only those features that fit the criteria.

Most GIS programs use a Standard Query Language (SQL) interface to conduct attribute queries. If one is using an external relational DBMS program (like Access or Oracle), SQL makes the call to the external database and isolates only the necessary records that you will use. SQL uses set algebra, Boolean algebra, and arithmetic operators (=, -, *, /) for attribute queries. Set Algebra includes the use of less than (<), greater than (>), equal to (=), and not equal to (<>) operations. You can create an expression like that found below (see Figure 4.9) to isolate only those records that fit your criteria. You can extend or constrain the selected features by using Boolean algebra, which uses the conditions OR (extend), AND (constrain), and NOT to further select or isolate records. Each record is queried and added to the set if it meets the criteria.



AREA > 2000000					
APN	AREA	PERIMETER	LANDUSE	ACCESSED_VAL	CITY
20100400020000	6474154.35276	10145.96973	HMAJCG	254316	SACRAMENTO
20100400010000	7076794.10172	10644.47635	HFAJAG	152800	SACRAMENTO
20100300200000	12993367.28984	15307.50117	HFAJAG	78013	ELVERTA
20100400030000	2942042.70203	7688.52193	HFAJAG	301890	SACRAMENTO
20100300190000	102725.86950	5216.30257	WBAC0A	298056	SACRAMENTO
20101000140000	78669.05521	2742.30260	WGAC0A	315000	ELVERTA
20100300170000	208715.26064	5238.44336	MROADA	295000	SACRAMENTO
20100300180000	12711260.26959	15068.44010	HFAJAG	110000	ELVERTA
20100200150000	8530649.18776	11583.31722	HFAJAG	276000	SACRAMENTO
20100200200000	2534604.48019	7728.17055	HFAJAG	192300	ELVERTA

AREA > 2000000 AND LANDUSE = HFAJAG					
APN	AREA	PERIMETER	LANDUSE	ACCESSED_VAL	CITY
20100400020000	6474154.35276	10145.96973	HMAJCG	254316	SACRAMENTO
20100400010000	7076794.10172	10644.47635	HFAJAG	152800	SACRAMENTO
20100300200000	12993367.28984	15307.50117	HFAJAG	78013	ELVERTA
20100400030000	2942042.70203	7688.52193	HFAJAG	301890	SACRAMENTO
20100300190000	102725.86950	5216.30257	WBAC0A	298056	SACRAMENTO
20101000140000	78669.05521	2742.30260	WGAC0A	315000	ELVERTA
20100300170000	208715.26064	5238.44336	MROADA	295000	SACRAMENTO
20100300180000	12711260.26959	15068.44010	HFAJAG	110000	ELVERTA
20100200150000	8530649.18776	11583.31722	HFAJAG	276000	SACRAMENTO
20100200200000	2534604.48019	7728.17055	HFAJAG	192300	ELVERTA

AREA > 2000000 AND LANDUSE = HFAJAG NOT CITY = ELVERTA					
APN	AREA	PERIMETER	LANDUSE	ACCESSED_VAL	CITY
20100400020000	6474154.35276	10145.96973	HMAJCG	254316	SACRAMENTO
20100400010000	7076794.10172	10644.47635	HFAJAG	152800	SACRAMENTO
20100300200000	12993367.28984	15307.50117	HFAJAG	78013	ELVERTA
20100400030000	2942042.70203	7688.52193	HFAJAG	301890	SACRAMENTO
20100300190000	102725.86950	5216.30257	WBAC0A	298056	SACRAMENTO
20101000140000	78669.05521	2742.30260	WGAC0A	315000	ELVERTA
20100300170000	208715.26064	5238.44336	MROADA	295000	SACRAMENTO
20100300180000	12711260.26959	15068.44010	HFAJAG	110000	ELVERTA
20100200150000	8530649.18776	11583.31722	HFAJAG	276000	SACRAMENTO
20100200200000	2534604.48019	7728.17055	HFAJAG	192300	ELVERTA

Figure 4.9: Select records based on their attributes by using SQL expressions.

Once the records are selected, you can work with just those records. This is helpful for viewing, sorting, editing, calculating fields, generating statistics, using the selected features to select features in another GIS layer, creating a new layer with only the selected features, and isolating specific records to perform analysis functions on (like buffering selected features).

In addition, spatial queries, selecting features based on their geographic location (see Chapter 5), can be combined with attribute queries for more sophisticated queries. There is more on attribute and spatial queries in Chapter 5.

Calculate Attributes

Within an open data file, you can create new attributes by using values in existing fields, mathematical expressions, and text functions (see Figure 4.10). Mathematical operations allow you to add, subtract, multiply, and divide existing fields or values to create new, derived attributes. Text functions allow you to populate fields with data, copy values from one field to another, concatenate fields (and or values), truncate attributes, and convert text to different formats. Before calculating the new field, however, you need to create a new attribute field, which includes defining its field name and its data properties). Calculations can

be performed on a single record, several selected records, or on every record in the data file. The calculate function can also be used to copy data from one field to another.

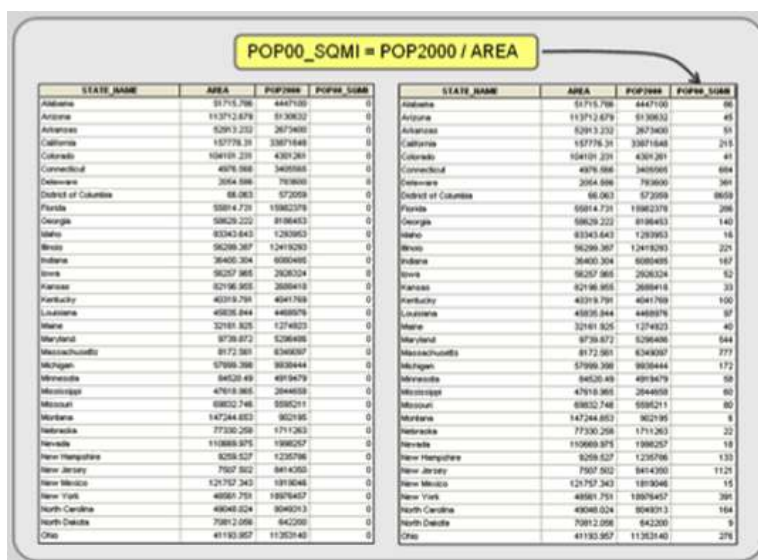


Figure 4.10: Calculating fields. In this example, population density is calculated by dividing population by area. First, the field must be added. Then, you calculate the results directly into the new field.

Geocoding

There is a way to create geographic data directly from attribute data. The process, called geocoding, assigns geographic locations to features directly from attribute fields that contain locational information within a data file. This is a popular way to create GIS feature layers; you create or obtain a spreadsheet or data file with location information, open the attribute table in your GIS, and direct the system toward the appropriate attribute fields. There are two types of geocoding: coordinate locations and address matching.

Spatial features can be created from data files containing fields with x,y coordinate values. The coordinates need to be separated into two separate fields: one for the x coordinate and one for the y coordinate. The process is straightforward; you direct the GIS to the data file's appropriate x,y fields, and it creates a spatial layer of point features from the coordinates. One possible complication is that the data file's coordinates are different than the coordinate system you are using. This requires that you open the file in a temporary workspace registered to the data file's coordinate system and then convert the new spatial layer to the desired coordinate system.

Address matching is another type of geocoding. It matches records in two data files—one containing a list of addresses and the other having street network attributes—to create a new layer (see Figure 4.11). In other words, it creates a layer of point features alongside street segments when addresses in the two data files match. It essentially looks up the address in the first record of the external data file and tries to find a match along the street network layer. If multiple possibilities exist, the routine will present them for user input. After the first record is matched or not, it moves to the second record and tries again. The resultant file is assigned the street network's coordinate system.

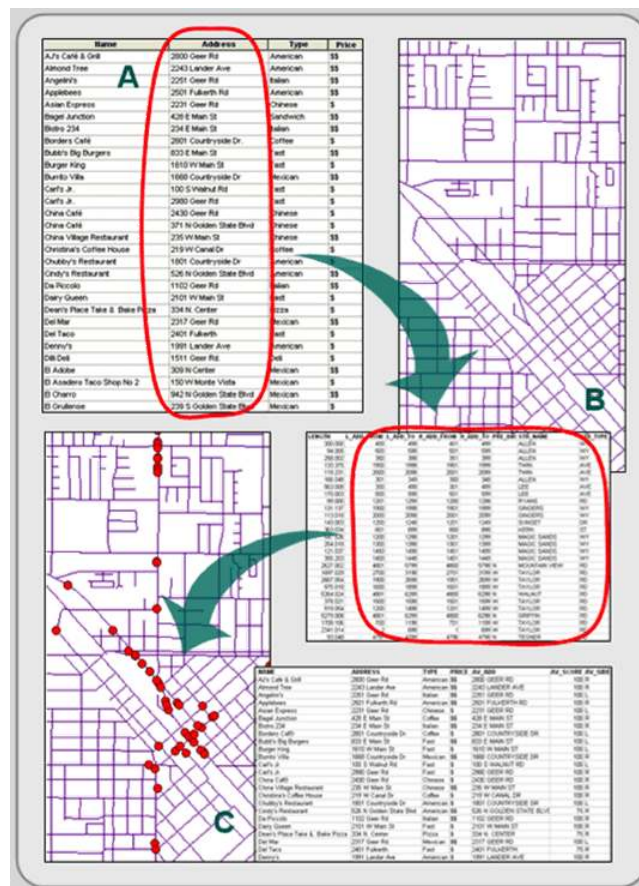


Figure 4.11: Address matching. The addresses in an external data file (A) are compared to a street network's (B) attribute fields, and if a match is made, the record in the external data file gets a point on the map (C).

Both the street network layer and the external data file need address data (street name, street type, and an address range for start and end of each line segment), and perhaps even more information like city, state, and Zip code attributes to make your address information unique (multiple cities will likely contain streets with the same name). The process works well if the addresses in both the external data file and the street network layer are accurate and complete, but address matching is a time consuming process.

Data Export

Exporting your GIS layers, including their geographic and attribute data files, are covered in Chapter 6. Most GIS programs can export your layer's attribute file in a number of formats including dBase and ASCII. The exported files can then be used in database, statistic, and spreadsheet programs for additional analysis.

This page titled [4.5: Attribute Preprocessing And Editing](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

Instead, the layer's data file can be displayed and sorted by each attribute in ascending order to identify missing attributes (see Figure 4.12). Map features missing a value for a particular field are revealed at the top of the table. Selecting those features from the data file and highlighting them on the screen can be a handy way to reference those features that have missing attributes. The selected features can then be investigated and updated. You can also sort the attributes alphabetically and glance down the field looking for spelling mistakes.

EMPLOYER_B	EMPLOYER_A	CITY	ZIP
	2100 STANFORD AVE SUITE #14	MCCLESTON	90390
	4001 E VENTNORE	CERES	90307
	2000 STANFORD	MCCLESTON	
	2001 ROMANS	MCCLESTON	90390
	100 STANFORD AVENUE #101	MCCLESTON	
	1401 J STREET	MCCLESTON	
	1717 SYLVAN AVE	MCCLESTON	90358
	1100 GARGALE RD	MCCLESTON	90305
	14 YOSEMITE BLVD	MCCLESTON	90307
	303 E HATCH ROAD	MCCLESTON	90301
	804 14TH STREET	MCCLESTON	90354
	428 14TH ST	MCCLESTON	90355
	125 S. EMERALD	MCCLESTON	90351
	3920 MC HENRY AVE	MCCLESTON	90390
CONTINENTAL PET SUPPLIES BARRETT BUSINESS SERVICES, INC.			
"VOUGHN WITH PETSON" PER HEADLER			
AHL PRODUCTS, INC.			
A.W.A.P. OF AMERICA "ZAPCO"			
ACE LATHING, INC.			
ADVANCED ROOF SAVERS TECH, INC.			
ALL VALLEY PAVING, INC.			
ALL VALLEY POULTRY AND LIVESTOCK EQUIP., INC.			
ALUMA PANEL, INC.			
AMERICAN PESTICIDE STATIONS, INC.			
AMERICAN CUSTOM FOAM, INC.			
AMERICAN MEATS, INC.			
AMERICAN MEDICAL RESPONSE, INC.			
AMERINE SYSTEMS, INC.			
ANGEL R WHITFIELD FARM LABOR			
ART TOYWEAR, INC.			
ASSOC. CLEANING SVC., INC.			
ATTARCO MOTOR, INC.			
AUTO PARTS DISTRIBUTORS, INC.			
AUTO PARTS DISTRIBUTORS, INC.			

More difficult to detect are incorrect attribute values. They require familiarity with the original source maps and an understanding of spatial patterns. For example, if you were working with income data, you should select low income values and display them on a map. Does their spatial location make sense? Do the same with high income values. Nominal data sets can be displayed the same way. For example, select different land use classes, and see if they make geographic sense. Display all heavy industrial sites and look at their locations. If heavy industry appears in the middle of wealthy residential areas or they are not located along highways, railroads, or rivers (which they need for transportation purposes) than these values may be inaccurate. More information may be needed; try looking at web-based aerial photographs or field check odd values.

This page titled [4.6: Attribute Verification](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

5: Analysis

[5.1: Introduction](#)

[5.2: Selection and Measurement](#)

[5.3: Overlay Analysis](#)

[5.4: Neighborhood Operations](#)

[5.5: Connectivity Analyses](#)

This page titled [5: Analysis](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

5.1: Introduction

Once the data input process is complete and your GIS layers are preprocessed, you can begin the analysis stage. Analyzing geographic data requires critical thinking and reasoning. You look for patterns, associations, connections, interactions, and evidence of change through time and over space. GIS helps you analyze the data sets and test for spatial relationships, but it does not replace the necessity for you to think spatially. First, you must conceive of the possible spatial patterns and relationships. This chapter provides an overview of the most common analysis functions and continues the spatial analysis discussion started in Chapter 1.

By integrating GIS layers, you can ask the spatial questions outlined in Chapter 1: “What is at...?”, “Where is it...?”, “What has changed since?”, “What spatial patterns exist?”, and “What if...?” (the scenario question). The first two of these questions inventory features and minimally examine feature location and relationships. The last three questions are more complex. To answer these questions, you must use or string together some of the analytical functions that you will learn about in this chapter. The particular analytical functions—and their order—are up to you.

This chapter focuses on the GIS functions that assist you in analysis—that help you evaluate, estimate, predict, interpret, and understand spatial data. The chapter breaks down these functions into four analyses presented in Figure 5.1:



Figure 5.1: Analysis functions categories.

Many of the specific analyses contained in the above categories have multiple names for the same function. This chapter uses the most frequently used terms, but attempts to note commonly used synonyms. It also notes whether the analytical processes are vector-based, raster-based, or can be used with both data models.

This page titled [5.1: Introduction](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

Attribute Query (Boolean Selection)

NAME	ADDRESS	TYPE	PRICE
Aly's Cafe & Grill	2600 Glen Rd	American	\$5
Almond Tree	2244 Lyndale Ave	American	\$5
Angelito	2201 Glen Rd	Italian	\$5
Apollonia	2501 Plymouth Rd	American	\$5
Asian Express	2231 Glen Rd	Chinese	\$
Bacon's London	420 E Main St	Coffee	\$5
Bakery 24	234 E Main St	Bakery	\$5
Bonkers Cafe	2801 Countyville Dr	Coffee	\$
Burger's Big Burgers	333 E Main St	Fast	\$5
Burger King	1800 E Main St	Fast	\$
Burger Villa	1800 Countyville Dr	Mexican	\$5
Carli's	1100 E Main Rd	Fast	\$
Carli's Jr	2600 Glen Rd	Fast	\$
China Cafe	2430 Glen Rd	Chinese	\$
China Village Restaurant	236 W Main St	Chinese	\$5
Devotional Coffee House	2114 Carroll Dr	Coffee	\$
Chubby's Restaurant	1801 Countyville Dr	American	\$
Crusty's Restaurant	426 N Golden State Blvd	American	\$5
De Pizzeria	1130 Glen Rd	Italian	\$5
Dairy Queen	2125 W Main St	Fast	\$
Dean's Place Tacos & Bites Pizzas	334 N Center	Pizza	\$
DeMar	2217 Glen Rd	Mexican	\$5
De Tacos	2429 Plymouth	Fast	\$
DeVito's	1801 Lyndale Ave	American	\$
D&D Deli	1511 Glen Rd	Deli	\$
DJ Amigos	330 N Center	Mexican	\$5
El Accordion Taco Shop No. 2	110 W Main Villa	Mexican	\$
El Obrero	3842 N Golden State Blvd	Mexican	\$
El Orlandino	208 E Golden State Blvd	Mexican	\$
El Jardin	400 E Olive	Mexican	\$
El Rose No. 3	3031 N Golden State Blvd	Mexican	\$
Evans Restaurant and Lounge	4402 N Golden State Blvd	American	\$5
Fire Station Family Grill	130 E Main Rd	American	\$5
Godfather's Pizza	2900 Glen Rd	Pizza	\$5
Golden China	2525 Glen Rd	Chinese	\$5
Golden Dragon	2300 W Main St	Chinese	\$5
Golden Hour	410 E Main St	Chinese	\$
Henry's	217 E Golden State Blvd	Eastern	\$5
Himalayan Buffet	1800 Countyville Dr	American	\$5
House Of Aves	2620 Glen Rd	Coffee	\$
In-and-Out Burger	3071 Countyville Dr	Fast	\$
International House of Pancakes	1525 Countyville Dr	American	\$5
Jack-In-The-Box	2601 Glen Rd	Fast	\$
Jack-In-The-Box	100 N Golden State	Fast	\$
Jack-In-The-Box	1801 Lyndale Ave	Fast	\$
Jewelry	1110 E Main Villa Ave	Mexican	\$5
Jerry's Pizzeria	144 N Golden State Blvd	Pizza	\$5
Jerry's Steak House	136 E Golden State Blvd	Steak	\$5
Joakim's Pizzeria	2640 Glen Rd	Eastern	\$5
Johnnyboy Fried Chicken	400 N Golden State Blvd	Fast	\$
Kawasaki	1110 W Main St	Mexican	\$5
La Crepe	201 E Main St	American	\$5
La Montaña	326 E Golden State Blvd	Mexican	\$5
La Panchera	1134 W Main St	Mexican	\$5

Spatial Selection (Spatial Searches)

<https://geo.libretexts.org/@go/page/44919>

an interactively drawn graphic polygon). Figure 5.3 is an example that uses the same restaurant layer as the previous figure. Again, the first layer consists of restaurants, some of which one wants to select. The second layer is composed of polygons radiating out from points of interest. After the selection process, the customers falling within the overlaying polygons are highlighted (selected). Ninety out of 112 restaurants fit the query expression and are selected within the attribute file and simultaneously on the map.

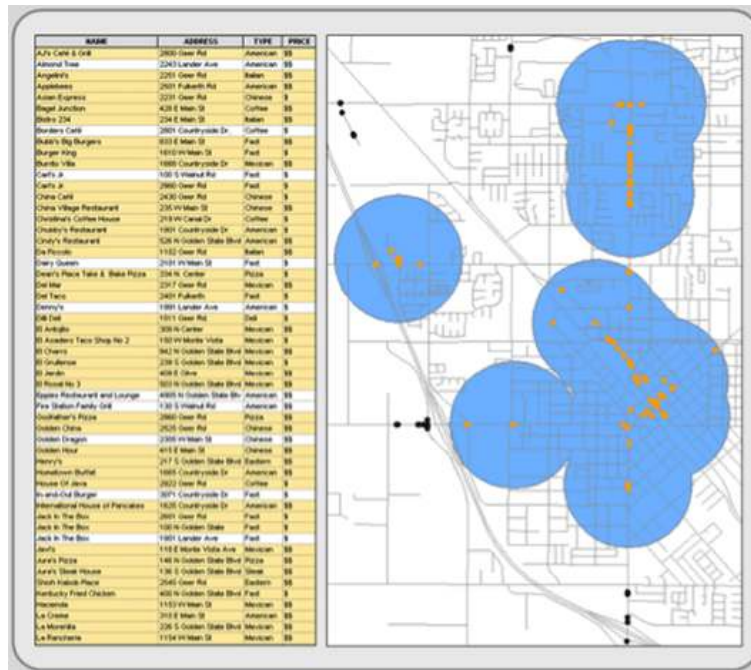


Figure 5.3: Spatial selection. Only those restaurants that fall within the blue polygons are selected.

There are many types of spatial selection. Point in polygon, perhaps the most used, selects the points of one layer if they are contained within a selected polygon (or polygons) of a different layer (or graphic). Line in Polygon, a similar operation, selects line features that are wholly or partially contained within a different layer's polygon. Polygon in Polygon is another variation that selects polygon features within (or overlapping) selected polygons from a second layer. Another type of spatial selection is point distance (which has line and polygon versions too), which identifies all the points in one layer that are within a specified distance of a selected point(s) in a different layer. Like any type of selection, you can perform analytical processes on those highlighted features or save them to a new layer.

You can mix spatial selections and attribute queries. Here you might spatially select features first, and then from the feature's attribute file, you would reduce (or alternatively increase) the selected records through attribute query expressions. Figure 5.4 is an example using the same restaurant data as above. Thirty seven out of 112 restaurants fit the query expression and fall within the overlaying polygons. They are highlighted both within the attribute file and on the map.

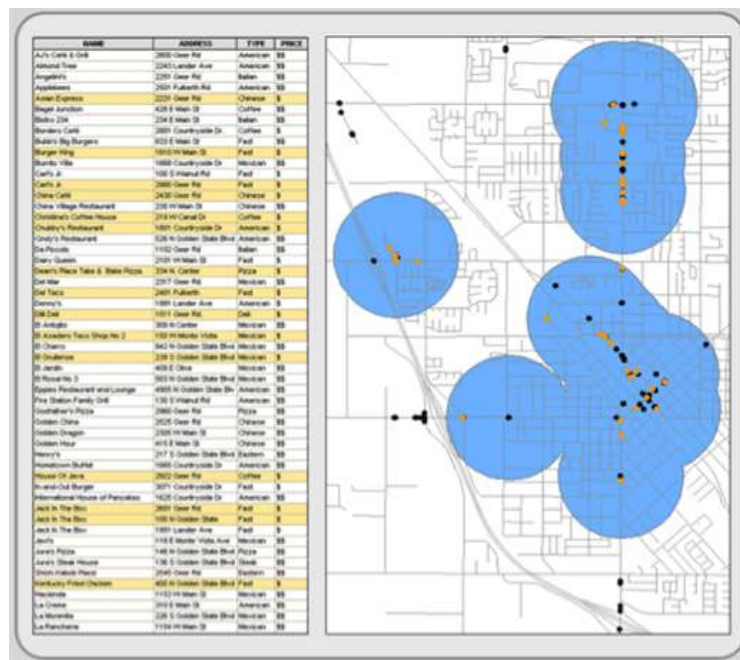


Figure 5.4: Spatial and attribute selection combined. In this example, restaurants that fall within the blue polygons and are inexpensive (PRICE = \$) are highlighted in orange.

Spatial selection is a vector process, but when you combine reclassification (discussed in Chapter 3) and overlay (discussed later in this chapter), the combination produces a raster version of spatial selection.

Measuring Distance

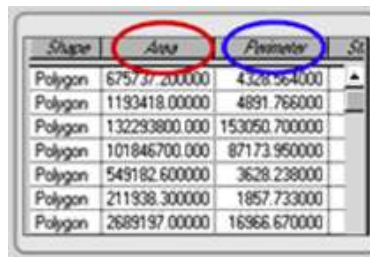
There are many ways to measure distance. Most GIS programs, both raster and vector, have a ruler button that allows you to measure distances across a map. After clicking the button, you point on the map where you want to begin your distance measurement and then click at the ending point (or intervening points that define the path you want to measure).

Many vector-based systems measure distances along existing vector line networks, like streets, sewers, and railroads. This type of distance measurement relies on topological network relationships, which are discussed later (see Connectivity Analysis). In addition, some vector systems automatically generate length measurements for line features as you enter them. They store the length result in an attribute field within the layer's data file. Those systems that do not have this automatic function usually provide a way for you to calculate line feature length and store the result in an attribute field that you define. Once calculated and stored, you can sum the length of multiple line features by selecting them and calculating their sum (see Calculating Descriptive Statistics below).

Raster-based systems allow you to generate distance measurements in all directions away from a selected pixel or group of pixels. These distances are placed in a new layer where each cell's value represents the distance from that cell to the nearest selected pixel. These "distance" layers are often used for spread functions (see Spread Functions below).

Measuring Area/Perimeter

Many vector systems automatically generate area and perimeter measurements for polygon features and store these values in prescribed attribute fields. Those systems that do not have this automatic function do provide a way for you to generate area and perimeter and store the results in user-defined fields. See Figure 5.5 for an example. Once calculated and stored, you can select multiple polygon features and sum their area and perimeter (see Calculating Descriptive Statistics below).



Shape	Area	Perimeter	SD
Polygon	675737.200000	4328564.000	
Polygon	1193418.000000	4891.766000	
Polygon	132293800.000	153050.700000	
Polygon	101846700.000	87173.950000	
Polygon	549182.600000	3628.238000	
Polygon	211938.300000	1857.733000	
Polygon	2689197.000000	16966.670000	

Figure 5.5: Area and perimeter contained as attributes in the layer's data file.

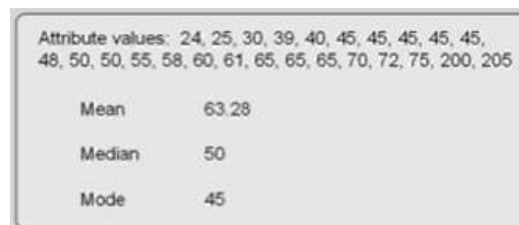
Calculating areas and perimeters are done differently in raster systems. Instead of measuring and storing each polygon's area and perimeter in the feature's pixels, raster systems already know the size—the area covered—by a single pixel. To calculate area, it simply adds up the number of pixels with a specified attribute and multiplies the count by the area contained in a single pixel. It is easy math. For example, your layer might have 100 polygons that possess one of twelve land cover categories. The routine finds each occurrence of the twelve categories (even if they are not contiguous) and sums the category's area and perimeter. Perimeter is usually equally easy if the pixels are square, and in the vast majority of cases they are. These measurements are provided either in standard tables or in new layers where the pixels exhibit the sums of the area and perimeter of the category to which it originally belonged.

Calculating Descriptive Statistics

Descriptive statistics summarize attribute data. They reduce the complexities of numerous individual values into a few meaningful numbers that describe the individual features collectively. Descriptive statistics are organized into two groups: measures of central tendency and measures of dispersion.

Central tendency describes the center of the attribute data's distribution. The mean, median, and mode are its three common measures, but which measure you should use depends largely on the attribute's level of measurement (described in Chapter 2). Figure 5.6 depicts the three central tendency measures for the attribute values of a single field.

- The most used measure is the *mean* (commonly referred to as average), which is calculated by adding together each feature's attribute value and dividing the sum by the number of features. For example, if you wanted to characterize the age of the people reading this e-text, you would sum the age of each reader and divide it by the number of readers. The result is the mean. It—like all measures of central tendency—is a surrogate used to describe all the values within a single attribute field. This measure requires interval or ratio data.
- If we placed the attribute values in ascending or descending order, the *median* is the middle score in the distribution (this works for an odd number of cases). In other words, half of the attribute values are above and half are below this value. In an even numbered distribution, the median is the average of the two middle scores. Median is used for ordinal and derived (aggregated) data.
- *Mode* is the most frequent score in a distribution. Of course, some distributions do not have a mode if there are no repeated values. At times, the only repeated value might be at the low or high end of the distribution, making this measure a bit unreliable and certainly un-central. The measure, however, is helpful in describing leading categories (for instance the different political parties). It is the only measure for describing the central tendency of nominal data.



Attribute values: 24, 25, 30, 39, 40, 45, 45, 45, 45, 48, 50, 50, 55, 58, 60, 61, 65, 65, 65, 70, 72, 75, 200, 205	
Mean	63.28
Median	50
Mode	45

Figure 5.6: Central tendency measures. These attribute values attained from a vector layer's attribute field or from selected pixels in a raster layer.

Dispersion, the second group of descriptive statistics, looks at the attribute data's spread. Its measures (including range, variance, and standard deviation) describe how much the attribute values vary around the distribution's center (its central tendency measures). Are the values clustered tightly or are they spread out? These measures help you judge how well the central tendency characterizes all of the values in the attribute field. If the measure of dispersion is small, the values are clustered and the central tendency measure describes the distribution well. There are several types of measures of dispersion (also see Figure 5.7):

- *Counts* and *frequencies* are not measures of dispersion, but they are basic ways to summarize data. Counting simply denotes quantity. Frequency is the number of times an attribute field has a particular value. A frequency distribution, usually in the form of a histogram, describes the shape (or structure) of the attribute data by tabulating the frequencies of each value (or range of values).
- *Range* is the distance between the minimum and maximum attribute values. To derive it, simply subtract the minimum value from the maximum value. It is the simplest measure of dispersion, but it is vulnerable to outliers (rogue values that are significantly different from the rest of the attribute values). If you think outliers affect the range, use the interquartile range instead. It divides the distribution, arranged from low to high, into four parts each containing 25 percent of the attribute values, and it is the difference between the 25th & 75th percentiles.
- The *variance* looks at the difference between the distribution's values and its central tendency measure (in this case the mean). It is more complex than computing the average difference that each attribute value falls from the mean. Such a score does not provide enough numeric emphasis to the attribute values on the low and high end of the distribution. The variance adjusts for this by squaring the difference, summing the squares, and dividing by the count.
- *Standard deviation* is the square root of the variance. Like the variance, it describes the dispersion around the mean and allows you to evaluate how closely the numbers in the dataset are packed around the mean (in other words, how well the mean describes or summarizes the set of numbers). Similarly, the smaller the number the tighter the values are clustered around the central tendency measure. Unlike the variance's higher values, however, standard deviation uses numbers that are similar to the original dataset. Still, the two are essentially the same thing.

Attribute values: 24, 25, 30, 39, 40, 45, 45, 45, 45, 45, 48, 50, 50, 55, 58, 60, 61, 65, 65, 65, 70, 72, 75, 200, 205	
Count	25
Range	205 - 24 = 181
Variance	1866.4416
Std. Dev.	43.2023

Figure 5.7: Measures of Dispersion.

In vector systems, descriptive statistics are usually generated within the attribute file interface. In raster layers, menu commands process the descriptive statistics. Each attribute field can be summarized in its entirety or confined to selected records or pixels. Which descriptive statistics are calculated depends on the attribute data's level of measurement (Figure 5.8).

Type of descriptive statistic	Level of measurement			
	Nominal	Ordinal	Interval	Ratio
Mode, counts, frequency	Yes	Yes	Yes	Yes
Median, minimum, maximum, range.	No	Yes	Yes	Yes
Mean, variance, standard deviation.	No	No	Yes	Yes

Figure 5.8: Depicts which descriptive statistics can be used with different data levels.

This page titled [5.2: Selection and Measurement](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

5.3: Overlay Analysis

Overlay is one of the most common and powerful GIS functions. It investigates the spatial association of features by “vertically stacking” feature layers to investigate geographic patterns and determine locations that meet specific criteria.

It is the best-known GIS function, but examples of overlay predate computers and GIS. A simple but powerful example was described in Chapter 1: Dr. Snow overlaid water pumps over cholera deaths to see a spatial pattern and infer a connection between water and the disease. Other, more sophisticated, overlay analyses also occurred before the advent of GIS. In the 1960s, Ian McHarg sought a better way to plan land use, taking into account the physical environment and human factors. In *Design with Nature* (1969), McHarg formalized his site planning process based on overlay transparencies. He created hard-copy transparent maps for each relevant human (historic values, scenic vistas, social costs, etc.) and physical (slope, surface drainage, riparian areas, susceptibility to erosion, etc) factor. Each transparency included shades ranging from dark tones (areas with high values) to light tones (areas with low values). Physically, the transparencies were superimposed upon each other over the study area’s base map. A composite map revealed dark tones over areas where multiple layers had high values (high impacts) and light tones in regions with low impact values. McHarg felt that planners needed to undertake this process to determine which areas should be left natural and which places were suitable for development. His book and method were so popular that many of the first GIS projects attempted to formalize his technique using GIS.

Today, there are many types of GIS overlay. Vector and raster models both perform overlay, but their overlay functions differ considerably and thus will be discussed separately.

Vector (Logical) Overlay

Vector overlay predominantly overlays polygons in one layer over polygons in another layer, but it can also be used to overlay point or line features over polygon layers. Sometimes referred to as topological or logical overlay, it is conceptually and mathematically more demanding than raster overlay. There are three types of vector overlay operations:

Polygon on polygon is where one polygon layer is superimposed over another polygon layer to create a new output polygon layer. The resultant polygons may contain some or all of the attributes from the polygons in which they were created. Several types of polygon on polygon overlay exist, including intersection (A and B), union (A or B), and clip (A not B). These Boolean operators work both on the attribute table and the geography.

- *Intersection* computes the geometric intersection of all of the polygons in the input layers (see Figure 5.9). Only those features that share a common geography are preserved in the output layer. Any polygon or portion of a polygon that falls outside of the common area is discarded from the output layer. The new polygon layer can possess the attribute data of the features in the input layers.

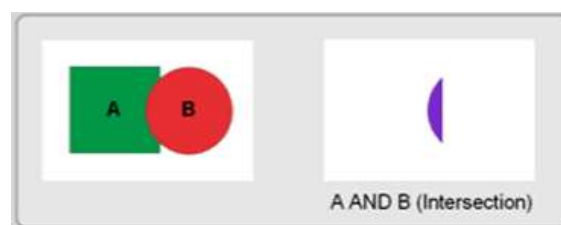


Figure 5.9: Intersection of two layers.

- *Union* combines the features of input polygon layers (see Figure 5.10). All polygons from the input layers are included in the output polygon layer. It can also possess the combined attribute data of the input polygon layers.

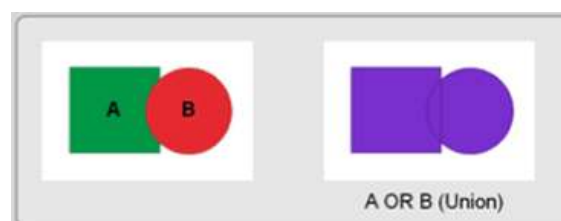


Figure 5.10: Union of two layers.

- *Clip* removes those features (or portions of features) from an input polygon layer that overlay with features from a clip polygon layer (Figure 5.11). The clip layer acts as a cookie cutter to remove features (and portions of features) that fall inside the clip layer.

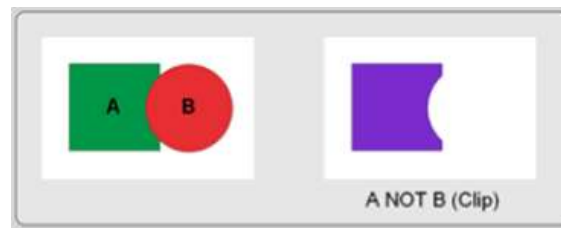


Figure 5.11: Clipping one layer from another.

Point in polygon is where a layer of point features is superimposed over a layer of polygon features. The two layers produce a point layer that includes attributes from the surrounding input layer polygons (Figure 5.12). Alternatively, you can tally the number of point features falling within each polygon and store the sum as a new attribute in the polygon layer. Other point attributes can be aggregated (summed, averaged, etc.) and included as attributes in the polygon's data file. The transferring of attributes based on their geographic position is called a spatial join.

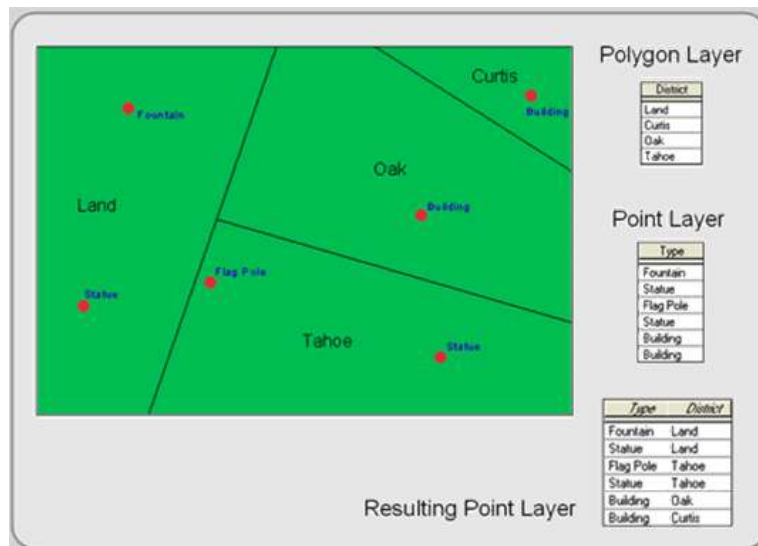


Figure 5.12: Point in Polygon.

Line on polygon is similar to point in polygon, but lines are superimposed on polygons. This type of spatial join either appends polygon attributes to line features falling within them or counts and aggregates line attribute data to the polygon layer's data file.

Raster (Arithmetic) Overlay

Raster overlay superimposes at least two input raster layers to produce an output layer. Each cell in the output layer is calculated from the corresponding pixels in the input layers. To do this, the layers must line up perfectly; they must have the same pixel resolution and spatial extent. If they do not align, they can be adjusted to fit by the preprocessing functions discussed in Chapter 3. Once preprocessed, raster overlay is flexible, efficient, quick, and offers more overlay possibilities than vector overlay.

Raster overlay, frequently called map algebra, is based on calculations which include arithmetic expressions and set and Boolean algebraic operators to process the input layers to create an output layer. The most common operators are addition, subtraction, multiplication, and division, but other popular operators include maximum, minimum, average, AND, OR, and NOT. In short, raster overlay simply uses arithmetic operators to compute the corresponding cells of two or more input layers together, uses Boolean algebra like AND or OR to find the pixels that fit a particular query statement, or executes statistical tests like correlation and regression on the input layers (see Figure 5.13).

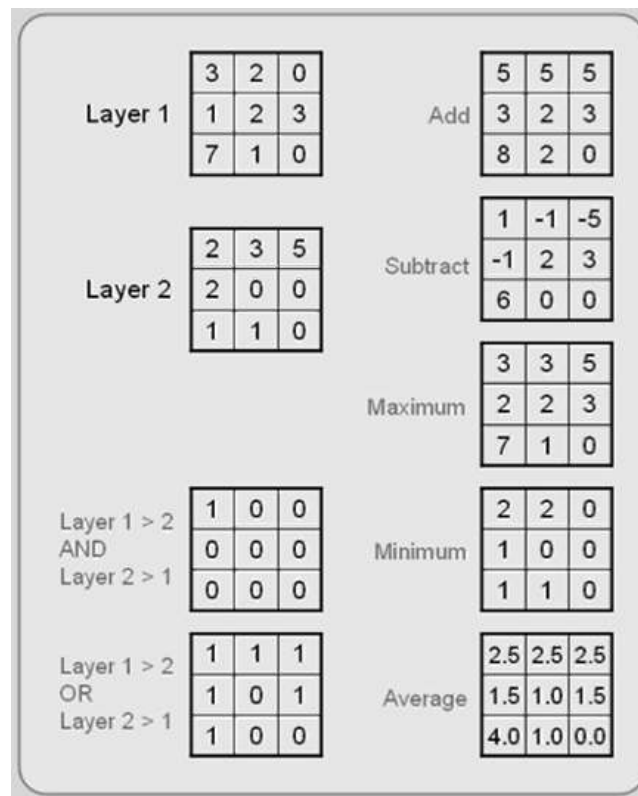


Figure 5.13: Raster Overlay. Using layers 1 and 2, all sorts of overlay are possible.

Correlation and Regression

Correlation and Regression are two ways to compute the degree of association between two (or sometimes more) layers. With correlation, you do not assume a causal relationship. In other words, one layer is not affecting the spatial pattern of the other layer. The patterns may be similar, but no cause and effect is implied.

Regression is different; you make the assumption that one layer (and its variable) influences the other. You specify an independent variable layer (sometimes more than one) that affects the dependent variable layer. Figure 5.14 shows a precipitation (dependent) and elevation (independent) as the layers.

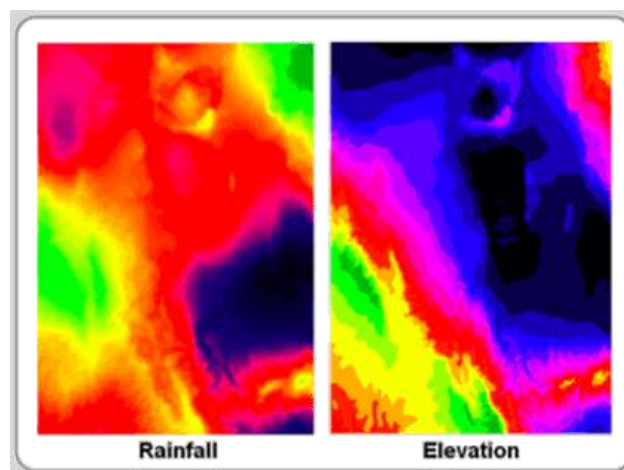


Figure 5.14: Is there a spatial relationship between these two layers? Correlation and regression tests allow you to overlay layers to test their spatial relationship.

With both statistical tests, you compute a correlation coefficient, which ranges from -1 to +1. Positive coefficients indicate that the two layer's variables are associated in the same direction. As one variable increases, the other variable increases (both can simultaneously decrease too). The values closer to +1 describe a stronger association than those closer to zero. A negative coefficient depicts two layer's variables that are associated but in opposite directions. As one variable increases, the other variable decreases. Values closer to -1 have a strong negative association. If the correlation coefficient is near zero, there is little to no association. Both of these processes are raster based.

This page titled [5.3: Overlay Analysis](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

5.4: Neighborhood Operations

Neighborhood operations, also called proximity analyses, consider the characteristics of neighboring areas around a specific location. These functions either modify existing features or create new feature layers, which are influenced, to some degree, by the distance from existing features. All GIS programs provide some neighborhood analyses, which include buffering, interpolation, Thiessen polygons, and various topographic functions.

Buffering

Buffering creates physical zones around features. These “buffers” are usually based on specific straight-line distances from selected features (like in Figure 5.15). Buffers, common to both raster and vector systems, are created around point, line, or polygon features. The resulting buffers are placed in an output polygon feature layer. Once complete, buffer layers are used to determine which features (in other layers) occur either within or outside the buffers (spatial queries), to perform overlay, or to measure the area of the buffer zone. They are the most used neighborhood operation.



Figure 5.15: Buffering around a selected line feature.

Most buffers use constant widths to generate zones, but some buffers have variable widths that depend on feature attribute values. As an example, the figure to the right depicts a variable distance buffer based on decibels from an airport.

Interpolation

Interpolation is a method of predicting or estimating pixel values at unsampled locations based on the known values of neighboring pixels (see Figure 5.16). Since it is impractical for you to take measurements at all locations across your study area due to money, time, legal, and physical constraints, you interpolate between known pixel values (sampled locations). With interpolation, you create a continuous surface like elevation, temperature, and soil characteristics that occur everywhere. Because of its continuous nature, interpolation is only available within raster-based systems.

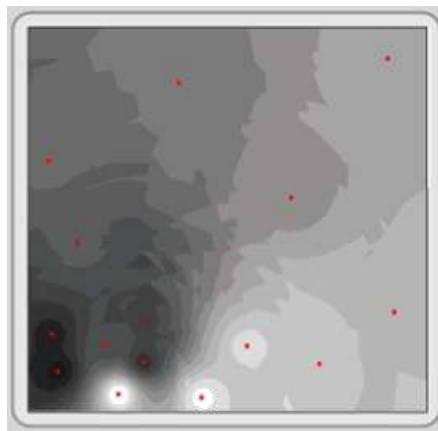


Figure 5.16: Interpolating between point features. The red dots are the points where values are known. The gray cells are the estimated data based on the known values.

There are many different types of interpolation: Linear interpolation, the simplest form, assumes that the value change over distance from recorded pixels is uniform. In other words, the value change per pixel is constant between two known points. This interpolation method is not always appropriate, so there are other methods including Fixed-radius Local Averaging, Inverse Distance Weighted, Trend Surface, Splines, and Kriging. All of these interpolation methods look at the values of the recorded pixels to generate the value of the pixels that fall in between. The methods differ in how they weigh the recorded attributes and in the number of observations used for each method. No method is accurate in every situation.

Theissen polygons (voronoi or proximal polygons)

Theissen polygons are boundaries created around points within a point layer (see Figure 5.17). The resultant polygons form around each of the points, and they delineate territories around which any location inside the polygon is closer to the internal point (that created it) than to any other point in the layer. Attributes associated with each point are assigned to the resultant polygon. It is a vector and raster process, but for more than one attribute, raster systems must use multiple layers.



Figure 5.17: Creating Theissen polygons from point features.

Topographic Functions

Topographic functions use Digital Elevation Models (DEMs) to illustrate the lay of the land. DEMs are raster layers containing elevation data in each pixel. From these values, you produce output layers to portray slope (inclination), aspect (direction), and hillshading (see Figure 5.18). These topographic functions are typical neighborhood processes; each pixel in the resultant layer is a product of its own elevation value as well as those of its surrounding neighbors.

- *Slope* layers exhibit the incline or steepness of the land. It is the change in elevation over a defined distance.
- *Aspect* is the compass direction in which a slope faces. From north, it is usually expressed clockwise from 0 to 360 degrees.
- *Hillshading*, which is cartographically called shaded relief, is a lighting effect which mimics the sun to highlight hills and valleys. Some areas appear to be illuminated while others lie in shadows.

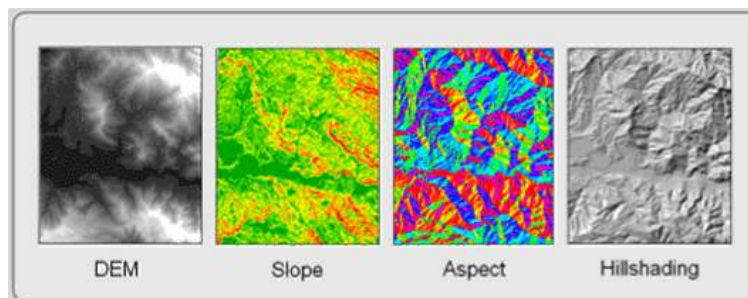


Figure 5.18: Topographic Functions. The DEM creates the slope, aspect, and hillshading layers.

While these functions are raster processes, most can be mimicked in a vector environment by Triangulated Irregular Networks (TIN). In addition, topographic functions can derive vector isolines (contours).

This page titled [5.4: Neighborhood Operations](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

5.5: Connectivity Analyses

Connectivity analyses use functions that accumulate values over an area traveled. Most often, these include the analysis of surfaces and networks. Connectivity analyses include network analysis, spread functions, and visibility analysis. This group of analytical functions is the least developed in commercial GIS software, but this situation is changing as commercial demand for these functions is increasing.

Vector-based systems generally focus on network analysis capabilities. Raster-based systems provide visibility analysis and sophisticated spread function capabilities.

Spread Functions (Surface Analysis)

Spread functions are raster analysis techniques that determine paths through space by considering how phenomena (including features) spread over an area in all directions but with different resistances. You begin with an origin or starting layer (a point where the path begins) and a friction layer, which represents how difficult—how much resistance—it is for the phenomenon to pass through each cell. From these two layers, a new layer is formed that indicates how much resistance the phenomenon encounters as it spreads in all directions (see Figure 5.19).

Add a destination layer, and you can determine the “least cost” path between the origin and the destination. “Least cost” can be a monetary cost, but it can also represent the time it takes to go from one point to another, the environmental cost of using a route, or even the amount of effort (calories) that is spent.

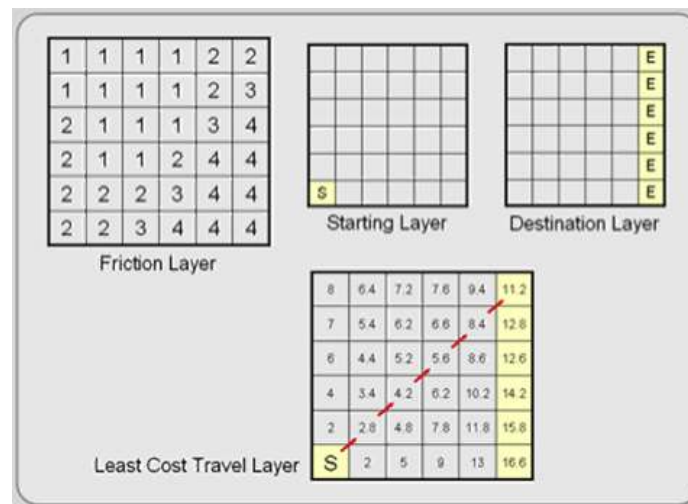


Figure 5.19: Spread Functions. This example shows that the shortest distance is not always the least cost distance.

Viewshed Modeling (Intervisibility Analysis)

Viewshed modeling uses elevation layers to indicate areas on the map that can and cannot be viewed from a specific vantage point. The non-obscured area is the viewshed. Viewsheds are developed from DEMs in raster-based systems and from TINs in vector systems. The ability to determine viewshed (and how they can be altered) is particularly useful to national and state park planners and landscape architects. Figure 5.20 depicts the areas within a park where a proposed radio antenna can be seen.

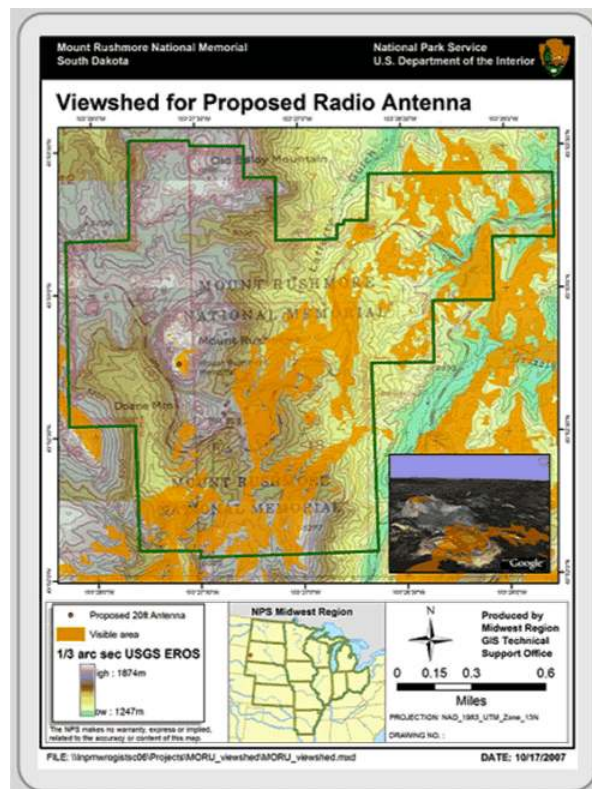


Figure 5.20: Viewshed Analysis. Map courtesy of the National Park Service, Department of Interior, 2007.

Network Analysis

Network analyses involve analyzing the flow of networks—a connected set of lines and point nodes (sometimes called centers or hubs). These linear networks most often represent features such as rivers, transportation corridors (roads, railroads, and even flight paths), and utilities (electric, telephone, television, sewer, water, gas). Point nodes usually represent pickup or destination sites, clients, transformers, valves, and intersections. People, water, consumer packages, kilowatts, and many other resources flow to and from nodes along linear features.

Each linear feature affects the resource flow. For example, a street segment might only provide flow in one direction (a one-way street) and at a certain speed. Nodes can also affect flow. A stuck valve might allow too much of a resource to stream out and away from its intended destination. Network analysis tools help you analyze the “cost” of moving through the network. Like spread functions, “cost” can represent money, time, distance, or effort. Network analyses are vector-based applications, but there are similarities with raster-based spread functions.

The three major types of network analyses include route selection (optimal path or shortest path), resource allocation, and network modeling.

- *Route Selection* attempts to identify the least “cost” route. As described above, cost can be defined a number of ways. You might want to find the shortest path between your home and a weekend destination or the least costly route that delivers UPS packages to their recipients. In any route selection routine, two or more nodes, including an origin and a destination point, must be identified and be able to be visited on the network. Sometimes there are a large number of possible routes. It is the job of the network analysis algorithm to determine the least cost route. Multiple paths are tested until the least cost path connects the starting and destination points.
- *Resource Allocation*, the second major type of network analysis, involves the apportionment of a network to nodes. To do this, you define one or more allocation nodes on the network. Territories of linear features, like streets, are defined around each of these allocation nodes. The linear features are usually assigned to the nearest node, where distance is measured in time, length, money, or effort. Figure 5.21 depicts 4-minute response times from six fire stations and three potential fire station locations. The polygon that is drawn around each station (triangle) represents the area that can be covered in 4 minutes.



Figure 5.21: Resource allocation. Map is courtesy of Tyler Schrag, Bellingham Fire Department, 2006.

- Network Modeling uses interconnected linear features and point nodes to analyze how resources travel through networks. The linear features, like streets or river channels, have attributes that might define travel speed, number of lanes, and volume of flow. Nodes also have attributes that might identify vehicular turns and the time or cost required for each turn. Resources like water or traffic are placed in the network and their movement modeled. This way, problems with the network load can be identified.

This page titled [5.5: Connectivity Analyses](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

6: Output

[6.1: Introduction](#)

[6.2: Map Communication](#)

[6.3: Map Types](#)

[6.4: Constructing Thematic Maps](#)

[6.5: Types Of Output](#)

This page titled [6: Output](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

6.1: Introduction

Mark Monmonnier's book, *How to Lie with Maps*, exposed how people and organizations overtly and intentionally misshaped map features to deceive. The concept that underlies his entire book is that all maps lie—to some extent. That statement is true. As you read in Chapter 3, projections introduce shape, area, distance, and directional distortions. In addition, all maps simplify the real world and thus lie by omission. Map symbols exaggerate and minimize the extent of features. These can all be considered lies. Now, before you stop reading and discard maps as fiction, remember that all models of reality (paintings, literature, statistics, and photographs) are abstractions and thus lie to some degree and that omission and abstraction can lead to greater communication and understanding. Simply put, maps aid in communication by emphasizing the location of features (and omitting many other features) to depict patterns across space. As Monmonnier states, "A good map tells a multitude of little white lies; it suppresses truth to help the user see what needs to be seen" (1996, p. 25).

Some mapmakers radically distort features to enhance communication, and they can do this legitimately. Perhaps the best example of manipulating the size and location of features is the London Underground map, which distorts subway line length and terminal location, but, in doing so, it enhances our spatial understanding of the subway system (see Figure 6.1). Harry Beck, the map's creator, devoted his life to this user-friendly map. He designed the map using only horizontal, vertical, and 45-degree lines that emphasized the relative position of terminals (connections) rather than adhering to strict spatial accuracy. The distances from suburban terminals were shortened to give more map room for central city terminals. He clearly understood that the traveling public perceived the network in terms of stops and transit line connections, not distance. The map communicated these principals clearly, and it has been copied by transit lines across the world.



Figure 6.1: London's Underground Map. Courtesy of London Regional Transport, 1991.

This chapter broadly focuses on GIS output and more narrowly on maps and their design. If accurate and clear communication is your goal, as it should be, you need to consider topics already discussed like projections and scale (Chapters 2 and 3) as well as subjects described in this chapter including symbolization, classification, generalization, and color.

This page titled [6.1: Introduction](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

6.2: Map Communication

The International Cartographic Association (ICA) defines a map as “a representation, normally to scale and on a flat medium, of a selection of material or abstract features on, or in relation to, the surface of the Earth.” In other words, maps are an approximation, a model, a summary of the real world.

Maps communicate; they represent and help us organize knowledge by representing a portion of the Earth’s surface. They are created for transmitting spatial information to a map reader, yet most maps are improperly designed and do not communicate easily nor effectively. This is not the fault of the map reader. The fault lies with the cartographer that makes the map. To design better maps, consider the cartographic communication process with its four stages (see Figure 6.2): 1) Real World, 2) Selection, 3) Generalization, and 4) Map.



Figure 6.2: Cartographic communication process.

Real World

As described in Chapter 1, the world is too complex for direct analysis and understanding, so we create models of the world by selecting and generalizing some of its features. Imagine, however, if we could record the world’s infinite detail on a map. Lewis Carroll, author of *Alice in Wonderland*, described such a detailed map in *Sylvie and Bruno Concluded*. In this fantasy, a Professor explains to another how his country’s cartographers experimented with ever-larger maps. The Professor states, “And then came the grandest idea of all! We actually made a map of the country, on the scale of *a mile to the mile!*” “Have you used it much?” the other person enquired.” No, says the professor, “It has never been spread out” ... “the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well.”

In reality, the “country itself” is a poor replacement for a map. If we could use the real world as our guide, then we would not need maps. Every map selects and generalizes the world’s features, and these little “white lies” help maps communicate.

Selection

Maps are selective. You determine what should be, and equally important, what should not be included on your map. If features do not aid your map’s purpose, nor orient the map reader, eliminate them. There must be a reason for the presence of every feature type. Guiding your selection process should be two essential considerations: the map’s purpose and its scale.

What is the purpose of your map? What are you trying to get across? Who is your audience? Addressing these questions helps you determine how much detail should be placed on your map. Selecting too many types of features obscures your map’s primary purpose.

Scale is the relationship between distances on the map (or screen) and corresponding distances in the real world. It is a major factor in determining which features are selected and which are omitted. Ask yourself, how is the map going to be presented to your audience? Will it be on an 8 ½” x 11” piece of paper, a 3” x 3” portion of a newspaper, or through a data projector onto a screen? The physical size of the presented map largely dictates the amount of detail that can be displayed on the map. The chosen scale affects not only the selection of features but also the degree of their generalization.

Generalization

Geographic data and detail are without limit. If you are flying high above a city, you will see certain features that define the city’s overall shape and its major neighborhoods. As you descend into a neighborhood, the homes, streets, parked cars, and sidewalks become clear. Descend into a backyard and you see a pool, a vegetable garden, chairs, and a redwood deck. Dogs and cats are

visible. Pull out a magnifying glass and investigate the redwood deck's grain to see its color, pits, splinters, and undulations. To capture all of the real world's features and their detail, you would need an infinitely large database and an infinite amount of time.

The features you select need to be generalized, but how much detail should they have? Map size being equal, large-scale maps that depict features in a small area can have more detail. Here are several generalizing tasks to make map reading easier and more effective:

1. *Smooth* features. For example, take some kinks out of a river or a road. Beck's London Underground map (Figure 6.1) smoothed and straightened subway routes, which made the network more intuitive to map readers.
2. *Abstract* features. Detail catches the map reader's eye. Abstraction removes detail. Remove the detail of a city's street to a single line. For example, represent the library's actual footprint with a black square.
3. *Aggregate* features. Some features may be lumped together to deemphasize them. For example, represent several school buildings with a single symbol.
4. *Exaggerate* features. While smoothing, abstraction, and aggregation seek to deemphasize features, exaggeration places greater emphasis on the feature. If it is important for the purpose of your map, enlarge the feature.
5. *Displace* features. Sometimes features need to be moved, perhaps slightly, to accentuate them and make the map more visually pleasing and intuitive. For example, Beck moved London's train stations.

The primary role of maps is to communicate, and this is impossible without selection and generalization. Still, people have difficulty reading maps. If you find maps easy to read, it is partly due to your familiarity with maps and their conventions. Once you understand these conventions, map reading becomes easier and reinforces that understanding.

Conventions, a form of abstraction, are signs and symbols that allow people to read maps. Cartographers rely on conventions for good cartographic communication. Some conventions are almost universally understood like the use of blue for a river's line work and to fill a body of water. Most adults comprehend that water is symbolically represented by blue. Even on maps with text in foreign languages, you can distinguish water based largely on color. Lines on a road map are another example. They are easily understood as roads even though they have no width, lanes, curbs, or gutters.

Universal examples, however, are rare. Culture and profession influence how one interprets conventions. For example, red, which symbolizes danger (traffic lights and fire for example) and anger in Western countries stands for courage, happiness, success, and Communism in China. The greatest number of conventions, however, comes from professional associations, which have developed complicated formal and informal symbols. For example, geologists use both solid and dashed lines with associated symbols to infer normal faults, strike-slip faults, thrust faults and their associated characteristics like foliation, bedding, and lineation.

This page titled [6.2: Map Communication](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

6.3: Map Types

Maps are the product, the output, of the cartographic communication process. There are several types of maps, usually divided into two categories: general purpose and thematic. General-purpose maps show the location of roads, rivers, institutions, and land covers. Thematic maps depict particular economic, social, demographic, political, or environmental themes like population density, age distribution, political party preference, income, or malaria. The discussion below describes some of the most frequently used thematic maps and includes some conditions for each of their use. It might be helpful to review the portion of Chapter 2 on data types (including levels of measurement) before reading this section.

Dot Density Map

Thematic dot maps use dots or points to show a comparative density of features over a base map (see Figure 6.3). The dots are all the same size. Most dot maps are vector based and usually do not originate from point layers. They derive their dots from values stored in polygon layer attribute fields. Each polygon's attribute value dictates the number of dots displayed across the polygon feature. For instance, if one of your polygon features had a value of 2,223 cattle and you decided to represent 500 cattle with one dot, the map would have four dots randomly draped over the polygon.

Data type: Interval

Feature type: Polygon (sometimes point)



Figure 6.3: Dot density map.

Isoline (Isarithmic) Map

Isoline maps use continuous lines (sometimes called isolines or contours) to reference differences across a continuous surface. Lines connect places that have the same value. They require at least ordinal data, but generally use interval or ratio data.

Two types of isoline maps exist: Isometric maps contain absolute data, which is based on scanning the entire surface. Remote sensing imagery is a good example. Isometric maps are largely raster-based due to the continuous nature of the layer. Isopleth maps, the second type, create continuous data from discrete data. In other words, it derives a continuous surface from multiple known locations where measurements were taken (locations in a point layer) (See Figure 6.4). Temperature and rainfall maps are good examples. These maps are both raster and vector based.

Data type: Interval or Ratio (sometimes ordinal)

Feature type: Raster or point

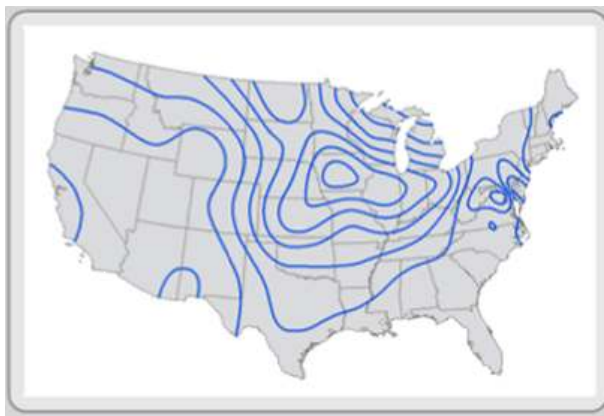


Figure 6.4: Isoline map.

Graduated Symbol Map

Graduated symbol maps use symbols that occur at points across a map, but unlike dot maps, the symbol size varies based on quantity or magnitude (see Figure 6.5). Usually one graduated symbol is generated from within each polygon feature, and its symbol size is determined by the polygon's attribute value. Higher values get larger symbols. Graduated symbol maps depict ordinal or interval data. The symbols can be circles, squares, or just about any form. Point feature layers can also be used to create graduated symbol maps.

There are two kinds of graduated symbol maps (both are vector based): Proportional symbol maps have symbols that are equivalent to the quantity represented. Range graded symbol maps use a user-defined number of classes each with a different-sized graduated symbol to represent its magnitude. Each symbol represents a range of values, not a single value.

Data type: Ordinal and Interval

Feature type: Polygon and point

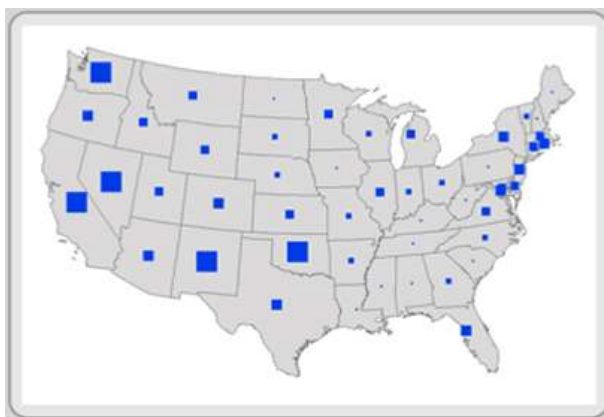


Figure 6.5: Graduated symbol map.

Choropleth Map

Choropleth maps are the most common and easily recognized of the thematic maps (see Figure 6.6). They show ratios, proportions, and percentages that are aggregated within polygon features. They use grays and colors to depict each polygon's (or each pixel's) attribute value. An election map, depicting shaded states of blue or red—based on the percentage of votes cast for a politician or a party—is an example. Like graduated symbol maps, choropleth maps have proportional and range graded variations, but true choropleths only use ratio data. Simpler “shade” or “color” maps use nominal or ordinal data.

Data type: Rate, proportion, or percentage

Feature type: Raster or Polygon

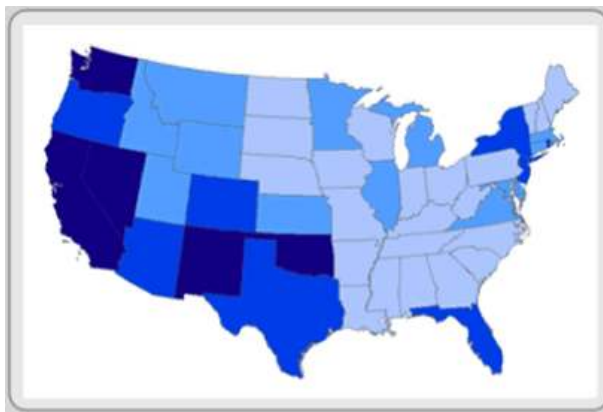


Figure 6.6: Choropleth map.

Cartogram

Cartograms distort polygon shape to depict the magnitude of attribute data (see Figure 6.7). A high value within a normally small geographic unit (polygon) creates a large geographic unit on the map because the size of the polygon is based on the feature's attribute value. There are different types of cartograms; they vary on the degree to which the geography is preserved. Broadly, there are two types of cartograms: Non-contiguous is the simplest and easiest to construct. The polygons do not need to touch each other. They grow and shrink, but they maintain their shape. Contiguous cartograms maintain their connections with each other, but to do this, they distort the shape of their polygons. Cartograms are vector-based, but most commercial software packages do not have a routine to create cartograms.

Data type: Interval and Ratio

Feature type: Polygon



Figure 6.7: Cartogram.

Flow Map

Flow maps show the movement of goods, people, and ideas between places (see Figure 6.8). Usually they depict interval data by differentiating the width of the lines connecting places. Simpler types of flow maps could depict nominal and ordinal data. Flow maps are vector-based, but most commercial software packages do not have sophisticated flow-mapping routines.

Data type: Interval

Feature type: Line

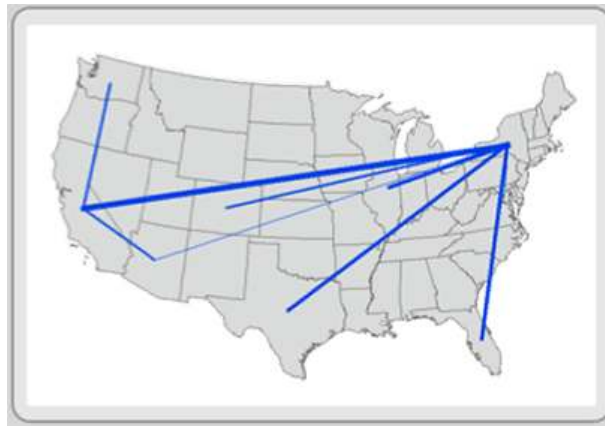


Figure 6.8: Flow map.

Density Map

Density maps depict the concentration of points (and less often lines) across a continuous surface (see Figure 6.9). Conceptually, each point in the feature layer spreads out its presence beyond its immediate location to include adjacent areas. Then, each cell in the raster output image makes a circular search around itself to determine how many points (or lines) fall within the circular radius. These maps most often depict feature counts, but density can also be derived from one of the point layer's attribute fields.

Data type: Interval

Feature type: Point

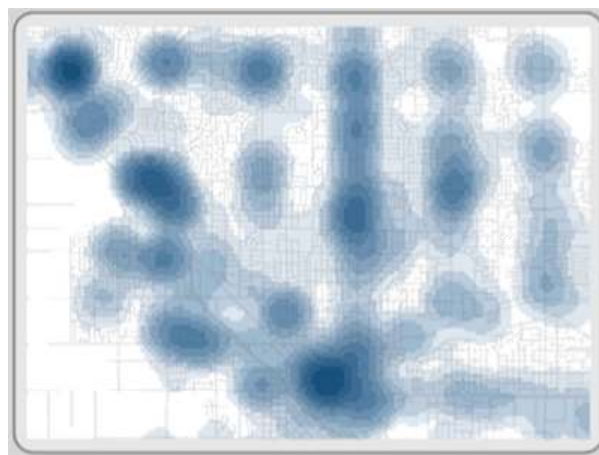


Figure 6.9: Density map.

This page titled [6.3: Map Types](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

6.4: Constructing Thematic Maps

Successful maps are understood by their audience. GIS programs provide quick and powerful ways to produce maps, but good maps are not always the result. Your task is to create maps that communicate well. The following are 14 rules you should consider when producing maps.

1. Pick an Appropriate Mapping Technique.

The maps described above indicate the necessary feature and data types needed to create them. You may need to change your feature type (points, lines, polygons, or raster surfaces) or your data type (nominal, ordinal, interval, ratio, rate, proportion, or percentage) to use a desired mapping technique, or you can use the proper map technique for the data and feature type you already have.

While choropleth maps are the most common thematic map, they are also the most misused. Instead of using rate, proportion, or percentage data, they often exhibit count data (interval and ratio data), which produces faulty impressions about the data's distribution across a map. The problem is that smaller geographic units may have lower counts and larger units may have higher values. In addition, the dark gray tint that a large state exhibits captures the eye more than the same shade on a small state. If you need to use count data, use graduated symbol maps.

2. Show the data.

Edward Tufte, in his book, *The Visual Display of Quantitative Information* (1983), states: "Above all else, show the data." Although he was primarily writing about graphic design, it is also the most important rule concerning map design. Maps must draw the reader's attention to their substance—their theme—not to something else. To "show the data", you should maximize the size of the map frame (which includes the base map and theme). Make it as large as possible. Then, focus the map's ink on the theme. In the advertisement in Figure 6.10, the "data"—the watch—is maximized.



Figure 6.10: "Show the data."

After the map frame, the most important elements are the title and then the legend. These elements need to occupy prominent positions. When given a map to read, your eyes should snap to the map frame (and its theme), title, and legend (in that order). As a first approximation, the most important elements should be located near the top or just inside the left-hand margin of the page. Less important and ancillary map elements can be positioned along the bottom and toward the right (see right-hand map in Figure 6.11).

3. Map balance.

Once the elements are arranged to reflect their importance, give attention to the map's overall balance (see Figure 6.11). Distribute the map elements as evenly as possible within the map's border to avoid unnecessary crowding or, conversely, large blank areas, called white space. When balancing a map, consider its color, shape, and the size of both the map frame and the other map elements.

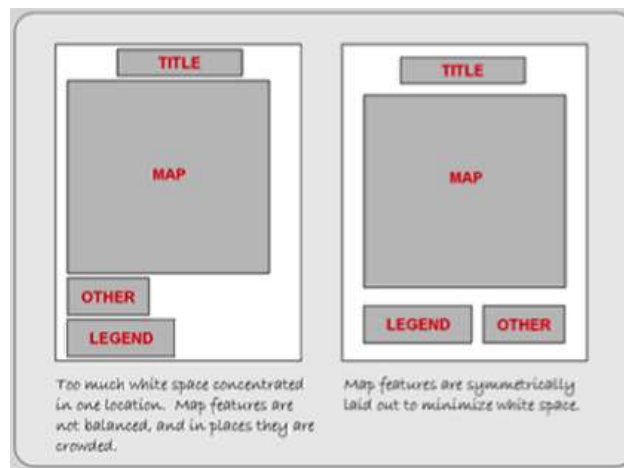


Figure 6.11: Map balance. The map on the right displays little white space with its symmetrical layout. Asymmetrical layouts also work as long as they minimize white space and present the map features in a logical interpretive order

4. Visual Contrast

Contrasts should be striking. It is important to establish a visual hierarchy both for the map features that depict the data—the darker and larger features will stand out (see Figures 6.5 and 6.6 as examples—and for the text (larger fonts should be used for the most important text). Good visual contrast shows the map reader what is most important. It catches their eye. Other features should fall into the background. Psychologists say that dark areas, heavy patterns, complex shapes, and recognizable figures stand out, so make sure on your map that you want them to.

5. Clarity and Legibility

Your map needs to be both clear and legible. For clarity, do not overload the map with too much information and place the map elements in logical positions. Ask yourself, does the map make sense? Is it clear? Is it laid out well?

For legibility, make your lines sharp and the grayscale shades in a choropleth map distinguishable from each other. When it is important to make distinctions, show the differences in your features by using different sizes (dot size, width of lines, font size), textures (solid lines versus dashed or dotted lines), and colors. For example, large cities might be identified with a big black square and highlighted with a larger font than smaller cities and towns. Freeways and major highways might be highlighted by a thicker and darker line than less used roads.

6. For choropleth and graduated-symbol maps, pick an appropriate classification method.

To create a graduated symbol or choropleth map, you need to answer two questions: 1) Which classification method displays your data most accurately? 2) How many classification categories (or classes) will you have? The first question addresses an important point: you need to depict as accurately as possible the underlying distribution of the field's attribute values. We call this simply the "data's shape." This is a somewhat difficult—but important—task because the whole point of displaying the attribute values cartographically is to generalize the data to present or reveal its spatial pattern. Using the wrong classification scheme distorts the data and obscures spatial patterns. There are many classification methods including equal interval, quantiles, arithmetic progression, standard deviation, and natural breaks (these are discussed below).

The answer to the second question may come more naturally. If you have too few categories, the map may obscure the data's distribution. Too many categories are just as fruitless and equally unlikely to reveal spatial patterns. In addition, it is difficult for most map readers to distinguish among more than seven classes. Except for some maps that display nominal data, seven or more classes are too many and the map becomes an illustrated table. Try to use between three and six classes. Let the data's shape help you decide. Create a histogram of the attribute field you want to map (see Calculating Descriptive Statistics in Chapter 5). Frequently, you might discover natural valleys (called cut points) that could define individual classes. It takes some training, but it will become easier. In other cases, the data's shape might give you very little feedback, and you will have more command over the number of classes your map uses.

Regarding classification methods, Figures 6.13 to 6.17 portray the median age of California's 58 counties. Even though they use the same attribute and have five classes apiece, they look somewhat different because each map uses a different classification

method. The first map uses equal intervals (Figure 6.13), the second is divided into quantiles (Figure 6.14), the third is arithmetic progression (Figure 6.15), the fourth uses standard deviation (Figure 6.16), and finally, the last map has user-assisted natural breaks (Figure 6.17).

Even though the maps use the same attribute field, they convey different spatial patterns. Other attribute fields might reveal greater variation among the maps. The five maps in this example are a bit subtle, but variation still exists. The equal interval method seems to stress the lowest population values while the arithmetic progression method accentuates the highest categories. You use different ranging methods to generalize different types of data distributions. Each classification method is suited to a particular data “shape”.

Therefore, the first step in preparing a choropleth or a graduated-symbol map is to explore the shape of any statistical dataset you plan to map (left hand graphic in Figures 6.13 to 6.17 are examples). To do this, you should plot a histogram (or a dot plot) of the data and employ basic descriptive statistics to explore the data’s shape and distribution. As stated above, many GIS programs provide options that graph data and automatically calculate descriptive statistics like mean, mode, median, range, and standard deviation (see Chapter 5). Figure 6.12 depicts various generalized dataset shapes.

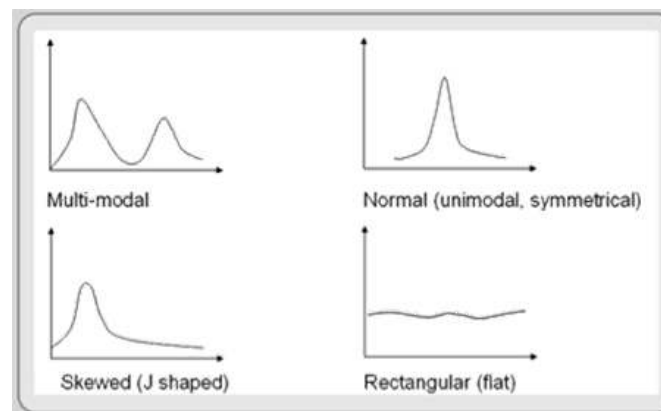


Figure 6.12: Generalized data shapes. The shape determines the classification method you use.

In generalizing statistical distributions, cartographers use the term “cutpoint” to refer to the boundaries between categories. The following classification methods differ in how they assign cutpoints. Some of the most commonly employed ranging methods include:

A. Equal Interval (also called Equal Steps or Equal Size) – This method takes the difference between the low and high values of a distribution and divides the difference into evenly spaced intervals. If 0 and 10 were the low and high values of a distribution, and you wanted to divide the data into five categories, the cut-points would be: 0, 2, 4, 6, 8, and 10. Each category will be differentiated by its own shade or pattern. The method is useful for mapping rectangular (even) distributions.

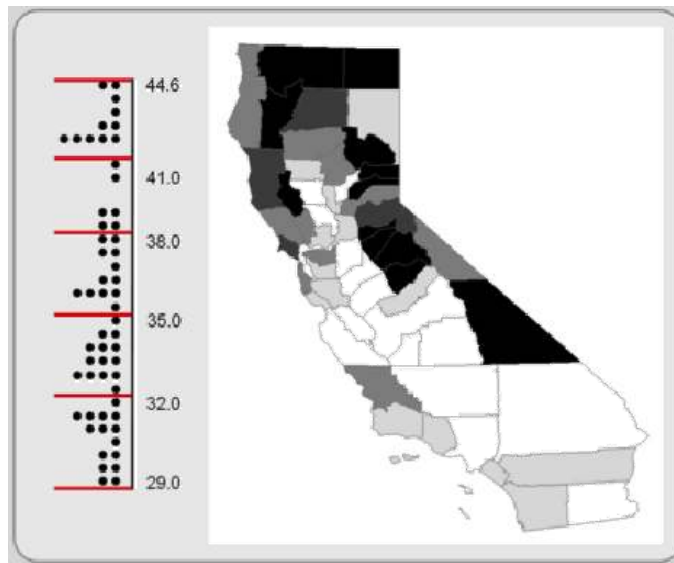


Figure 6.13: Equal interval classification method.

B. Quantiles (and Percentiles) – This method arranges your attribute values from low to high and places equal numbers of observations in each category. If your data included one hundred observations and you wished to divide the data into five categories, the lowest twenty observations would be placed into category one, the next twenty into the second, and so on until the highest twenty observations were placed into category five. The percentile method is similar to the quantile method described above; it, however, places equal percentages of the observations in each of the categories. Both of these methods are also useful for mapping rectangular distributions.

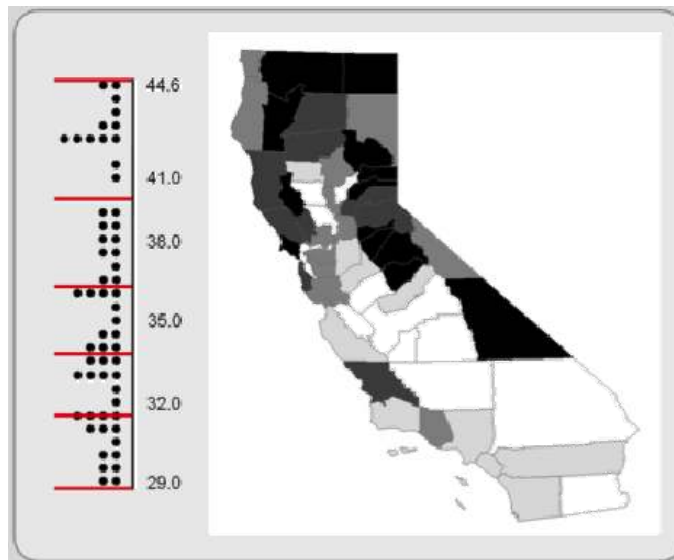


Figure 6.14: Quantiles (Quintiles) classification method.

C. Arithmetic Progression – In this method, the widths of the category intervals are increased in size at an arithmetic (that is, at an additive, multiplied, etc.) rate. If your first category is one unit wide and you choose to increment the width one unit at a time, the second category would be two units wide, the third three units wide, and so forth to the end of the distribution. This method can be applied to J-shaped or heavily skewed (either positive or negatively skewed) data.

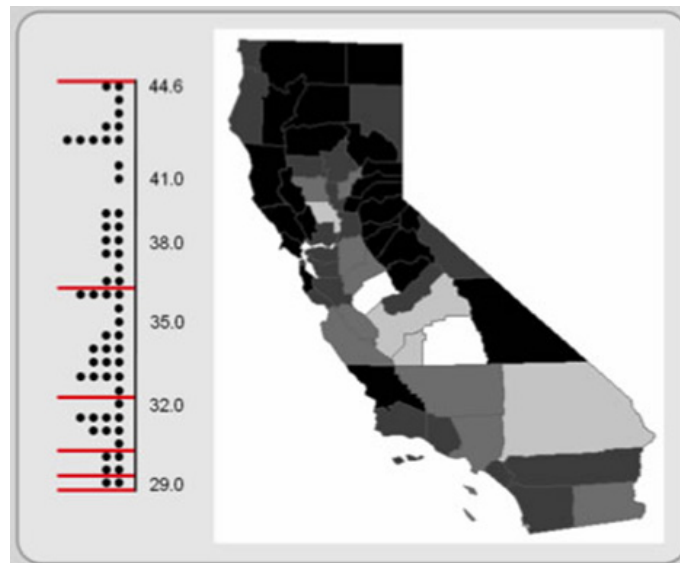


Figure 6.15: Arithmetic Progression classification method

D. Standard Deviation – In this method, the standard deviation of the distribution is used to set the cutpoints above and below the mean. This method should only be applied to distributions that approximate a normal or bell-shaped distribution.

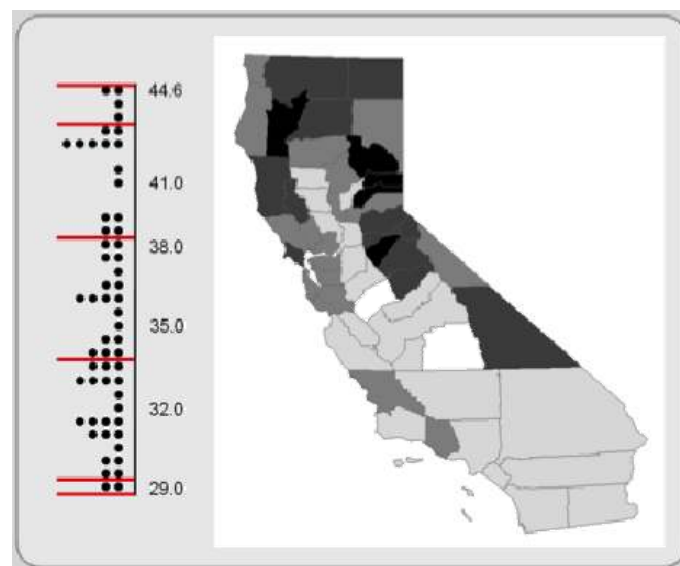


Figure 6.16: Standard deviation classification method.

E. Natural Breaks (Discontinuous) – This classification method creates classes by grouping clusters of similar values. If your data has several peaks, the valleys in between are the breaks (or cutpoints). You can manually adjust your cut points when visually inspecting the data's distribution. This is a good choice for data sets that are multi-modal or those that do not approximate some of the before mentioned data curves.

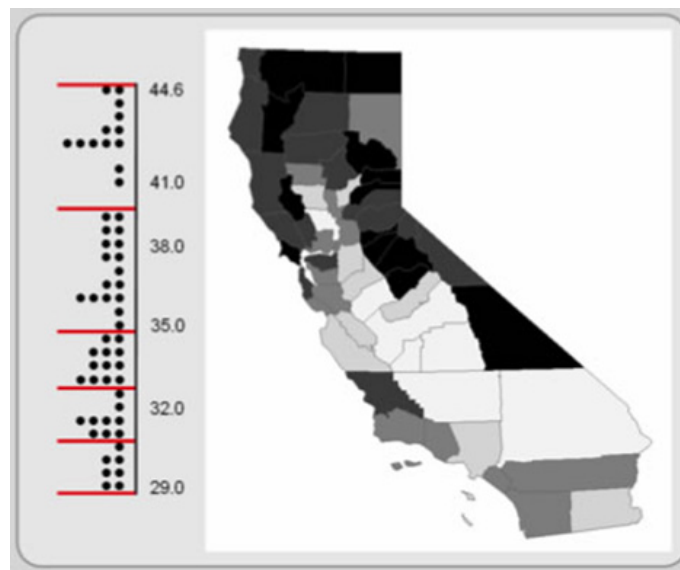


Figure 6.17: Natural Breaks classification method.

7. General Considerations about Text.

The lettering that goes on a map is important. Plan your text just as you plan the rest of your map. All text should be map-reader friendly, so spell out words, run words from left to right, avoid running words vertically, and do not use words that are mysterious and grandiose. An exception to this is the use of U.S., which at least to a U.S. audience is understandable.

Also, include little messages on the map that help explain some of the patterns. Ask people to read and interpret your map's text. If anyone is confused, rewording is necessary.

Typography is the appearance of type. Type should be clear, precise, and modest. Like your map, it should be legible and clear.

There are four major type elements: style, form, size, and contrast. The type's style refers to fonts and the presence or absence of serifs. Fonts are the shape and pattern of letters. There are thousands of fonts, and a short list includes Helvetica, Arial, Universal, Times, and Courier. Generally try to limit your map to one font. Two can work well, but you should have a good reason for additional fonts. Do not pick a font that is too ornate. It catches the map reader's eye, when you want the map's theme (the data) to catch the eye first.

Serifs are the attachments found at the end of some letters. Sans-Serif means that the font is without serifs. Studies indicate that Sans-Serif fonts can be read more rapidly, but less information is retained. Serifs provide greater comprehension, possibly because there is greater differentiation among letters that have serifs. As a guideline, use serifs, but nothing too ornate, for your larger text. Most Sans-serif fonts, however, are a better choice for smaller text because they are easier to see.

Serif – Times New Roman	Sans-Serif – Arial
Serif – Clarendon	Sans-Serif – Calibri
Serif – Garamond	Sans-Serif – CG Omega
<i>Serif – Monotype Corsiva</i>	Sans-Serif – Helvetica
Serif – Palatino	Sans-Serif – Universal

Figure 6.18: Serif and SanSerif fonts.

The form of type includes case (upper, lower, mixed), italics, and boldfacing. Capitalization indicates something is important, but capitalized text is harder to read, so try using mostly mixed case. Book titles, water bodies, and referenced titles should be italicized (not underlined). Underlining will catch the map reader's eye too much.

The size of type has been traditionally measured in points. 1 point is $\frac{1}{72}$ of an inch. Increasingly, some systems use conventional units like inches or millimeters to measure size.

The contrast of the type with the background is important for legibility. Your type should always stand out from the background unless you want a small message on your map to be recessed. Shadings frequently fades-out lettering, so you can remove the shading underneath the type if it is too difficult to see.

On maps, text is used for titles, legends, explanations, source statements, labels, scales, and other map features. Specifics regarding text for these map elements are discussed in the following sections.

8. Titles and Subtitles.

Titles must relate to the map's topic clearly. Good titles require thought, and the following should be considered when constructing a title: make it clear, keep it brief, and start the title off with the subject of the map (the noun) if you can. Never start the title with "Map of..."

Subtitles are frequently used in a series of maps to denote changing locations, time, or conditions. Subtitles should be reserved for detailed data and should always be smaller in size than the title.

9. Legends.

Legends contain explanatory text and list the symbols used on a map. The symbols in the legend should appear exactly as they are on the map. Not all maps require legends; sometimes the necessary information is found in a caption or in the text of an article.

When annotating your legend, be clear and concise; try not to repeat the words you have in your title and subtitle. The legend text should explain the legend's symbols, but it should not use the word "Legend" or "Key". It should be obvious.

10. Scale.

You must provide a scale unless the audience is intimately familiar with the geographic area or distance is of no importance. Scales should not contain the word "Scale". There are verbal, numeric (1:24,000), and graphic scales, but graphic scales are preferred because they maintain their relative size when maps are shrunk or enlarged.

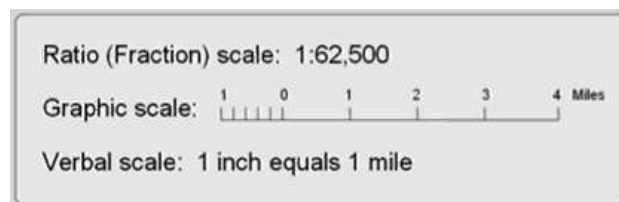


Figure 6.19: Different types of scales.

11. Source Statements.

Most thematic maps need a statement about the data's source and date. If you are using copyrighted material, you must get permission, and the publisher may specify the wording of the statement. Additional text you might provide includes the name of the map's cartographer and the map's projection (coordinate system). Most of this information should be located at the bottom of the map (perhaps near the right-hand boarder) and the text should be small (usually the smallest on your map).

12. Labeling.

Labeling gives descriptions to various features and provides the flexibility to attach descriptions to individual point, line and polygon features. The Swiss cartographer Eduard Imhof (1975) provides six basic rules for label placement. Labels should:

- be legible.
- have clear association with the features they describe.
- not disturb other map contents.
- assist to show the feature's extent.
- reflect the classification and hierarchy of features.
- not be evenly dispersed nor clustered together.

Additional rules might include:

- place the names on the land or at sea

- not be curved unless necessary (unless following a river's course)
- take precedence over line work
- be connected with the feature by the use of an arrow if the label cannot be placed near the feature.

Their positioning can be difficult and frustrating at times. It may not be possible to satisfy all of these labeling rules. As Imhof emphasizes: “no rule without exception” (1975, 29). Two rules often stand in opposition. You must consider each conflict individually, and choose the best solution.

13. North Arrows.

Unless it is obvious, provide a north arrow, but keep it small and simple.

14. Use of color.

Color can dramatically heighten your message, or it can be a distraction for the map reader. There are so many pitfalls, that Edward Tufte warns us about the use of color: “Above all”, he says, “do no harm” (1990, 81).

Color can be used throughout your map, but the most important color decisions involve your map features and especially the theme portrayed. For your theme, color choices depend on whether it's based on qualitative or quantitative data (see Chapter 2). Qualitative data are best mapped using unlike colors (or different patterns if the map is black and white), each having the same weight or brightness. A sequential color scheme, using a single hue, ranging from light to dark, should not be used. This type of color scheme insinuates that one color (and thus, one category) is greater than the other.

If the theme uses quantitative data, you must decide if the data is uni-polar or bi-polar. Most data are uni-polar, which means there is no natural or meaningful dividing point. A parcel's acreage and price are examples. For uni-polar data, a sequential color scheme is recommended. Imagine a palette starting with a very light blue and graduating through medium blues to conclude with a dark blue. The eye tends to be led by these patterns of light and dark, and dark (Tufte's ink) catches our eye first. For this reason (as well as tradition), darker shades should mean “more”.

Bi-polar data, on the other hand, has a dividing point that makes sense. The natural increase rates of countries are an example. Most countries have more births than deaths, so they have positive growth rates. Several countries, however, possess negative natural increase rates. In this case, zero is a clear divide in the attribute values. An attribute's mean and medium can also be used to divide data (but it must make sense). For bi-polar data, diverging color schemes—those that have two distinct hues—are recommended. An example might start dark green, progressively get lighter until it becomes gray-like, emerge as light red, and progressively get darker red.

In all cases, the color of each class must be clearly distinguishable from each other. Sometimes the category colors may appear to be discrete on the computer screen, but if the purpose is a printed map, you must test the colors with a hard-copy print (because they will vary). Also note that colors vary on different printers (even different makes from the same manufacturer) so beware. If you use a data projector to display your map, understand that some colors (especially yellow) get washed out and may be hard to see against a screen. Also, color features against a light gray or muted color background stand out more. It can also reduce eye strain in the map reader.

Much theory has been developed that aids cartographers in choosing colors. As discussed earlier, different colors have emotional and cultural associations. In addition, many colors are not perceived equally by all people, and some of these colors should be avoided. The geographer Cindy Brewer has developed ColorBrewer (<http://colorbrewer2.org/>) to help you choose different color schemes, and see how they change your perception of your results.

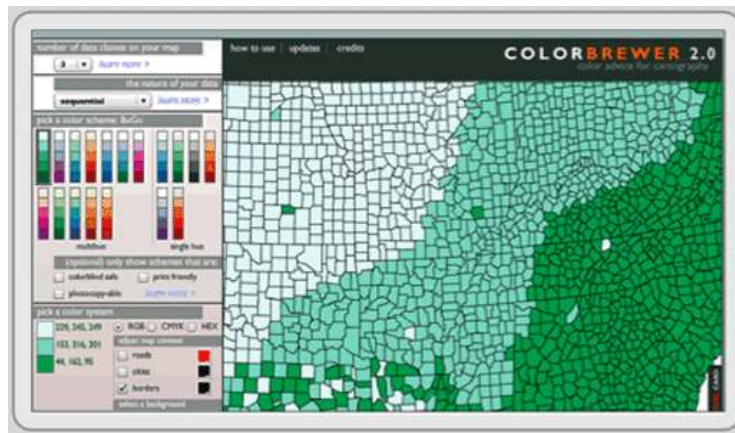


Figure 6.20: Colorbrewer 2.0 (<http://colorbrewer2.org/>).

This page titled [6.4: Constructing Thematic Maps](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

6.5: Types Of Output

Although there are statistics, graphs, and reports, the map is the main way to depict output from a GIS. They are easily displayed, printed, modified, and re-displayed. Maps are taking new forms, which is bending our conception of what maps are. While printed, projected (over a data projector or computer screen), and web-based static (pdfs, jpegs, etc.) maps dominate, animated and interactive maps have appeared and are changing GIS and cartography. Think of rotating themes on an interactive map, and looking at spatial relationships from different angles. Multimedia technology even makes it possible to integrate sound and video, and it encourages different ways to facilitate map presentation and spatial analysis.

This page titled [6.5: Types Of Output](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [Michael Schmandt](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

Index

A

accessible
InfoPage

accuracy
1.5: Concepts
3.4: Map Verification and Editing

analysis
5.5: Connectivity Analyses

analysis functions
1.6: Data Model
5.2: Selection and Measurement

analytical functions
5.1: Introduction

analytical operations
2.1: Introduction

aspect
5.4: Neighborhood Operations

attribute data
1.6: Data Model
2.4: Phase 3- Data Capture
4: Attribute Tables and Data Preprocessing
4.1: Introduction
4.5: Attribute Preprocessing And Editing

attribute fields
5.2: Selection and Measurement

attribute preprocessing
4: Attribute Tables and Data Preprocessing

attribute values
4.4: Principles Of Database Management – Raster

attribute verification
4: Attribute Tables and Data Preprocessing

attributes
3.1: Introduction

attributeverification
4.6: Attribute Verification

audience
6.2: Map Communication

B

buffering
5.4: Neighborhood Operations

C

career services
1.4: GIS Education And Training

cartograms
6.3: Map Types

cartography
1.1: Introduction
1.3: Manual Gis And Computer-Based GIS
1.5: Concepts
6.2: Map Communication
6.3: Map Types
6.4: Constructing Thematic Maps
6.5: Types Of Output

choropleth maps
6.3: Map Types

city planning
1.5: Concepts

classification settings
4: Attribute Tables and Data Preprocessing

collaboration
InfoPage

color schemes
6.4: Constructing Thematic Maps

communication
6.1: Introduction
6.2: Map Communication

community
InfoPage

Concepts
1: Introduction to GIS and Mapping

connectivity
5.5: Connectivity Analyses

coordinate
3.1: Introduction

coordinate systems
3.2: Earth

critical thinking
5.1: Introduction

D

data acquisition
2.1: Introduction

data analysis
1.2: Definitions Of Geographic Information Systems (GIS)
2.2: Phase 1- GIS Database Design
5.2: Selection and Measurement

data capture
2.1: Introduction
2.4: Phase 3- Data Capture

data editing
2.1: Introduction
3.4: Map Verification and Editing

data files
4.1: Introduction
4.3: Principles Of Database Management – Vector

data formats
4.3: Principles Of Database Management – Vector
4.5: Attribute Preprocessing And Editing

data input
2.1: Introduction

data integration
1.3: Manual Gis And Computer-Based GIS

data management
1.2: Definitions Of Geographic Information Systems (GIS)
4.1: Introduction

data manipulation
4.3: Principles Of Database Management – Vector

Data Model
1: Introduction to GIS and Mapping

data organization
4.3: Principles Of Database Management – Vector

data sets
5.1: Introduction

data storage
2.1: Introduction
4.4: Principles Of Database Management – Raster

data visualization
1.2: Definitions Of Geographic Information Systems (GIS)
1.3: Manual Gis And Computer-Based GIS
1.5: Concepts
1.6: Data Model
6.4: Constructing Thematic Maps
6.5: Types Of Output

dataacquisition
2.3: Phase 2- Data Acquisition

dataanalysis
2.3: Phase 2- Data Acquisition
4.2: Attribute Data

database management
4: Attribute Tables and Data Preprocessing
4.1: Introduction
4.3: Principles Of Database Management – Vector

datacollection
2.3: Phase 2- Data Acquisition

datacorrection
4.6: Attribute Verification

dataevaluation
2.3: Phase 2- Data Acquisition

dataquality
2.3: Phase 2- Data Acquisition

datasets
2.2: Phase 1- GIS Database Design

datavalidation
4.6: Attribute Verification

decision making
2.2: Phase 1- GIS Database Design

DEM
5.5: Connectivity Analyses

descriptive statistics
5.2: Selection and Measurement

design
6.2: Map Communication

design phase
2.1: Introduction

Digital Elevation Models
5.4: Neighborhood Operations

digital maps
4.1: Introduction

distortion
3.2: Earth
6.1: Introduction

dot density maps
6.3: Map Types

E

editing attributes
4.5: Attribute Preprocessing And Editing

education
InfoPage
1.4: GIS Education And Training

environmental analysis
2.2: Phase 1- GIS Database Design

Environmental Issues
1.2: Definitions Of Geographic Information Systems (GIS)

environmental planning
5.3: Overlay Analysis

external data files
4.3: Principles Of Database Management – Vector

F

feature editing
3.4: Map Verification and Editing

feature layers

- 4.3: Principles Of Database Management – Vector
- 4.5: Attribute Preprocessing And Editing

feature location

- 5.1: Introduction

feature selection

- 5.2: Selection and Measurement

file size

- 4.4: Principles Of Database Management – Raster

flow maps

- 6.3: Map Types

G

general purpose maps

- 6.3: Map Types

geocoding

- 4.1: Introduction
- 4.5: Attribute Preprocessing And Editing

geographic

- 4.2: Attribute Data

geographic data

- 5.1: Introduction
- 6.2: Map Communication

geographic features

- 4.3: Principles Of Database Management – Vector

geographic information system

- 2.2: Phase 1- GIS Database Design

geographic information systems

- TitlePage
- 1: Introduction to GIS and Mapping
- 1.6: Data Model
- 3.2: Earth
- 5.3: Overlay Analysis
- 6.1: Introduction

geographic patterns

- 1.2: Definitions Of Geographic Information Systems (GIS)

geographical

- 3.1: Introduction

geographical features

- 1.5: Concepts

geographicdata

- 4.6: Attribute Verification

geography

- 1: Introduction to GIS and Mapping
- 1.3: Manual Gis And Computer-Based GIS
- 1.4: GIS Education And Training

geoid

- 3.2: Earth

geometric conversion

- 3.4: Map Verification and Editing

georeferencing

- 3.3: Map Preprocessing

geospatial analysis

- 5.3: Overlay Analysis

geotechnology

- 1.1: Introduction
- 1.4: GIS Education And Training

GIS

- TitlePage
- 1: Introduction to GIS and Mapping
- 1.1: Introduction
- 1.2: Definitions Of Geographic Information Systems

(GIS)

- 1.3: Manual Gis And Computer-Based GIS
- 1.4: GIS Education And Training
- 1.5: Concepts
- 1.6: Data Model
- 2.1: Introduction
- 2.2: Phase 1- GIS Database Design
- 2.3: Phase 2- Data Acquisition
- 2.4: Phase 3- Data Capture
- 3.1: Introduction
- 3.3: Map Preprocessing
- 3.4: Map Verification and Editing
- 4.1: Introduction
- 4.2: Attribute Data
- 4.3: Principles Of Database Management – Vector
- 4.5: Attribute Preprocessing And Editing
- 5.1: Introduction
- 5.2: Selection and Measurement
- 5.3: Overlay Analysis
- 5.4: Neighborhood Operations
- 6.1: Introduction
- 6.4: Constructing Thematic Maps
- 6.5: Types Of Output

GIS capabilities

- 2.1: Introduction

GIS Education

- 1: Introduction to GIS and Mapping

GIS Training

- 1: Introduction to GIS and Mapping

GPS

- 1.1: Introduction
- 2.4: Phase 3- Data Capture

graduated symbol maps

- 6.3: Map Types

grid structure

- 4.4: Principles Of Database Management – Raster

H

historical GIS

- 1.3: Manual Gis And Computer-Based GIS

I

image processing

- 4.4: Principles Of Database Management – Raster

incomevalues

- 4.6: Attribute Verification

incorrectvalues

- 4.6: Attribute Verification

interactive

- InfoPage

interactive maps

- 6.5: Types Of Output

interpolation

- 5.4: Neighborhood Operations

introduction

- 4: Attribute Tables and Data Preprocessing

isoline maps

- 6.3: Map Types

J

job opportunities

- 1.4: GIS Education And Training

joining data files

- 4.5: Attribute Preprocessing And Editing

L

land use planning

- 5.3: Overlay Analysis

landuseclasses

- 4.6: Attribute Verification

landuses

- 4.2: Attribute Data

latitude

- 3.1: Introduction
- 3.2: Earth

learning path

- 4: Attribute Tables and Data Preprocessing

LibreTexts

- InfoPage

LIDAR

- 2.4: Phase 3- Data Capture

longitude

- 3.1: Introduction
- 3.2: Earth

M

Manual GIS

- 1: Introduction to GIS and Mapping

map analysis

- 1.1: Introduction

map design

- 1.1: Introduction
- 6.1: Introduction

map elements

- 6.4: Constructing Thematic Maps

map layers

- 3.3: Map Preprocessing

map production

- 1.1: Introduction

map projections

- 3.2: Earth

map verification

- 3.4: Map Verification and Editing

mapping

- 1: Introduction to GIS and Mapping
- 1.1: Introduction
- 1.2: Definitions Of Geographic Information Systems (GIS)

- 1.3: Manual Gis And Computer-Based GIS
- 1.6: Data Model
- 2.1: Introduction
- 2.2: Phase 1- GIS Database Design
- 2.4: Phase 3- Data Capture
- 4.6: Attribute Verification
- 6.4: Constructing Thematic Maps

mapping process

- 6.2: Map Communication

mapping techniques

- 5.3: Overlay Analysis

mappreprocessing

- 3.3: Map Preprocessing

maps

- 1.5: Concepts
- 4.2: Attribute Data
- 6.1: Introduction
- 6.2: Map Communication
- 6.3: Map Types
- 6.5: Types Of Output

measurements

- 5.2: Selection and Measurement

metadata

- 2.3: Phase 2- Data Acquisition

Michael Schmandt

[TitlePage](#)

missingattributes

[4.6: Attribute Verification](#)

model

[5.3: Overlay Analysis](#)

modeling

[5.5: Connectivity Analyses](#)

multimedia technology

[6.5: Types Of Output](#)

N

neighborhood operations

[5.4: Neighborhood Operations](#)

network

[5.5: Connectivity Analyses](#)

nonprofit

[InfoPage](#)

O

OER

[InfoPage](#)

overlay

[5.3: Overlay Analysis](#)

P

photogrammetry

[2.4: Phase 3- Data Capture](#)

pixels

[4.4: Principles Of Database Management – Raster](#)

populationdensity

[4.2: Attribute Data](#)

predictive modeling

[1.2: Definitions Of Geographic Information Systems \(GIS\)](#)

preprocessing

[3.1: Introduction](#)

[4.1: Introduction](#)

primarydata

[2.3: Phase 2- Data Acquisition](#)

printed maps

[6.5: Types Of Output](#)

project goals

[2.2: Phase 1- GIS Database Design](#)

projection

[6.1: Introduction](#)

projections

[3.2: Earth](#)

proximity analyses

[5.4: Neighborhood Operations](#)

public health

[1.3: Manual Gis And Computer-Based GIS](#)

Q

quadtree

[4.4: Principles Of Database Management – Raster](#)

qualitative

[4.2: Attribute Data](#)

quantitative

[4.2: Attribute Data](#)

querying attributes

[4.5: Attribute Preprocessing And Editing](#)

R

raster

[1.6: Data Model](#)

[4: Attribute Tables and Data Preprocessing](#)

[4.1: Introduction](#)

[5.5: Connectivity Analyses](#)

raster data model

[4.4: Principles Of Database Management – Raster](#)

raster images

[3.4: Map Verification and Editing](#)

raster layers

[4.4: Principles Of Database Management – Raster](#)

raster process

[5.2: Selection and Measurement](#)

raster systems

[5.4: Neighborhood Operations](#)

rasterization

[3.3: Map Preprocessing](#)

remote sensing

[1.1: Introduction](#)

[1.6: Data Model](#)

[2.4: Phase 3- Data Capture](#)

representation

[6.1: Introduction](#)

resampling

[3.3: Map Preprocessing](#)

research assistantships

[1.4: GIS Education And Training](#)

research question

[2.2: Phase 1- GIS Database Design](#)

route_selection

[5.5: Connectivity Analyses](#)

S

Sacramento State University

[TitlePage](#)

satellite imagery

[3.2: Earth](#)

secondarydata

[2.3: Phase 2- Data Acquisition](#)

slope

[5.4: Neighborhood Operations](#)

software

[1.4: GIS Education And Training](#)

software applications

[1.3: Manual Gis And Computer-Based GIS](#)

sortingdata

[4.6: Attribute Verification](#)

spatial

[3.1: Introduction](#)

spatial accuracy

[3.2: Earth](#)

spatial analysis

[1.2: Definitions Of Geographic Information Systems \(GIS\)](#)

[1.3: Manual Gis And Computer-Based GIS](#)

[1.5: Concepts](#)

[3.3: Map Preprocessing](#)

[4.4: Principles Of Database Management – Raster](#)

[5.1: Introduction](#)

[5.3: Overlay Analysis](#)

[6.4: Constructing Thematic Maps](#)

[6.5: Types Of Output](#)

spatial data

[1.2: Definitions Of Geographic Information Systems \(GIS\)](#)

[1.4: GIS Education And Training](#)

[1.6: Data Model](#)

[2.4: Phase 3- Data Capture](#)

[3.3: Map Preprocessing](#)

[3.4: Map Verification and Editing](#)

[4.3: Principles Of Database Management – Vector](#)

spatial errors

[3.4: Map Verification and Editing](#)

spatial extent

[2.2: Phase 1- GIS Database Design](#)

spatial information

[6.2: Map Communication](#)

spatial patterns

[5.3: Overlay Analysis](#)

spatial queries

[4.5: Attribute Preprocessing And Editing](#)

spatial relationships

[5.1: Introduction](#)

spatial selection

[5.2: Selection and Measurement](#)

spatial understanding

[6.1: Introduction](#)

spatialdata

[2.3: Phase 2- Data Acquisition](#)

[4.2: Attribute Data](#)

sql

[4.5: Attribute Preprocessing And Editing](#)

standards

[3.1: Introduction](#)

statisticalmanipulation

[4.2: Attribute Data](#)

statistics

[6.5: Types Of Output](#)

surveying

[2.4: Phase 3- Data Capture](#)

symbolization

[6.1: Introduction](#)

symbols

[6.2: Map Communication](#)

systems

[3.1: Introduction](#)

T

textbook

[TitlePage](#)

textbooks

[InfoPage](#)

thematic maps

[6.3: Map Types](#)

[6.4: Constructing Thematic Maps](#)

topology

[1.5: Concepts](#)

[3.3: Map Preprocessing](#)

training

[1.4: GIS Education And Training](#)

U

urban studies

[1.5: Concepts](#)

user interface

[4: Attribute Tables and Data Preprocessing](#)

user readability

[6.4: Constructing Thematic Maps](#)

V

vector

[1.6: Data Model](#)

[4: Attribute Tables and Data Preprocessing](#)

[4.1: Introduction](#)

[5.5: Connectivity Analyses](#)

vector layers

[3.4: Map Verification and Editing](#)

vector process

[5.2: Selection and Measurement](#)

vector systems

[5.4: Neighborhood Operations](#)

vectorization

[3.3: Map Preprocessing](#)

visibility_analysis

[5.5: Connectivity Analyses](#)

visual hierarchy

[6.4: Constructing Thematic Maps](#)

visual representation

[6.2: Map Communication](#)

visualization

[1.1: Introduction](#)

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: [GIS Commons: An Introductory Textbook on Geographic Information Systems](#)

Webpages: 46

All licenses found:

- [CC BY-SA 3.0](#): 93.5% (43 pages)
- [Undeclared](#): 6.5% (3 pages)

By Page

- [GIS Commons: An Introductory Textbook on Geographic Information Systems](#) - [CC BY-SA 3.0](#)
 - [Front Matter](#) - [CC BY-SA 3.0](#)
 - [TitlePage](#) - [CC BY-SA 3.0](#)
 - [InfoPage](#) - [CC BY-SA 3.0](#)
 - [Table of Contents](#) - [Undeclared](#)
 - [Licensing](#) - [CC BY-SA 3.0](#)
 - [1: Introduction to GIS and Mapping](#) - [CC BY-SA 3.0](#)
 - [1.1: Introduction](#) - [CC BY-SA 3.0](#)
 - [1.2: Definitions Of Geographic Information Systems \(GIS\)](#) - [CC BY-SA 3.0](#)
 - [1.3: Manual Gis And Computer-Based GIS](#) - [CC BY-SA 3.0](#)
 - [1.4: GIS Education And Training](#) - [CC BY-SA 3.0](#)
 - [1.5: Concepts](#) - [CC BY-SA 3.0](#)
 - [1.6: Data Model](#) - [CC BY-SA 3.0](#)
 - [2: Data Input](#) - [CC BY-SA 3.0](#)
 - [2.1: Introduction](#) - [CC BY-SA 3.0](#)
 - [2.2: Phase 1- GIS Database Design](#) - [CC BY-SA 3.0](#)
 - [2.3: Phase 2- Data Acquisition](#) - [CC BY-SA 3.0](#)
 - [2.4: Phase 3- Data Capture](#) - [CC BY-SA 3.0](#)
 - [3: The Earth and Map Preprocessing](#) - [CC BY-SA 3.0](#)
 - [3.1: Introduction](#) - [CC BY-SA 3.0](#)
 - [3.2: Earth](#) - [CC BY-SA 3.0](#)
 - [3.3: Map Preprocessing](#) - [CC BY-SA 3.0](#)
 - [3.4: Map Verification and Editing](#) - [Undeclared](#)
 - [4: Attribute Tables and Data Preprocessing](#) - [CC BY-SA 3.0](#)
 - [4.1: Introduction](#) - [CC BY-SA 3.0](#)
 - [4.2: Attribute Data](#) - [CC BY-SA 3.0](#)
 - [4.3: Principles Of Database Management – Vector](#) - [CC BY-SA 3.0](#)
 - [4.4: Principles Of Database Management – Raster](#) - [CC BY-SA 3.0](#)
 - [4.5: Attribute Preprocessing And Editing](#) - [CC BY-SA 3.0](#)
 - [4.6: Attribute Verification](#) - [CC BY-SA 3.0](#)
 - [5: Analysis](#) - [CC BY-SA 3.0](#)
 - [5.1: Introduction](#) - [CC BY-SA 3.0](#)
 - [5.2: Selection and Measurement](#) - [CC BY-SA 3.0](#)
 - [5.3: Overlay Analysis](#) - [CC BY-SA 3.0](#)
 - [5.4: Neighborhood Operations](#) - [CC BY-SA 3.0](#)
 - [5.5: Connectivity Analyses](#) - [CC BY-SA 3.0](#)
 - [6: Output](#) - [CC BY-SA 3.0](#)
 - [6.1: Introduction](#) - [CC BY-SA 3.0](#)
 - [6.2: Map Communication](#) - [CC BY-SA 3.0](#)
 - [6.3: Map Types](#) - [CC BY-SA 3.0](#)
 - [6.4: Constructing Thematic Maps](#) - [CC BY-SA 3.0](#)
 - [6.5: Types Of Output](#) - [CC BY-SA 3.0](#)
 - [Back Matter](#) - [CC BY-SA 3.0](#)
 - [Index](#) - [CC BY-SA 3.0](#)
 - [Glossary](#) - [CC BY-SA 3.0](#)
 - [Detailed Licensing](#) - [Undeclared](#)