GEOGRAPHIC INFORMATION SYSTEMS AND CARTOGRAPHY

Adam Dastrup
Salt Lake Community College

LibreTexts™

Geographic Information Systems and Cartography

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, 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.

This text was compiled on 03/26/2025

TABLE OF CONTENTS

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

Introduction

1

Textbook Introduction and Description

Introduction to Geographic Information Systems, including Geographic Information Systems (GIS), Global Positioning Systems (GPS), cartography, remote sensing, and spatial analysis. Readers will learn how to utilize geospatial technology to address social and environmental issues. This course is designed to be used as a stand-alone course to complement other disciplines or as an entry-level course in a geospatial program. Course content is based upon the United States Department of Labor's Geospatial Technology Competency Model for entry-level geospatial occupations including Geospatial or GIS Technicians and Technologists.

- The student will describe the fundamental concepts of Geographic Information Science and Technology.
- The student will demonstrate proficiency in the basic functions of geospatial software and hardware.
- The student will demonstrate awareness of fundamental remote sensing and spatial analysis techniques.
- The student will demonstrate basic proficiency in map creation and design principles, including thematic map display, employment of map projections, and cartographic design.
- The student will demonstrate proficiency in the creation and acquisition of spatial data including the use of the Global Position System.
- The student will demonstrate how to access different sources of data, demonstrate the process of creating data, and discuss the fundamental concepts of data quality.

Author's Notes

This textbook is a composite of multiple geographic information systems resources, along with additions by the author. The textbook is meant to be mobile, interactive, and multimodal. Special thanks to the authors of the GTCM Model Courses from the National Geospatial Technology Center for Excellence (GeoTech Center), and the authors of the Essentials to Geographic Information Systems textbook. Both of these resources and the images used are openly licensed as Creative Commons. All videos are embedded and not downloaded into the textbook in order to not violate any copyright privileges.

CHAPTER OVERVIEW

1: Introduction to Geographic Information Systems

1.1: Introduction

Introduction

Knowing something about where something happens can help us to understand what happened when it happened, how it happened, and why it happened. Whether it is analyzing the spatial and temporal distribution of the COVID pandemic, understanding the loss of biodiversity or climate change, the path of a deadly tornado or hurricane, or better understanding food deserts, knowing something about where things happen is essential to how we understand and relate to our local environment and the world at large. Modern digital mapping technology has revolutionized how society analyzes and understands the spatial and temporal aspects of our physical and cultural environments.

Much of this textbook will focus on digital mapping technology called a **geographic information system** (GIS). GIS is information technology that can help us understand and relate to the world's what, when, how, and why by answering where. Geographic information systems are about digital maps, but they are also about much more.

GIS is used to organize, analyze, visualize, and share data and information from different historical periods (temporal) and at various scales (spatial) of analysis. From climatologists trying to understand the causes and consequences of sea rise to epidemiologists locating ground zero of COVID-19 to archaeologists reconstructing ancient Rome, to politicians and law enforcement trying to understand better how the political consultants are developing campaign strategies for the next presidential election, GIS is a potent tool.

More importantly, GIS is about the *science of geography* as a way to better learn and understand our world. As GIS technology evolves and society becomes ever more geospatially enabled, people are rediscovering the importance of geographic science and the power of maps.

To take full advantage of GIS and related geospatial technology, reflecting on how we already think spatially and temporally concerning the world is helpful. In other words, by recognizing and increasing our geographical awareness about how we relate to our local environment and the world, we will benefit more from using and applying GIS.

Learning Objectives

- Illustrate how we think geographically and spatially daily with mental maps to highlight the importance of asking geographic questions.
- Explain how the fundamental concepts of scale, location, direction, distance, space, and navigation are relevant to geography and geographic information systems.
- Define how a geographic information system is applied, its development, and its future.

Geospatial Technology Competency Model Alignment

- Demonstrate understanding of the conceptual foundations on which geographic information systems (GIS) are based, including the problem of representing change over time and the imprecision and uncertainty that characterizes all geographic information. Geospatial Technology Industry Competency Model

Chapter Sections

Click the "**Previous**" button on the lower left or the '**Next**" button on the lower right to navigate throughout the textbook.

1.2: Geographic Science

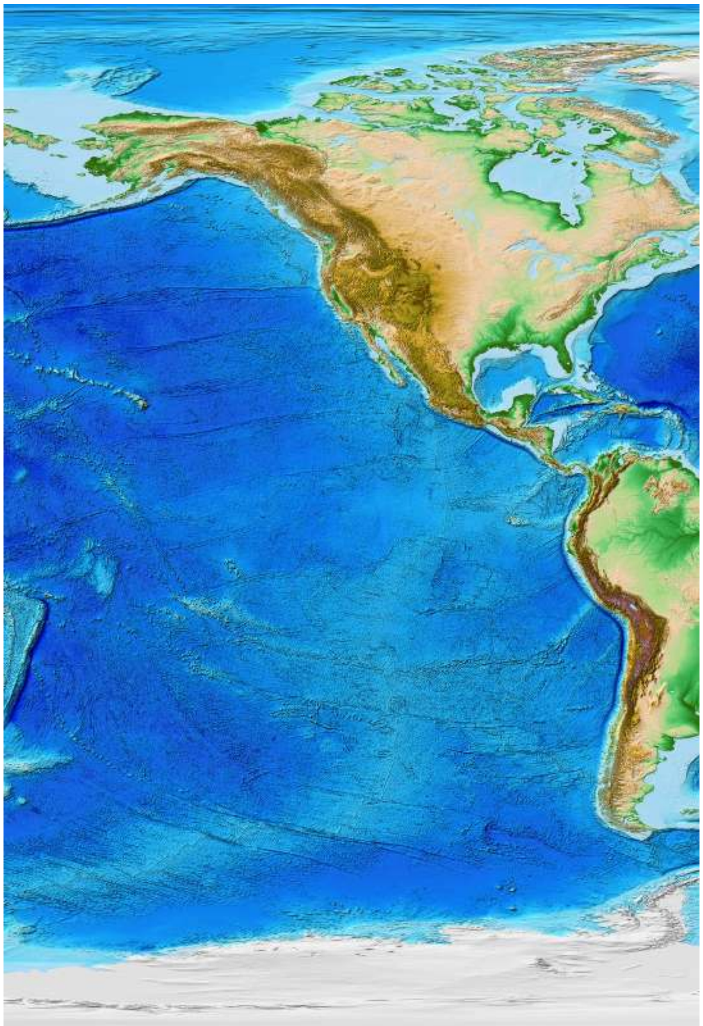
Most individuals define **geography** as a field of study that deals with maps, yet this definition is only partially correct. A better definition of geography may be the study of natural and human-constructed phenomena relative to a spatial and temporal dimension.

The discipline of geography has multiple intersectionalities that empower our understanding of physical and cultural landscapes. Geography bridges the social sciences with physical sciences and can provide a framework for understanding our world. By studying geography, we can begin to understand the relationships and common factors that tie our human community together. The world is undergoing globalization on a massive scale because of the rapid transfer of information and technology. The more we understand our world, the better prepared we will be to address the issues that confront our future. There are many approaches to studying geography. This textbook takes a geospatial approach, focusing on geographic information systems, remote sensing, and global positioning systems.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=28#oembed-1

Geography helps us make sense of the world through four historical traditions:

- **Physical geography** – the study of the natural features of the earth's surface, especially in its current assets, including land formation, climate, currents, and distribution of flora and fauna.
- **Human geography**: the study of the relationships between the physical environment and the activities of humans.
- **World regional geography** – the study of the physical and cultural landscapes of specific regions of the world.
- **Spatial analysis** – the study of the physical and cultural environments using various geospatial technologies, including geographic information systems (GIS), global positioning systems (GPS), and remote sensing such as satellite and aerial imagery.



The discipline of geography has a history that stretches over many centuries. Over this period, geography has become a basic form of human scholarship. Examining the historical evolution of geography as a discipline provides some essential insights concerning its character and methodology. These insights are also helpful in gaining a better understanding of the nature of physical geography.

Geographers seek to answer the *where*, *why*, and *how*. For example, knowing a country's location is undoubtedly helpful, but geographers dig deeper by asking:

- Why is it located there?
- Why does it have a particular shape or spatial pattern, and how does this affect how it interacts with the surrounding area?
- Why do the people of the country have certain cultural features?
- Why does the country have a specific style of government?
- How do we analyze the spatial and temporal patterns in human-environment interactions?

The term "geography" comes from the Greek term geo, meaning "the earth," and graphia, meaning "to write," and many early geographers did precisely that: they wrote about the world. Ibn Battuta, for example, was a scholar from Morocco who traveled extensively across Africa and Asia in the 14th century C.E. Eratosthenes is commonly considered the "Father of Geography," and he authored the book on the subject in the third century BCE. His three-volume text, Geographica, included maps of the entire known world, including different climate zones, the locations of hundreds of different cities, and a coordinate system. This was a revolutionary and highly regarded text, especially for the time. Eratosthenes is also credited as the first person to calculate the circumference of the

Earth. Many early geographers, like Eratosthenes, were primarily cartographers, referring to people who scientifically studied and created maps, and early maps, such as those used in Babylon, Polynesia, and the Arabian Peninsula, were often used for navigation. In the Middle Ages, as academic inquiry in Europe declined with the fall of the Roman Empire, Muslim geographer Muhammad al-Idrisi created one of the most advanced maps of pre-modern times, inspiring future geographers from the region.

Though using more advanced tools and techniques, geography today draws on the foundations laid by these predecessors. Attention to the spatial and temporal perspective unites all geographers. As geographer Harm deBlij once explained, there are three main ways to look at the world. One way is chronological, as a historian might examine the sequence of world events. A second way is systematical, as a sociologist might explore the societal systems that help shape a given country's structures of inequality. The third way is spatially, and this is the *geographic perspective*. When confronted with a global problem, geographers immediately ask *where* and *why*.

Spatial and Temporal Thinking

Although geography is a broad discipline that includes quantitative techniques like statistics and qualitative methods like interviews, all geographers share this common way of looking at the world from a *spatial* and *temporal* perspective. As geospatial technology has advanced, geographers can better analyze the world spatially and over time. For example, the Landsat satellites repeated going over a particular location on Earth roughly every 16 days for the past 50 years, starting in 1972. So, this allows us to understand the world spatially and over time (temporally).

At no other time in the history of the world has it been easier to create or acquire a map of anything. Maps and geospatial technology are literally and virtually everywhere. Though the modes and means of making and distributing maps have been revolutionized with recent advances in computing like the Internet, the art and science of map-making dates back centuries. This is because humans are inherently spatial organisms, and for us to live in the world, we must first somehow relate to it. Enter the mental map.

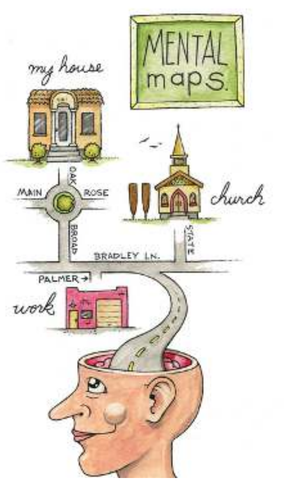
One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=28#oembed-2

Mental Maps

Mental or **cognitive maps** are psychological tools that we all use every day. As the name suggests, mental maps are maps of our environment stored in our brains. We rely on our mental maps to get from one place to another, plan our daily activities, or understand and situate events we hear about from our friends, family, or news. Mental maps also reflect the amount and extent of our geographic and spatial awareness of our local and global environment. What you choose to include and exclude on your map provides insights into what places you think are important and how you move through your place of residence.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=28#oembed-3

What you include and omit on your map, by choice, speaks volumes about your geographical knowledge and spatial awareness, or lack thereof. Recognizing and identifying what we do not know is essential for learning. Only when we identify the unknown can we ask questions, collect information to answer those questions, develop knowledge through answers, and begin understanding the world where we live.



"Mental Map" is licensed under a Creative Commons Attribution 4.0 License.

Asking Spatial and Temporal Questions

Filling in our mental maps and, more generally, the gap in our geographic knowledge requires us to ask questions about the world where we live and how we relate to it. Such questions can be simple with a local focus or more complex with a global perspective. The thread that unifies such questions is geography. For instance, the question of where is an essential part of the questions "Where is the nearest hospital?" and "Where are the biodiversity hotspots concerning cities?" Articulating questions clearly and breaking them into manageable pieces are valuable skills when using and applying geospatial technology.

Spatial and temporal questions, called geographic literacy, can help us identify complex issues and solve difficult problems to understand better and improve our environment. These geographic questions are listed here and followed by a few examples (Nyerges, 1991). Nyerges, T. 1991. "Analytical Map Use." Cartography and Geographic Information Systems (formerly The American Cartographer) 18: 11–22.

Questions about *geographic location*:

- Where is it?
- Why is it here or there?
- How much of it is here or there?

Questions about *geographic distribution*:

- Is it distributed locally or globally?
- Is it spatially clustered or dispersed?
- Where are the boundaries?

Questions about the *geographic association*:

- What else is near it?
- What else occurs with it?
- What is absent in its presence?

Questions about *geographic interaction*:

- Is it linked to something else?
- What is the nature of this association?
- How much interaction occurs between the locations?

Questions about *geographic change*:

- Has it always been here?
- How has it changed over time and space?
- What causes its diffusion or contraction?

These and related geographic questions are frequently asked by people from various areas of expertise, industries, and professions. For instance, urban planners, traffic engineers, and demographers may be interested in understanding the commuting patterns between cities and suburbs (geographic interaction). Biologists and botanists may be curious about why one animal or plant species flourish in one place and not another (geographic location/distribution). Epidemiologists and public health officials are undoubtedly interested in where disease outbreaks occur and how, why, and where they spread (geographic change/interaction/location).

Geospatial technology can assist in answering all these questions and many more. Furthermore, a GIS often opens additional avenues of inquiry when searching for answers to geographic questions. Herein is one of the greatest strengths of the GIS. While a GIS can answer specific questions or solve problems, it often unearths even more exciting questions. It presents more problems to be solved in the future.

Click the "**Previous**" button on the lower left or the '**Next**" button on the lower right to navigate throughout the textbook.

1.3: Geospatial Technology

Geospatial technology is one of the primary driving technologies used by geographers to understand the spatial and temporal aspects of the planet. This textbook will focus on geographic information systems (GIS) and have chapters on remote sensing and global positioning systems (GPS). Each of these technologies has its strengths and disadvantages, but they are some of the most powerful tools humans have ever developed.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=30#oembed-1

Geographic Information Systems (GIS)

Suppose you have launched a new business that manufactures solar panels for homeowners. You are planning a mail campaign to bring this revolutionary new product to the attention of prospective buyers. However, since it is a small business, you cannot afford to sponsor coast-to-coast television commercials or to send brochures by mail to more than 100 million U.S. households. So instead, you plan to target the most likely customers – those who are environmentally conscious, have higher than average family incomes, and live in areas with enough sunshine to support solar power.

Fortunately, data is available to help you define your mailing list. For example, household incomes are routinely reported to banks and other financial institutions when families apply for mortgages, loans, and credit cards. In addition, personal tastes related to environmental issues are reflected in behaviors such as magazine subscriptions, credit card purchases, and types of community development. Market research companies collect such data and transform it into information by creating "lifestyle segments" – household categories with similar incomes and tastes. Your solar company can purchase lifestyle segment information by 5-digit ZIP code or ZIP+4 codes, which designate individual households.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=30#oembed-2

It is astonishing how valuable information is from the millions of daily transactions. The fact that lifestyle information products are often delivered by geographic areas, such as ZIP codes, speaks to the appeal of GIS. The scale of these data and their potential applications are increasing continually with the advent of new mechanisms for sharing information and purchasing linked to our GPS-enabled smartphones. GIS is a computer-based tool to help people transform geographic data into geographic information.

GIS arose from the need to perform spatial queries on geographic data, spatial and nonspatial data that can be imputed into a **geodatabase**. A **spatial query** requires knowledge of locations and attributes of that location. For example, an environmental analyst might want to know which public drinking water sources are located within one mile of a known toxic chemical spill. Alternatively, a planner might be called upon to identify property parcels in areas subject to flooding.

Defining a Geographic Information System

GIS consists of a particular computer program capable of storing, editing, processing, and presenting geographic data and information as maps from a software perspective. Several GIS software providers, such as Environmental Systems Research Institute Inc. (Esri), distribute the ArcGIS platform. Though companies like Google provide online mapping services and interfaces like Google Earth and Google Maps, such services are not currently considered fully-fledged GIS platforms. There are also open-source GIS options, such as QGIS, and OpenStreetMap, which are freely distributed and maintained by the open-source community. All GIS software, regardless of vendor, consists of a database management system capable of handling and integrating two types of data: spatial and attribute data.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=30#oembed-3

Spatial data refer to real-world geographic objects of interest, such as streets, buildings, lakes, countries, and locations. In addition to location, each object also possesses certain traits of interest, or attributes, such as a name, number of stories, depth, or population. GIS software keeps track of the spatial and attribute data and permits us to link the two types of data together to create information and facilitate analysis. For example, one way to describe and visualize a GIS is by picturing it as a cake with many layers. Each layer of the cake represents a different geographic theme and geodatabase, such as water features, buildings, and roads, and each layer is stacked on top of another.

As hardware, a GIS consists of a computer, memory, storage devices, scanners, printers, global positioning system (GPS) units, and other physical components. Suppose the computer is situated on a network. In that case, the network can also be considered an integral component of the GIS because it enables us to share data and information that the GIS uses as inputs and creates as outputs.

A GIS permits us to maintain, analyze, and share a wealth of data and information as a tool. GIS is used across the public and private sectors, from mapping the path of a hurricane to the more complex task of determining the most efficient U.S. Postal Service routes in a city. Online and mobile mapping, navigation, and location-based services also personalize and democratize GIS by bringing maps and mapping to the masses.

These are just a few definitions of a GIS. Like several geographic concepts discussed previously, there is no single or universally accepted definition of a GIS. There are just as many definitions of GIS as people who use GIS. In this regard, people like you learn, apply, develop, and study GIS in new and compelling ways that unify it.

Three Approaches to GIS

In addition to recognizing the many definitions of a GIS, it is also constructive to identify three general and overlapping approaches to understanding GIS – the *application* approach, the *developer* approach, and the *scientific* approach. Though most GIS users would identify with one approach more than another, they are not mutually exclusive. As GIS and, more generally, information technology advance, the following categories will be transformed and reshaped accordingly.

Application Approach

The application approach to GIS considers a GIS primarily to be a tool. This is also perhaps the most common view of a GIS. From this perspective, a GIS is used to answer questions, support decision-making, maintain an inventory of geographic data and information, and make maps. As a tool, specific skills should be acquired and required to use and apply a GIS properly. The application approach to a GIS is more concerned with using and applying GIS to solve problems than the GIS itself.

For instance, suppose we want to determine the best location for a new supermarket. What factors are essential behind making this decision? Information about neighborhood demographics, existing supermarkets, the location of suppliers, zoning regulations, and available real estate are all critical to this decision. A GIS platform can integrate information from the census bureau, realtors, the local zoning agency, and even the Internet. A suitability analysis can then be conducted with the GIS, showing the best locations for the supermarket given the various local geographic opportunities (e.g., demographics/consumers) and constraints (e.g., supply chain, zoning, and real estate limitations).

Several professional communities and organizations, such as the Urban and Regional Information Systems Association (URISA) and the United States Census Bureau, are concerned with using and applying a GIS.

Developer Approach

Unlike the previous example in which a GIS is applied to answer or solve a particular question, the developer approach to GIS is concerned with developing the GIS as a software or technology platform. Rather than focusing on how a GIS is used and applied, the developer approach is concerned with improving, refining, and extending the tool and technology and is in the realm of computer programmers and software developers.

The ongoing integration and evolution of GIS, maps, the Internet, and web-based mapping can be considered an outcome of the developer approach to GIS. Delivering maps, navigation tools, and user-friendly GIS to people via the Internet is the central challenge. The domain of GIS programmers and developers is the underlying logic and computer code that permits us to ask questions about getting from point A to point B on a navigation website or seeing where a new restaurant or open house is located on a web-based map. The Open-Source Geospatial Foundation is another example of a community of GIS developers working to build and distribute open-source GIS software.

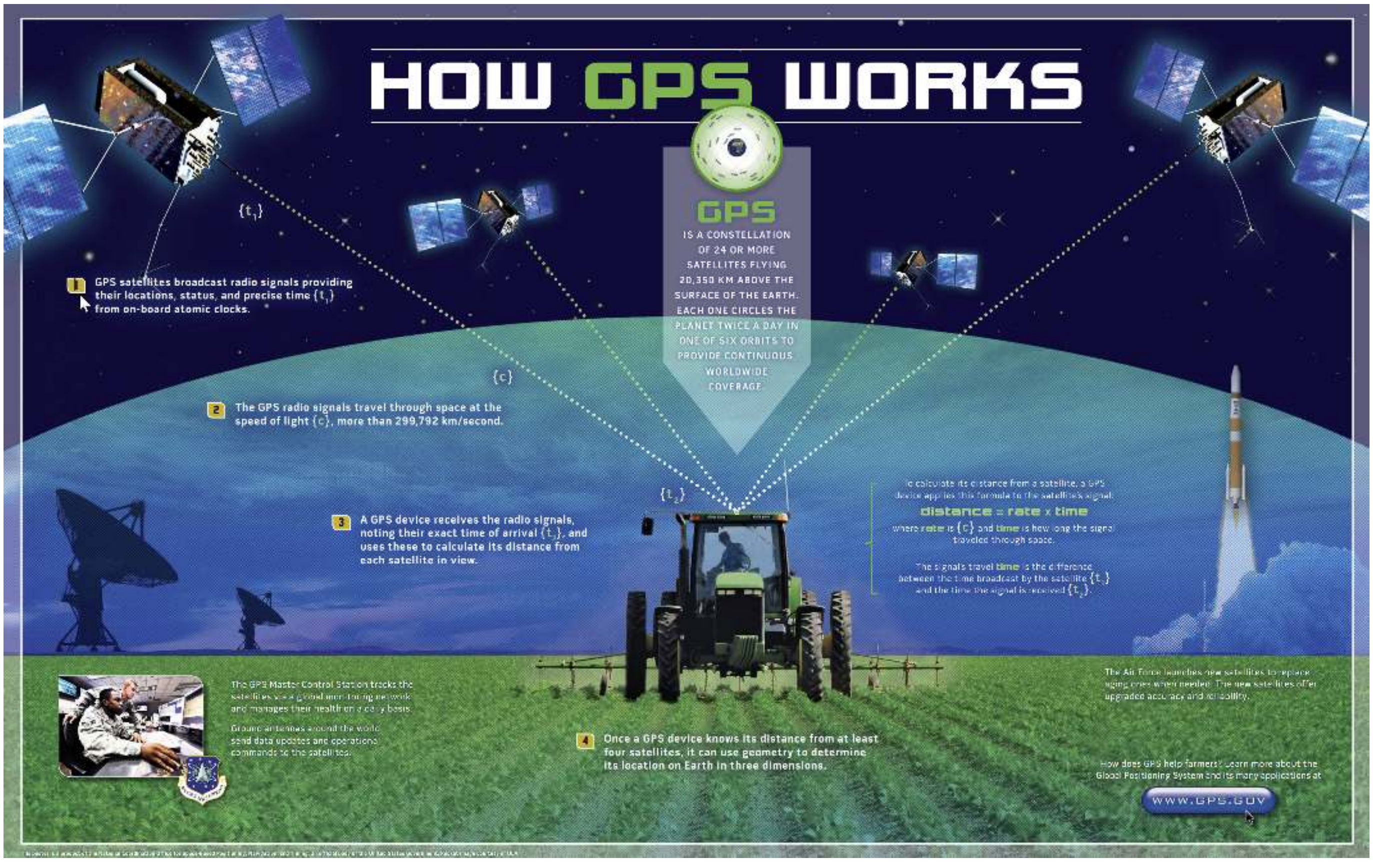
The developer approach to GIS drives and introduces innovation and is informed and guided by the application approach's existing needs and future demands. It is indeed at the forefront; it is dynamic and represents an area for considerable growth in the future.

Scientific Approach – GIScience

The scientific approach to GIS not only unites with the applications and developer approaches and is more concerned with broader questions and how geography, cognition, map interpretation, and other geospatial issues such as accuracy and errors are relevant to GIS and vice versa. This approach is often referred to as **geographic information science** (GIScience), and it is also interested in the social consequences and implications of the use and diffusion of GIS technology. From exploring the propagation of error to examining how GIS and related technology are redefining privacy, GIScience is, at the same time, an agent of change and understanding.

Considering the rapid rate of technological and GIS innovation, in conjunction with the widespread application of GIS, new questions about GIS technology and its use are continually emerging. One of the most discussed topics concerns privacy and what is referred to as locational privacy. Location privacy was of little concern to society. However, with the advent of GPS and its integration into cars and other mobile devices, questions, debates, and even lawsuits concerning locational privacy and who has the right to such information are rapidly emerging.

Global Positioning Systems



The use of location-based technologies has reached unprecedented levels. **Global Positioning Systems** (GPS) has been around since the 1980s when the military developed it. It slowly became mainstream with hiking and vehicle devices. However, in the 2010s, GPS has been going through a new revolution with the technology now in smartphones, cameras, microchips for pets, and transportation companies such as Uber and Lyft.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=30#oembed-4

GPS is a satellite-based navigation system made up of a network of twenty-four satellites placed into orbit by the U.S. Department of Defense. GPS was originally intended for military applications, but in the 1980s, the government made the system available for civilian use. GPS works in any weather conditions, anywhere globally, 24 hours a day.

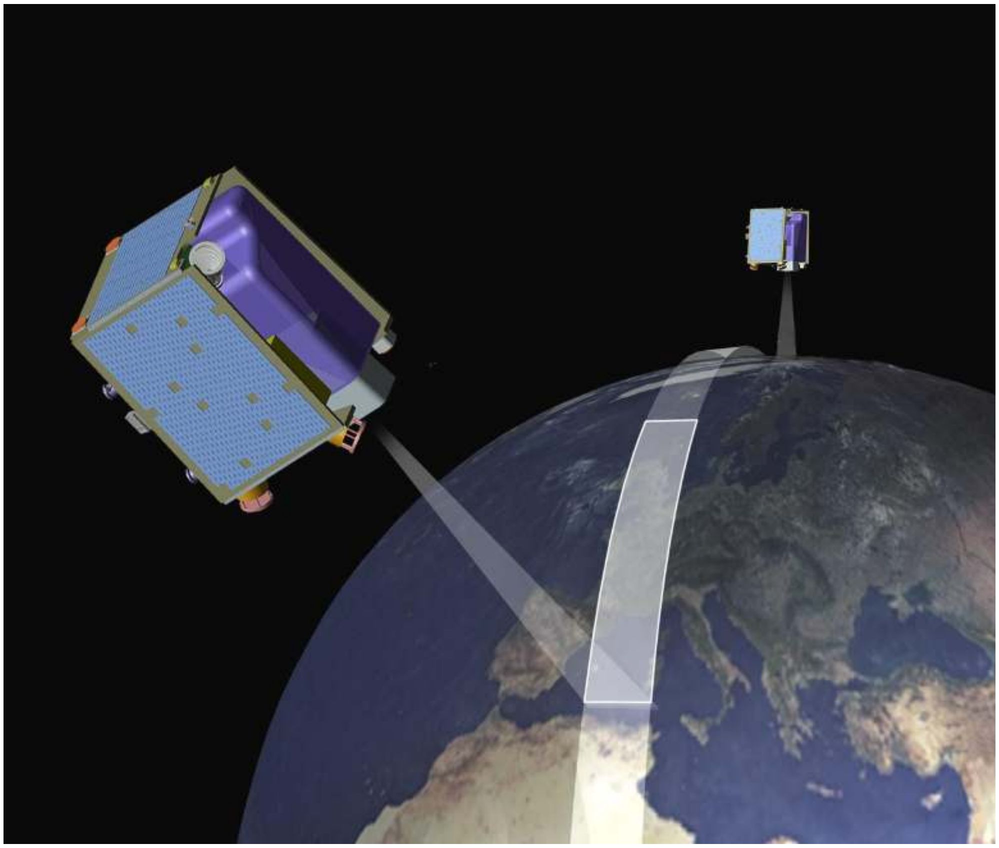
Using GPS to determine your location is not particularly useful if you do not know your landscape. For instance, your GPS could tell you that you are in the mall, but you may not know how to get to the door without a map.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=30#oembed-5

Concisely, GPS works like this: satellites circle the Earth twice a day in a very and transmit a signal to Earth. GPS receivers (or smartphones and watches) take this information and use trilateration to calculate the user's exact location. With distance measurements from a few satellites, the receiver can determine the user's position and display it on the unit's electronic map.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=30#oembed-6

Remote Sensing



“Rapid Eye Satellites” is licensed under Creative Commons Attribution-Share Alike 3.0 Unported.

The distance between the object and observer can be considerable, for example, imaging from the newly launched James Web Space Telescope, or minor, as in microscopes for examining bacterial growth. In geography, **remote sensing** takes on a specific connotation dealing with space-borne and aerial imaging systems used to remotely sense electromagnetic radiation reflected and emitted from Earth’s surface.

Remote sensing systems work the same way as a desktop scanner connected to a personal computer. A desktop scanner creates a digital image of a document by recording the intensity of light reflected from the document, pixel by pixel. For example, color scanners may have three light sources and three sets of sensors for visible light’s blue, green, and red wavelengths. Like the images, a desktop scanner produces, remotely sensed data consists of reflectance values arrayed in rows and columns that make up raster grids.

Remote sensing is used to solve a host of problems across various disciplines. For example, Landsat imagery monitors plant health and foliar changes. In contrast, the imagery produced by IKONOS is used for geospatial intelligence applications (yes, that means spying) and monitoring urban infrastructure. Other satellites, such as AVHRR (Advanced High-Resolution Radiometer), monitor the effects of global warming on vegetation patterns on a global scale. The MODIS (Moderate Resolution Imaging Spectroradiometer) *Terra* and *Aqua* sensors are designed to monitor atmospheric and oceanic composition in addition to the typical terrestrial applications. Today, **Unmanned Aerial Systems** (UAS), also called *drones*, extended and mainstreamed aerial imagery collection to the masses.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=30#oembed-7

Click the “**Previous**” button on the lower left or the ‘**Next**” button on the lower right to navigate throughout the textbook.

1.4: Geographic Concepts

Scale

The actual location size is reduced when representing the Earth on a manageable map. **Scale** is the ratio between the distance between two locations on a map and the corresponding distance on Earth's surface. A 1:1000 scale map, for example, would mean that one meter on the map equals 1000 meters, or one kilometer, on Earth's surface. Scale can sometimes be confusing, so it is important to remember that it refers to a ratio. It does not refer to the size of the map itself but rather how zoomed in or out the map is. A 1:1 scale map of your room would be the same size as yours – plenty of room for considerable detail but hard to fit into a glove compartment.

As with map projections, the "best" scale for a map depends on its use. If you are going on a walking tour of a historic town, a 1:5,000 scale map is commonly used. If you are a geography student looking at a map of the entire world, a 1:50,000,000 scale map would be appropriate. "Large" and "small" scales refer to the ratio, not the size of the landmass on the map. One divided by 5,000 is 0.0002, a larger number than one divided by 50,000,000 (which is 0.00000002). Thus, a 1:5,000 scale map is considered "large" scale while 1:50,000,000 is considered "small" scale.

Location

The concept distinguishing geography from other fields is location, which is central to a GIS. **Location** is simply a position on the surface of the Earth. What is more, everything can be assigned a geographic location. Once we know the location of something, we can put it on a map, for example, with a GIS.

We tend to define and describe locations in nominal or absolute terms. In the case of the former, locations are simply defined and described by name. For example, city names such as New York, Tokyo, or London refer to nominal locations. **Toponymy**, or the study of place names and their respective history and meanings, is concerned with nominal locations.

Though we tend to associate the notion of location with points on the Earth's surface, locations can also refer to geographic features (e.g., Rocky Mountains) or large areas (e.g., Siberia). The United States Board on Geographic Names maintains geographic naming standards and keeps track of such names through the Geographic Names Information Systems (GNIS). The GNIS database also provides information about which state and county the feature is in and its geographic coordinates.

Contrasting nominal locations are absolute locations that use a *reference system* to define positions on the Earth's surface. For instance, defining a location on the Earth's surface using latitude and longitude is an example of an **absolute location**. Postal codes and street addresses are other examples of absolute locations that usually follow local logic. Though there is no global standard for street addresses, we can determine the geographic coordinates (i.e., latitude and longitude) of street addresses, zip codes, place names, and other geographic data through geocoding.

Location can also be defined in relative terms. **Relative location** refers to defining and describing places about other known locations. For instance, Cairo, Egypt, is north of Johannesburg, South Africa; New Zealand is southeast of Australia; and Kabul, Afghanistan, is northwest of Lahore, Pakistan. Unlike nominal or absolute locations that define single points, relative locations provide more information and situate one place concerning another.

Direction

Like location, the concept of direction is central to geography and GIS. **Direction** refers to the position of something relative to something else, usually along a line. A **reference point** or benchmark from which direction will be measured must be established to determine direction. One of the most common benchmarks used to determine direction is ourselves. **Egocentric direction** refers to when we use ourselves as a directional benchmark. Describing something as "to my left," "behind me," or "next to me" are examples of egocentric direction.

As the name suggests, **landmark direction** uses a known landmark or geographic feature as a benchmark to determine direction. Such landmarks may be a busy city intersection, a prominent point of interest like the Colosseum in Rome, or some other feature like a mountain range or river. The critical thing to remember about landmark direction, especially when providing directions, is that the landmark should be well-known.

In geography and GIS, three more standard benchmarks are used to define the directions of true north, magnetic north, and grid north. **True North** is based on the point at which the axis of the Earth's rotation intersects the Earth's surface. In this respect, the North and South Poles serve as the geographic benchmarks for determining direction. **Magnetic North** (and south) refers to the point on the Earth's surface where the Earth's magnetic fields converge. This is also the point to which magnetic compasses point. Note that magnetic North falls in northern Canada and is not geographically coincident with true North or the North Pole. **Grid north** simply refers to the northward direction that the grid lines of latitude and longitude on a map called a graticule point to.

Distance

Complementing the concepts of location and direction is distance. **Distance** is the degree or amount of separation between locations and can be measured in nominal or absolute terms with various units. For example, we can describe the distances between locations nominally as "large" or "small," or we can describe two or more locations as "near" or "far apart."

Calculating the distance between two locations on the Earth's surface can be quite involving because we are dealing with a three-dimensional object. However, moving from the three-dimensional Earth to two-dimensional maps on paper, computer screens, and mobile devices is not trivial and is discussed in greater detail later.

We also use a variety of units to measure distance. For instance, the distance between London and Singapore can be measured in miles, kilometers, flight time on a jumbo jet, or days on a cargo ship. Whether or not such distances make London and Singapore "near" or "far" from each other is a matter of opinion, experience, and patience. Hence the use of absolute distance metrics, such as that derived from the distance formula, provide a standardized method to measure how far or close locations are from each other.

Space

Where distance suggests a measurable quantity in terms of how far apart locations are situated, space is a more abstract concept that is more commonly described than measured. For example, space can be described as "empty," "public," or "private."

Within the scope of a GIS, we are interested in **space**, and we are interested in what fills particular spaces and how and why things are distributed across space. Space is an ambiguous and generic term used to denote the general geographic area of interest.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=32#oembed-1

One kind of space relevant to a GIS is topological space. **Topological space** concerns the relationships between nature and location connectivity within a given space. What is essential within topological space are (1) how locations are (or are not) related or connected and (2) the rules that govern such geographic relationships.

Transportation maps for subways provide some of the best illustrations of topological spaces. When using maps, we are primarily concerned with how to get from one stop to another along with a transportation network. Specific rules also govern how we can travel along the network (e.g., transferring lines is possible only at a few key stops; we can travel only one direction on a particular line). Such maps may be of little use when traveling around a city by car or foot. However, they show the local transportation network and how locations are linked together effectively and efficiently.

Navigation

Like those discussed previously, transportation maps illustrate how we move through the environments where we live, work, and play. This movement and destination-oriented travel are referred to as **navigation**. How we navigate space is a complex process that blends our various motor skills, technology, mental maps, and awareness of locations, distances, directions, and the space we live in. Our geographical knowledge and spatial awareness are continuously updated and changed as we move from one location.

The acquisition of geographic knowledge is a lifelong endeavor. Though several factors influence the nature of such knowledge, we tend to rely on the three following types of geographic knowledge when navigating through space:

- Landmark knowledge refers to our ability to locate and identify unique points, patterns, or features (e.g., landmarks) in space.
- Route knowledge lets us connect and travel between landmarks by moving through space.
- Survey knowledge enables us to understand where landmarks are concerning each other and take shortcuts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=32#oembed-2

Each type of geographic knowledge is acquired in stages, one after the other. For instance, when we find ourselves in a new or unfamiliar location, we usually identify a few unique points of interest (e.g., hotel, building, fountain) to orient ourselves. Next, we are building up our landmark knowledge. Using and traveling between these landmarks develops our route knowledge and reinforces our landmark knowledge and overall geographical awareness. Finally, survey knowledge develops once we understand how routes connect landmarks and locations in space. At this point, when we are comfortable with our survey knowledge, we can take shortcuts from one location to another. Though there is no guarantee that a shortcut will be successful, we are at least expanding our local geographic knowledge if we get lost.

Landmark, route, and survey knowledge are the cornerstones of having a sense of direction and framing our geographical learning and awareness. While some would argue that they are born with a good sense of direction, others admit to always getting lost. The popularity of personal navigation devices and online mapping services speaks to the overwhelming desire to know and situate where we are in the world. Though developing and maintaining a keen sense of direction matters less and less as such devices and services continue to develop and spread, it can also be argued that the more we know about where we are in the world, the more we will want to learn about it.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=32#oembed-3

Click the “**Previous**” button on the lower left or the ‘**Next**” button on the lower right to navigate throughout the textbook.

1.5: Map Fundamentals

A **map** can be defined as a graphic representation of the real world. Because of the infinite nature of our universe, it is impossible to capture all the complexity found in the real world. For example, topographic maps abstract the three-dimensional real world at a reduced scale on a two-dimensional plane of the paper.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=34#oembed-1

Maps are used to display both cultural and physical features of the environment. Standard topographic maps, also called **reference maps**, show various information, including roads, land-use classification, elevation, rivers and other water bodies, political boundaries, and the identification of houses and other types of buildings. With the growth of GIS, thematic maps are becoming quite common. As the name implies, **thematic maps** focus on a particular theme, such as U.S. Census data, COVID-19 cases, and deaths, shifting temperatures because of climate change, biodiversity hotspots, or loss, are just examples. Esri has created a new commons database called The Living Atlas, allowing GIS users to upload dynamic maps into a "living atlas of the world."

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=34#oembed-2

Great and Small Circles

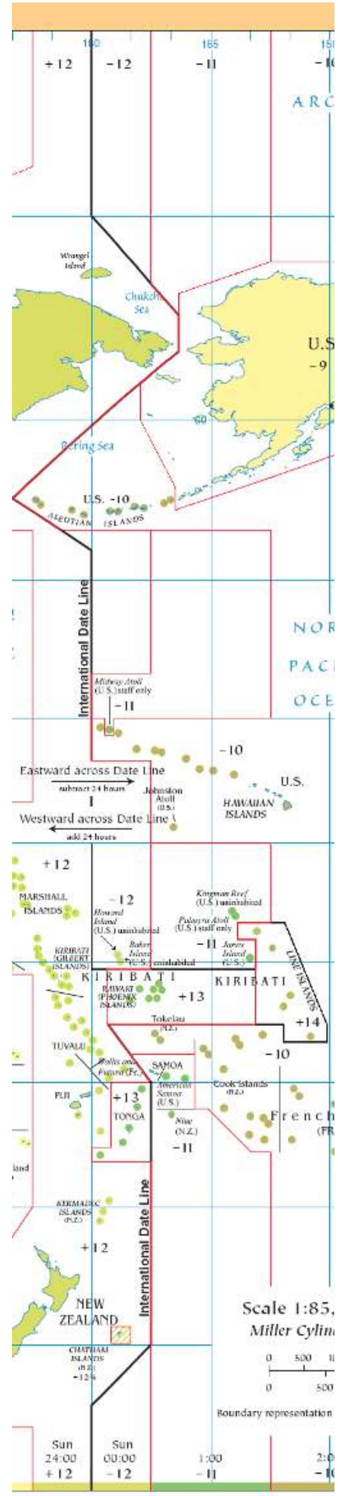
Much of Earth's grid system is based on the location of the North Pole, South Pole, and the Equator. The poles are an imaginary line running from the axis of Earth's rotation. The **Plane of the Equator** is an imaginary horizontal line that cuts the Earth into two halves. This brings up the topic of great and small circles. A **great circle** is any circle that divides the Earth into a circumference of two halves. It is also the largest circle that can be drawn on a sphere. The line connecting any points along a great circle is also the shortest distance between those two points.

Examples of great circles include the **Equator**, all lines of longitude, the line that divides the Earth into day and night called the circle of illumination, and the plane of the ecliptic, which divides the Earth into equal halves along the Equator. **Small circles** are circles that cut the Earth but not into equal halves. All lines of latitude, except the Equator, are made up of small circles.

Time Zones

Before the late nineteenth century, timekeeping was primarily a local phenomenon. Each town would set its clocks according to the motions of the Sun. For example, noon was defined as the time when the Sun reached its maximum altitude above the horizon. Cities and towns would assign a clockmaker to calibrate a town clock to these solar motions. This town clock would represent "official" time, and the citizens would set their watches and clocks accordingly.

The ladder half of the nineteenth century was a time of increased movement of humans. In the United States and Canada, large numbers of people were moving west, and settlements in these areas began expanding rapidly. Railroads moved people and resources between the various cities and towns to support these new settlements. However, the railroads experienced significant problems constructing timetables for the various stops because of how local time was kept. Timetables could only become more efficient if the towns and cities adopted some standard method of keeping time.



In 1878, Canadian Sir Sanford Fleming suggested a system of worldwide time zones that would simplify the keeping of time across the Earth. Fleming proposed dividing the globe into twenty-four time zones, every 15 degrees of longitude in width. Since the world rotates once every 24 hours on its axis and there are 360 degrees of longitude, each hour of Earth rotation represents 15 degrees of longitude.

Railroad companies in Canada and the United States began using Fleming's time zones in 1883. In 1884, an International Prime Meridian Conference was held in Washington D.C. to adopt the standardized timekeeping method and determine the prime Meridian's location. Conference members agreed that the longitude of Greenwich, England, would become zero degrees longitude and established the twenty-four time zones relative to the Prime Meridian. It was also proposed that the time measurement on the Earth would be made relative to the astronomical measurements at the Royal Observatory at Greenwich. This time standard was called Greenwich Mean Time (GMT).

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=34#oembed-3

Today, many nations operate on variations of the time zones suggested by Sir Fleming. This system measures time in the various zones relative to the Coordinated Universal Time (UTC) standard at the Prime Meridian. **Coordinated Universal Time** became the standard legal reference of time worldwide in 1972. UTC is determined from atomic clocks coordinated by the **International Bureau of Weights and Measures** (BIPM) in France. The numbers at the bottom of the time zone map indicate how many hours each zone is earlier (negative sign) or later (positive sign) than the Coordinated Universal Time standard. Also, note that national boundaries and political matters influence the shape of the time zone boundaries. For example, China uses a single time zone (eight hours ahead of Coordinated Universal Time) instead of five different time zones.

Geographic Coordinate Systems

Two types of coordinate systems are currently widely used in geography: the *geographical coordinate system* and the *rectangular* (also called *Cartesian*) *coordinate system*. The **geographic coordinate system** measures location from only two values, even though the locations are described for a three-dimensional surface. The two values that define location are measured relative to the Earth's polar axis. The two measures used in the geographic coordinate system are latitude and longitude.

Latitude and Longitude

Latitude is an angular measurement north or south of the Equator relative to a point found at the center of the Earth. This **central point** is also located on the Earth's rotational or polar axis. The Equator is the starting point for the measurement of latitude. The Equator has a value of zero degrees. A line of latitude or parallel of 30° North has an angle of 30° north of the plane represented by the Equator. The maximum latitude value can attain 90° North or South. These lines of latitude run parallel to the rotational axis of the Earth.

image

"Latitude and Longitude of the Earth" is licensed under the Creative Commons CC0 1.0 Universal Public Domain Dedication."

Lines connecting points of the same latitude, called **parallels**, have parallel lines. The only parallel that is also a great circle is the Equator. All other parallels are small circles. The following are the essential parallel lines:

- Equator, 0 degrees
- Tropic of Cancer, 23.5 degrees N
- Tropic of Capricorn, 23.5 degrees S
- Arctic Circle, 66.5 degrees N
- Antarctic Circle, 66.5 degrees S
- North Pole, 90 degrees N (infinitely small circle)
- South Pole, 90 degrees S (infinitely small circle)

Longitude is the angular measurement east and west of the Prime Meridian. Lines connecting points of the same longitude are called **meridians**. The position of the **Prime Meridian** was determined by international agreement to be in line with the location of the former astronomical observatory at Greenwich, England. Because the Earth's circumference is like a circle, it was decided to measure longitude in degrees. The number of degrees found in a circle is 360. The Prime Meridian has a value of zero degrees. A line of longitude or Meridian of 45° West has an angle of 45° west of the plane represented by the Prime Meridian. The maximum value that a meridian of longitude can have is 180°, the distance halfway around a circle. This meridian is called the **International Date Line**. The line determines where the new day begins in the world. Because of this, the International Date Line is not a straight line; it follows national borders so that a country is not divided into two separate days.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=34#oembed-4

When parallel and meridian lines are combined, a **geographic grid system** allows users to determine their exact location on the planet.

Coordinate Systems and Map Projections

Depicting the Earth's three-dimensional surface on a two-dimensional map creates various distortions involving distance, area, and direction. It is possible to create maps that are equidistance. However, even these types of maps have some form of distance distortion. Equidistance maps can only control distortion along either latitude or longitude. Distance is often correct on equidistance maps only in the direction of latitude.

Distance distortion is usually insignificant on a map with a large scale, 1:125,000 or larger. An example of a large-scale map is a standard topographic map. On these maps measuring straight line distance is simple. Distance is first measured on the map using a ruler. Then, the map's scale converts this measurement into a real-world distance. For example, if we measured ten centimeters on a map with a scale of 1:10,000, we would multiply 10 (distance) by 10,000 (scale). Thus, the actual distance in the real world would be 100,000 centimeters.

Measuring distance along map features that are not straight is more complicated. One technique employed for this task is to use several straight-line segments. This method's accuracy depends on the number of straight-line segments used. Another method for measuring curvilinear map distances is to use a mechanical device called an opisometer. This device uses a small rotating wheel that records the distance traveled. The recorded distance is measured by this device either in centimeters or inches.

Historically, most maps were hand-drawn, but with the advent of computer technology, more advanced maps were created with satellite technology. Geographic information science (GIS), sometimes referred to as geographic information systems, uses computers and satellite imagery to capture, store, manipulate, analyze, manage, and present spatial data. GIS primarily uses layers of information and is often used to make decisions in various contexts. For example, an urban planner might use GIS to determine the best location for a new fire station, while a biologist might use GIS to map the migratory paths of birds. In addition, we use GIS to get navigation directions from one place to another, layering place names, buildings, and roads.

One difficulty with map-making is that the Earth is a sphere while maps are flat even with advanced technology. As a result, some distortion always occurs when converting the spherical Earth to a flat map. A **map projection**, or a representation of Earth's surface on a flat plane, always distorts at least one of these four properties: *area*, *shape*, *distance*, and *direction*. Some maps preserve three of these properties while significantly distorting another, while other maps seek to minimize overall distortion but distort each property. So, which map projection is best? That depends on the purpose of the map. For example, while significantly distorting the size of places near the poles, the **Mercator Projection** preserves angles and shapes, making it ideal for navigation. The **Winkel Tripel Projection** is so named because its creator, Oswald Winkel, sought to minimize three kinds of distortion: area, direction, and distance. The National Geographic Society has used it since 1998 as the standard projection of world maps.

1.6: Cartographic Fundamentals

Have you ever found driving directions and maps online, used a smartphone to 'check in' to your favorite restaurant, or entered a town name or zip code to retrieve the local weather forecast? Every time you and millions of other users perform these tasks, you use Geographic Information Science (GIScience) and related spatial technologies. Many of these technologies, such as Global Positioning Systems (GPS) and in-vehicle navigation units, are very well-known, and you can recall the last time you used them.

Other applications and services that are the products of GIScience are a little less obvious, but they are every bit as common. For example, you use geospatial technologies if you are connected to the Internet. A geographic lookup occurs whenever your browser requests a web page from a Content Delivery Network (CDN). The server you are connected to contacts other servers closest to it and retrieves the information. This happens so that the delay between your request to view the data and the data sent to you is as short as possible.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=36#oembed-1

GIScience and related technologies are everywhere, and we use them every day. However, when it comes to information, "spatial is special." Reliance on spatial attributes separates **geographic information** from other types of information. There are several distinguishing properties of geographic information. Understanding them and their implications for the practice of geographic information science is a key to utilizing **geographic data**.

- Geographic data represent spatial locations and nonspatial attributes measured at certain times.
- Geographic space is continuous.
- Geographic space is nearly spherical.
- Geographic data tend to be spatially dependent.

Spatial attributes tell us where things are or where things were when the data were collected. Geographic data allow us to ask many geographic questions by merely including spatial attributes. For example, we might ask, "are gas prices in Puyallup high?" The interactive map from GasBuddy.com can help us with such a question while enabling us to generate many other spatial inquiries related to the geographic variation in fuel prices.

Another essential characteristic of geographic space is that it is "continuous." Although the Earth has valleys, canyons, caves, oceans, and more, there are no places on Earth without a location, and connections exist from one place to another. Outside of science fiction, there are no tears in the fabric of space-time. Modern technology can measure location precisely, making it possible to generate incredibly detailed depictions of geographic feature locations (e.g., of the coastline of the eastern U.S.). Unfortunately, it is often possible to measure so precisely that we collect more location data than we can store and much more than is helpful for practical applications. How much information is useful to store or display on a map will depend on the map scale (how much of the world we represent within a fixed display, such as the size of your computer screen) and the map's purpose.

In addition to being continuous, geographic data tend to be spatially dependent. More simply, "everything is related to everything else, but near things are more related than distant things" (which implies that things near one another tend to be more alike than things far apart). A statistical calculation known as spatial autocorrelation can measure how alike things are concerning their proximity to other things. Without this fundamental property, geographic information science as we know it today would not be possible.

Data Collection

Geographic data comes in many types, from diverse sources, and is captured using many techniques; they are collected, sold, and distributed by many public and private entities. In general, we can divide geographic data collection into two main types.

Directly collected data are generated at the source of the phenomena being measured. Examples of directly collected data include temperature readings at specific weather stations, elevations recorded by visiting the location of interest, or the position of a grizzly bear equipped with a GPS-enabled collar. Also included here are data derived from surveys (e.g., the census) or observation (e.g., Audubon Christmas bird count).

Remotely sensed data are measured remotely without direct contact with the phenomena or needing to visit the locations of interest. Satellite images, sonar readings, and radar are all forms of remotely sensed data.

Maps are both raw material and geographic information systems (GIS) products. All maps represent features and characteristics of locations, and that representation depends upon data relevant at a particular time. All maps are also selective; they do not show us everything about the place depicted; they show only the features and characteristics that their maker decided to include. Maps are often categorized into reference or thematic maps based upon the producer's decision about what to include and the expectations about how the map will be used. The prototypical reference map depicts the location of "things" usually visible worldwide; examples include road maps and topographic maps depicting terrain.

Thematic maps, in contrast, typically depict "themes." They are more abstract, involve more processing and interpretation of data, and often depict not directly visible concepts; examples include maps of income, health, climate, or ecological diversity. There is no clear-cut line between reference and thematic maps. However, the categories are helpful to recognize because they relate directly to how the maps are intended to be used and their cartographers' decisions in shrinking and abstracting aspects of the world to generate the map. Several types of thematic maps include:

- **Choropleth** – a thematic map that uses tones or colors to represent spatial data as average values per unit area
- **Proportional symbol –** uses symbols of assorted sizes to represent data associated with different areas or locations within the map
- **Isopleth –** also known as contour maps or isopleth maps, depict smooth continuous phenomena such as precipitation or elevation
- **Dot –** uses a dot symbol to show the presence of a feature or phenomenon – dot maps rely on a visual scatter to show a spatial pattern
- **Dasymetric** – an alternative to a choropleth map but instead of mapping the data so that the region appears uniform, ancillary information is used to model the internal distribution of the data

---

Click the "**Previous**" button on the lower left or the '**Next**" button on the lower right to navigate throughout the textbook.

1.7: Future of Mapping Technology

The definitions and approaches to GIS described above illustrate this information technology's scope and breadth. Furthermore, as GIS becomes more accessible and widely distributed, there will always be new questions to be answered, new applications to be developed, and innovative technologies to integrate.

One notable development is the emergence of web-based GIS or Web GIS. **Web GIS** refers to integrating the vast amounts of content available on the Internet (e.g., text, photographs, video, and music) with geographic information, such as location. Adding such geographic information to such content is called geotagging, like geocoding. Integrating geographic information with such content opens new ways to access, search, organize, share, and distribute information.

Web-based mapping applications combine data and information from one source and map it with online mapping applications There are mashups for everything that can be assigned a location, from restaurants and music festivals to your photographs and favorite hikes. GPS technology within smartphones has also revolutionized the way geographic data is collected and distributed worldwide.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=38#oembed-1

The diffusion of GIS and the emergence of Web GIS have increased geographic awareness by lowering the barriers to viewing, using, and even creating maps and related geographic data and information. Though there are several benefits to this democratization of GIS, and more generally, information and technology, it should also be recognized that there are consequences and implications.

As with any other technology, great care must be taken in using and applying GIS. For instance, when was the last time you questioned what appeared on a map? For better or worse, maps are among the most authoritative forms of information. As tomorrow's GIS practitioners, you can influence how decisions are made and how others view and relate to the world with the maps you create in a GIS environment. Therefore, what and how you choose to map is a nontrivial exercise. Becoming more aware of our biases, limitations, and preferences allows us to take advantage of geographic information systems confidently.

Click the "**Previous**" button on the lower left or the '**Next**" button on the lower right to navigate throughout the textbook.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

1.9: Attributions and References

Creative Commons Attributions

References

BRIGHT SIDE (Director). (2019, June 17). *How GPS Works Today*. https://www.youtube.com/watch?v=wCcARVbL_Dk

Bureau, U. C. (n.d.). *Census.gov*. Census.Gov. Retrieved August 20, 2022, from https://www.census.gov/en.html

Campbell, J., & Shin, M. (n.d.). *Essentials of Geographic Information Systems—Open Textbook Library*. Retrieved January 20, 2020, from https://open.umn.edu/opentextbooks/textbooks/essentials-of-geographic-information-systems

Casual Navigation (Director). (2018, August 17). *How Does GPS Work?* https://www.youtube.com/watch?v=U3eX6QKS9kY

CNET Highlights (Director). (2021, June 8). *New Apple Maps vs New Google Maps (Watch their latest reveals)*. https://www.youtube.com/watch?v=t2kcgX9VdjY

Competency Model Clearinghouse—Geospatial Technology Industry Competency Model. (n.d.). Retrieved August 20, 2022, from https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx

CrashCourse (Director). (2020a, December 14). *What is a Map? Crash Course Geography #2*. https://www.youtube.com/watch?v=iHEMOdRo5u8

CrashCourse (Director). (2020b, December 22). *What is space and how do we study it? Crash Course Geography #3*. https://www.youtube.com/watch?v=mtq2PG67w4E

CrashCourse (Director). (2022, April 12). *Geographies of the Future: Crash Course Geography #50*. https://www.youtube.com/watch?v=T_vLXgj2pvo

Dastrup, R. A., MA, & GISP. (2018). *Introduction to Physical Geography*. https://physicalgeog.pressbooks.com/

Dastrup, R. A., MA, & GISP. (2019). *Introduction to Human Geography*. https://humangeography.pressbooks.com/

DiBiase, D. (n.d.). *Nature of Geographic Information Systems—Open Textbook Library*. Retrieved January 20, 2020, from https://open.umn.edu/opentextbooks/textbooks/nature-of-geographic-information-systems

Earth Science Data Systems, N. (n.d.). *AVHRR*. Earthdata; Earth Science Data Systems, NASA. Retrieved August 20, 2022, from http://www.earthdata.nasa.gov/sensors/avhrr

Esri. (n.d.). Retrieved January 20, 2020, from https://www.esri.com/

Esri (Director). (2016, December 5). *What Is GIS?* https://www.youtube.com/watch?v=LHDCRjAxpI0

Esri Events (Director). (2021, December 3). *Latest & Greatest from ArcGIS Living Atlas of the World*. https://www.youtube.com/watch?v=iGHUSY6oYg4

Geographic Names Information System (GNIS) | U.S. Geological Survey. (n.d.). Retrieved August 20, 2022, from https://www.usgs.gov/tools/geographic-names-information-system-gnis

Geography | The Science of Our World. (n.d.). Retrieved August 20, 2022, from https://www.geography.com/

geospatialrev (Director). (2010, March 25). *The Geospatial Revolution Project—Trailer*. https://www.youtube.com/watch?v=mBlUwPcFjCw

GeoTech Center. (n.d.). GeoTech Center. Retrieved January 20, 2020, from http://www.geotechcenter.org/

GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri. (n.d.). Retrieved August 20, 2022, from https://www.esri.com/en-us/home

Google Earth. (n.d.). Retrieved August 20, 2022, from https://earth.google.com/web/@0,-18.0013,0a,22251752.77375655d,35y,0h,0t,0r

Google Earth (Director). (2021, April 15). *Our Cities | Timelapse in Google Earth*. https://www.youtube.com/watch?v=v74_mf2usc0

Google Maps. (n.d.). Retrieved August 20, 2022, from https://www.google.com/maps/@40.6927563,-111.5587427,15z

GPS: The Global Positioning System. (n.d.). Retrieved August 20, 2022, from https://www.gps.gov/

How Does GPS Work? (n.d.). Retrieved January 20, 2020, from https://www.youtube.com/watch?v=FU_pY2sTwTA&feature=emb_logo

How GPS Works. (n.d.). Retrieved January 20, 2020, from https://www.youtube.com/watch?v=IoRQiNFzT0k&feature=emb_logo

IKONOS-2—Earth Online. (n.d.). Retrieved August 20, 2022, from https://earth.esa.int/eogateway/missions/ikonos-2

Johnny Harris (Director). (2020, May 26). *The International Date Line, Explained*. https://www.youtube.com/watch?v=aBppb2quqkE

Kevin Stratvert (Director). (2022, March 9). *Top 20 Google Maps Tips & Tricks: All the best features you should know!* https://www.youtube.com/watch?v=beeNMoXuxPg

Landsat Science. (n.d.). Retrieved August 20, 2022, from https://landsat.gsfc.nasa.gov/

MODIS Web. (n.d.). Retrieved August 20, 2022, from https://modis.gsfc.nasa.gov/

National Geographic. (n.d.). National Geographic. Retrieved August 20, 2022, from https://www.nationalgeographic.com

National Geospatial-Intelligence Agency (Director). (2020, November 3). *NGA Explains: What is Cartography?* https://www.youtube.com/watch?v=fSbBmUxYgEs

Open Source Geospatial Foundation. (n.d.). OSGeo. Retrieved August 20, 2022, from https://www.osgeo.org/

OpenStreetMap. (n.d.). OpenStreetMap. Retrieved August 20, 2022, from https://www.openstreetmap.org/

QGIS. (n.d.). Retrieved January 20, 2020, from http://qgis.com/

TED-Ed (Director). (2015, January 29). *How does your smartphone know your location? – Wilton L. Virgo*. https://www.youtube.com/watch?v=70cDSUI4XKE

TEDx Talks (Director). (2017, December 13). *Want to learn better? Start mind mapping | Hazel Wagner | TEDxNaperville*. https://www.youtube.com/watch?v=5nTuScU70As

The 60-second case against time zones. (n.d.). Retrieved January 20, 2020, from https://www.youtube.com/watch?v=c3EQDfqqHsI&feature=emb_logo

The Importance of Maps. (n.d.). Retrieved January 20, 2020, from https://www.youtube.com/watch?time_continue=1&v=mR2CWUOSC08&feature=emb_logo

URISA. (n.d.). Retrieved August 20, 2022, from https://www.urisa.org/

U.S. Board on Geographic Names. (n.d.). Retrieved January 20, 2020, from https://www.usgs.gov/core-science-systems/ngp/board-on-geographic-names

U.S. Board on Geographic Names | U.S. Geological Survey. (n.d.). Retrieved August 20, 2022, from https://www.usgs.gov/us-board-on-geographic-names

vlogbrothers (Director). (2013, September 20). *42 Amazing Maps*. https://www.youtube.com/watch?v=dldHalRY-hY

Webb Space Telescope GSFC/NASA. (n.d.). Retrieved August 20, 2022, from https://www.jwst.nasa.gov/

Welcome to the QGIS project! (n.d.). Retrieved August 20, 2022, from https://www.qgis.org/en/site/

What is Geography? Crash Course Geography. (n.d.). Retrieved August 20, 2022, from https://www.youtube.com/watch?v=93LLwiMjDko

What Is GIS? (n.d.-a). Retrieved January 20, 2020, from https://www.youtube.com/watch?v=LHDCRjAxpI0&feature=emb_logo

What is GIS? (n.d.-b). Retrieved January 20, 2020, from https://www.esri.com/en-us/what-is-gis/overview

What is Remote Sensing? (n.d.-a). Retrieved January 20, 2020, from https://www.youtube.com/watch?v=xIsUP1Ds5Pg&feature=emb_logo

What is Remote Sensing? Understanding Remote Sensing. (n.d.-b). Retrieved January 20, 2020, from https://www.youtube.com/watch?v=N49PzLDUIFQ&feature=emb_logo

CHAPTER OVERVIEW

2: Spatially Representing Earth

2.1: Introduction

Maps and mapping are essential components of any geographic information system (GIS). For instance, maps constitute both the input and output of a GIS. Hence a GIS utilizes many concepts and themes from cartography, the formal study of maps and mapping. Therefore, in order for us to become proficient with GIS, we need to learn more about cartography, maps, and mapping. The first part of this chapter defines a map and describes a few key map types. Next, cartographic or mapping conventions are discussed, emphasizing map scale, coordinate systems, and map projections. The chapter concludes with a discussion of the process of map abstraction as it relates to GISs. This chapter provides the foundations for working with, integrating, and making maps with GIS.

Learning Outcomes

- Define a map and describe reference, thematic, and dynamic maps.
- Describe the concepts of map scale, coordinate systems, and map projections.
- Explain why they are central to maps, mapping, and geographic information systems.
- Highlight the decision-making process behind maps and underscore the need to be explicit and consistent when mapping and using geographic information systems.Add Chapter

GTCM Alignment

Chapter Sections

2.2: Fundamentals of Maps

Maps are among the most compelling forms of information for several reasons. Maps are artistic. Maps are scientific. Maps preserve history. Maps clarify. Maps reveal the invisible. Maps inform the future. Regardless of the reason, maps capture people's imaginations around the world. As one of the most trusted forms of information, map-makers and geographic information system (GIS) practitioners hold considerable power and influence. Therefore, understanding and appreciating maps and how maps convey information are essential aspects of GIS. The appreciation of maps begins with exploring various map types.

There are just as many definitions of maps as people who use and make them like GIS. We can define a map simply as a representation of the world. Such maps can be stored in our brains (i.e., mental maps), printed on paper, or online. Notwithstanding the actual medium of the map (e.g., our fleeting thoughts, paper, or digital display), maps represent and describe various aspects of the world. For clarity, the three types of maps are the reference map, the thematic map, and the dynamic map.

Reference Maps

The primary purpose of a **reference map** is to deliver location information to the map user. Therefore, geographic features and map elements on a reference map tend to be treated and represented equally. In other words, no single aspect of a reference map takes precedence over any other aspect. Moreover, reference maps represent geographic reality accurately. Examples of standard reference maps include topographic maps created by the United States Geological Survey (USGS) and image maps obtained from satellites or aircraft available through online mapping services.

The accuracy of a given reference map is indeed critical to many users. For instance, local governments need accurate reference maps for land use, zoning, and tax purposes. National governments need accurate reference maps for political, infrastructure, and military purposes. People who depend on navigation devices like global positioning systems (GPS) units also need accurate and up-to-date reference maps to arrive at their desired destinations.

Thematic Maps

Contrasting the reference map are thematic maps. As the name suggests, **thematic maps** are concerned with a particular theme or topic of interest. For example, while reference maps emphasize the location of geographic features, thematic maps are more concerned with how things are distributed across space. Such things are often abstract concepts such as life expectancy worldwide, per capita gross domestic product (GDP) in Europe, or literacy rates across India. One of the strengths of mapping and thematic mapping is that it can make such abstract and invisible concepts visible and comparable on a map.

It is important to note that reference and thematic maps are not mutually exclusive. In other words, thematic maps often contain and combine geographical reference information, and conversely, reference maps may contain thematic information. What is more, when used in conjunction, thematic and reference maps often complement each other.

For example, public health officials in a city may be interested in providing equal access to emergency rooms to the city's residents. Insights into this and related questions can be obtained through visual comparisons of a reference map that shows the locations of emergency rooms across the city to thematic maps of various segments of the population (e.g., households below poverty, percent elderly, underrepresented groups).

Within the context of a GIS, we can overlay the reference map of emergency rooms directly on top of the population maps to see whether access is uniform across neighborhood types. Of course, there are other factors to consider when looking at emergency room access (e.g., access to transport), but through such map overlays, underserved neighborhoods can be identified.

GIS Overlay Process

When presented in hardcopy format, both reference and thematic maps are static or fixed representations of reality. Such permanence on the page suggests that geography and the things that we map are also, in many ways, fixed or constant. This is far from reality. The integration of GIS with other forms of information technology like the Internet and mobile telecommunications is rapidly changing this view of maps and mapping and geography.

The diffusion of GIS and the popularity of online mapping tools and applications speak to this shift in thinking about maps and map use. In this regard, it is worthwhile to discuss the diffusion of dynamic maps. **Dynamic maps** are simply changeable or interactive representations of the earth. Dynamic mapping refers more to how maps are used and delivered to the map user today (e.g., online, via mobile phone) than to the map's content. Both reference and thematic maps can be dynamic, and such maps are an integral component of any GIS. The critical point about dynamic maps is that more people, not just GIS professionals, have access to such maps.

Unlike a hardcopy map with features and elements that users cannot modify or change, dynamic maps encourage and sometimes require user interaction. Such interaction can include changing the scale or visible area by zooming in or out, selecting which features or layers to include or remove from a map (e.g., roads, imagery), or even starting and stopping a map animation.

Just as dynamic maps will continue to evolve and require more user interaction in the future, map users will demand more interactive map features and controls. Furthermore, as this democratization of maps and mapping continues, map users' geographic awareness and appreciation will also increase. Therefore, it is critical to understand maps' nature, form, and content to support the changing needs, demands, and expectations of map users in the future.

2.3: Datums, Coordinate Systems, and Map Projections

All map users and viewers have certain expectations about what is contained on a map. Such expectations are formed and learned from previous experience by working with maps. However, it is essential to note that such expectations also change with increased exposure to maps. Understanding and meeting the expectations of map viewers is a challenging but necessary task because such expectations provide a starting point for creating any map.

The central purpose is to provide relevant and valuable information to the map user. For a map to be of value, it must convey information effectively and efficiently. Mapping conventions facilitate the delivery of information in such a manner by recognizing and managing the expectations of map users. Mapping or cartographic conventions refer to the accepted rules, norms, and practices behind the making of maps. For example, one of the most recognized mapping conventions is that “north is up” on most maps. Though this may not always be the case, many map users expect the north to be oriented or coincide with the top edge of a map or viewing device like a computer monitor.

Several other formal and informal mapping conventions and characteristics can be identified, many of which are taken for granted. The essential cartographic considerations are map scale, coordinate systems, and map projections. Map scale is concerned with reducing geographical features of interest to manageable proportions. Coordinate systems help us define the positions of features on the earth’s surface. Moreover, map projections are concerned with moving from the three-dimensional world to the two dimensions of a flat map or display.

Map Scale

One of the most significant challenges behind mapping the world and its resident features, patterns, and processes is reducing it to a manageable size. What exactly is meant by “manageable” is open to discussion and depends on the purpose and needs of the map at hand. Nonetheless, all maps reduce or shrink the world and its geographic features of interest by some factor. **Map scale** refers to the factor of reduction of the world so it fits on a map.

Map scale can be represented by text, a graphic, or some combination of the two. For example, it is common to see “one inch represents one kilometer” or something similar written on a map to give map users an idea of its scale. Map scale can also be portrayed graphically with a scale bar. Scale bars are usually used on reference maps and allow map users to approximate distances between locations and features on a map and get an overall idea of the map’s scale.

The representative fraction (RF) describes scale as a simple ratio. The numerator, which is always set to one (i.e., 1), denotes map distance, and the denominator denotes ground or “real-world” distance. One of the benefits of using a representative fraction to describe scale is unit neutral. In other words, any unit of measure can be used to interpret the map scale. For example, consider a map with an RF of 1:10,000. This means that one unit on the map represents 10,000 units on the ground. Such units could be inches, centimeters, or even pencil lengths; it does not matter.

Map scales can also be described as “small” or “large.” Such descriptions are usually made about representative fractions and the amount of detail represented on a map. For instance, a map with an RF of 1:1,000 is considered a large-scale map when compared to a map with an RF of 1:1,000,000 (i.e., 1:1,000 > 1:1,000,000). Furthermore, while the large-scale map shows more detail and less area, the small-scale map shows more area but less detail. Determining the thresholds for small- or large-scale maps is a judgment call.

All maps possess a scale, whether it is formally expressed or not. Though some say that online maps and GIS are “scaleless” because we can zoom in and out at will, it is more accurate to say that GIS and related mapping technology are multiscale. Therefore, understanding map scale and its overall impact on how the earth and its features are represented is critical for map-making and GIS.

Coordinate Systems

Just as all maps have a map scale, all maps have locations, too. **Coordinate systems** are frameworks that are used to define unique positions. For instance, in geometry, we use x (horizontal) and y (vertical) coordinates to define points on a two-dimensional plane. The coordinate system used to define locations on the three-dimensional earth is the geographic coordinate system (GCS), based on a sphere or spheroid. A spheroid (a.k.a. ellipsoid) is simply a slightly wider sphere than it is tall and approximates the actual shape of the earth more closely. Spheres are commonly used as models of the earth for simplicity.

The unit of measure in the GCS is degrees, and their respective latitude and longitude define locations within the GCS. Latitude is measured relative to the equator at zero degrees, with a maximum of ninety degrees north at the North Pole or ninety degrees south. Longitude is measured relative to the prime meridian at zero degrees, with a maximum of 180 degrees west or 180 degrees east.

Note that latitude and longitude can be expressed in **degrees-minutes-seconds** (DMS) or **decimal degrees** (DD). When using decimal degrees, latitudes above the equator and longitudes east of the prime meridian are positive, and latitudes below the equator and longitudes west of the prime meridian are negative (see the following table for examples).

Converting from DMS to DD is a straightforward exercise. For example, since there are sixty minutes in one degree, we can convert 118° 15 minutes to 118.25 (118 + 15/60). An online search of the term "coordinate conversion" will return several coordinate conversion tools.

When we want to map things like mountains, rivers, streets, and buildings, we need to define how the lines of latitude and longitude will be oriented and positioned on the sphere. A **datum** serves this purpose and specifies precisely the orientation and origins of the lines of latitude and longitude relative to the center of the earth or spheroid.

Depending on the need, situation, and location, there are several datums to choose from. For instance, local datums try to match the spheroid closely to the earth's surface in a local area and return accurate local coordinates. A typical local datum used in the United States is **NAD83** (i.e., North American Datum of 1983). For locations in the United States and Canada, NAD83 returns relatively accurate positions, but positional accuracy deteriorates when outside of North America.

The global **WGS84** datum (i.e., World Geodetic System of 1984) uses the center of the earth as the origin of the GCS and is used for defining locations across the globe. Because the datum uses the earth's center as its origin, locational measurements tend to be more consistent regardless of where they are obtained. However, they may be less accurate than those returned by a local datum. Note that switching between datums will alter the coordinates (i.e., latitude and longitude) for all locations of interest.

Map Projections

The earth is not flat or round but has a spherical shape called a **spheroid**. A globe is an excellent representation of the three-dimensional, spheroid earth. However, one of the problems with globes is that they are not very portable (i.e., you cannot fold a globe and put it in your pocket), and their small scale makes them of limited practical use (i.e., geographic detail is sacrificed). To overcome these issues, it is necessary to transform the three-dimensional shape of the earth into a two-dimensional surface like a flat piece of paper, computer screen, or mobile device display to obtain more helpful map forms and map scales. Enter the map projection.

Map projections refer to the methods and procedures used to transform the spherical three-dimensional earth into two-dimensional planar surfaces. Specifically, map projections are mathematical formulas used to translate latitude and longitude on the earth's surface to x and y coordinates on a plane. Since there is an infinite number of ways this translation can be performed, there is an infinite number of map projections.

To illustrate the concept of a map projection, imagine that we place a light bulb in the center of a translucent globe. On the globe are outlines of the continents and the lines of longitude and latitude called the graticule. When we turn the light bulb on, the outline of the continents and the graticule will be "projected" as shadows on the wall, ceiling, or any other nearby surface. This is what is meant by map "projection."

Within the realm of maps and mapping, there are three surfaces used for map projections (i.e., surfaces on which we project the shadows of the graticule). These surfaces are the plane, the cylinder, and the cone. Referring again to the previous example of a light bulb in the center of a globe, note that we can situate each surface in many ways during the projection process. For example, surfaces can be tangential to the globe along the equator or poles, and they can pass through or intersect the surface and be oriented at any number of angles.

Naming conventions for many map projections include the surface and its orientation. For example, as the name suggests, "planar" projections use the plane, "cylindrical" projections use cylinders, and "conic" projections use the cone. For cylindrical projections, the "normal" or "standard" aspect refers to when the cylinder is tangential to the equator (i.e., the axis of the cylinder is oriented north-south). When the cylinder's axis is perfectly oriented east-west, the aspect is called "transverse," All other orientations are called "oblique." Regardless of the orientation or the surface on which a projection is based, several distortions will be introduced that will influence the choice of map projection.

When moving from the three-dimensional surface of the earth to a two-dimensional plane, distortions are not only introduced but also inevitable. For example, map projections introduce distance, angles, and area distortions. Depending on the map's purpose, trade-offs will need to be made concerning such distortions.

Map projections that accurately represent distances are referred to as equidistant projections. Note that distances are only correct in one direction, usually running north-south, and are not correct everywhere across the map. Equidistant maps are frequently used for small-scale maps that cover large areas because they do an excellent job of preserving the shape of geographic features such as continents.

Maps that represent angles between locations, also referred to as bearings are called conformal. Conformal map projections maintain a bearing or head when traveling great distances. However, the cost of preserving bearings is that areas tend to be quite distorted in conformal map projections. Though shapes are preserved over small areas, areas become wildly distorted at small scales. The Mercator projection is an example of a conformal projection and is famous for distorting Greenland.

As the name indicates, **equal area** or **equivalent projections** preserve the area's quality. Such projections are of particular use when accurate measures or comparisons of geographical distributions are necessary (e.g., deforestation, wetlands). However, to maintain accurate proportions on the earth's surface, features sometimes become compressed or stretched depending on the orientation of the projection. Moreover, such projections distort distances as well as angular relationships.

As noted earlier, there is theoretically an infinite number of map projections to choose from. One of the key considerations behind the choice of map projection is to reduce the amount of distortion. The geographical object being mapped and the respective scale at which the map will be constructed are essential factors. For instance, maps of the North and South Poles usually use planar or azimuthal projections, and conical projections are best suited for the middle latitude areas of the earth. Features that stretch east-west, such as the country of Russia, are represented well with the standard cylindrical projection. In contrast, countries-oriented north-south (e.g., Chile, Norway) are better represented using a transverse projection.

If a map projection is unknown, sometimes it can be identified by working backward and carefully examining the nature and orientation of the graticule (i.e., grid of latitude and longitude) and the varying degrees of distortion. There are trade-offs made concerning distortion on every map. There are no fixed rules as to which distortions are more preferred over others. Therefore, the selection of map projection depends on the map's purpose.

Within the scope of GIS, knowing and understanding map projections are critical. For instance, to perform an overlay analysis like the earlier one, all map layers need to be in the same projection. If they are not, geographical features will not be appropriately aligned, and any analyses performed will be inaccurate. Most GIS include functions to assist in the identification of map projections, as well as to transform between projections to synchronize spatial data. Awareness of the potential pitfalls surrounding map projections is essential despite the technology capabilities.

2.4: Representing Geographic Features

Maps are a representation of the earth. Central to this representation is reducing the earth's features of interest to a manageable size (i.e., map scale) and its transformation into a functional two-dimensional form (i.e., map projection). The choice of both map scale and, to a lesser extent, map projection will influence the content and shape of the map.

In addition to the objective decisions made by cartographers and GIS users regarding map scale and map, projections are those concerning what to include and what to omit from the map. The purpose of a map will certainly guide some of these decisions, but other choices may be based on space limitations, map complexity, and desired accuracy. Furthermore, decisions about how to classify, simplify, or exaggerate features and how to symbolize objects of interest simultaneously fall under the realms of art and science.

Moving from the real world to the world of maps is **map abstraction**. This process involves making choices about how to represent features. More concerning geographic information systems (GIS), we must be explicit, consistent, and precise in defining and describing geographical features of interest. Failure to be explicit, consistent, and precise will return incorrect, inconsistent, and error-prone maps, analyses, and decisions based on such maps and GIS.

One of the most pressing environmental issues facing the world is deforestation. Deforestation refers to the reduction of forest area. This is an essential issue because it has implications for climate change, global warming, biodiversity, and the earth's water balance, among other things. In the last century, deforestation has increased at an alarming rate and is mainly attributed to human activity. Therefore, mapping forests regularly with a GIS is a logical way to monitor deforestation and has the potential to inform policies regarding forest conservation efforts. Easy enough, so let us get started.

So, what exactly is a forest? How do we know where a forest begins and where it ends? How can natural wildfires be differentiated from those started by humans? Can a forest exist in a swamp or wetland? For that matter, what is the difference between a swamp and a wetland? Such questions are not trivial in the context of mapping and GIS. Consistent and precise definitions of features like forests or swamps increase the reliability and efficiency of maps, mapping, and analysis with GIS.

The world comprises various features or entities within maps, cartography, and GIS. Such entities include but are not restricted to fire hydrants, caves, roads, rivers, lakes, hills, valleys, oceans, and the occasional barn. Moreover, such features have a form, and more precisely, a geometric form. For instance, fire hydrants and geysers are considered point-like features; rivers and streams are linear features; lakes, countries, and forests are areal features.

Features can also be categorized as either discrete or continuous. Discrete features are well defined and are easy to locate, measure, and count, and their edges or boundaries are readily defined. Examples of discrete features in a city include buildings, roads, traffic signals, and parks. Continuous features, on the other hand, are less well defined and exist across space. The most cited examples of continuous features are temperature and elevation. Changes in temperature and elevation tend to be gradual over relatively large areas.

Geographical features also have several characteristics, traits, or attributes that may or may not be attractive. For instance, to continue the deforestation example, determining whether a forest is a rainforest or whether a forest is in a protected park may be necessary. More general attributes may include measurements such as tree density per acre, average canopy height in meters, or proportions like percent palm trees or invasive species per hectare in the forest.

Notwithstanding the purpose of the map or GIS project, definitions of features must be clear and remain consistent. Similarly, it is essential that the attributes of features are also consistently defined, measured, and reported to generate accurate and valuable maps efficiently. Defining features and attributes of interest is often an iterative process of trial and error. Being able to associate a feature with a particular geometric form and determine the feature type are central to map abstraction, facilitating mapping, and GIS application.

Map Content and Generalization

The shape and content of maps vary according to purpose, need, and resources, among other factors. What is familiar to most maps and those within a GIS is that they are graphical representations of reality. Put another way, various graphical symbols represent geographical features or entities. Annotation or text is also commonly used on maps and facilitates map interpretation. Learning about map content and map generalization is essential because they serve as the building blocks for spatial data used within a GIS.

Building upon the previous discussion about the geometric form of geographic features, maps typically rely on three geometric objects: the point, the line, and the polygon or area. A **point** is defined by x and y coordinates, two points define a line, and a minimum of three points defines a polygon. The vital thing to note is that the definition of a point is analogous to a location that is defined by longitude and latitude. Furthermore, since lines and polygons are made up of points, location information (i.e., x and y, or longitude and latitude, coordinates) is intrinsic to points, lines, and polygons.

Simple and complex maps can be made using these three simple geometric objects. Additionally, by changing the graphical characteristics of each object, an infinite number of mapping possibilities emerge. Such changes can be made to the respective size, shape, color, and patterns of points, lines, and polygons. For instance, different sized points can reflect variations in population size. Line color or line size (i.e., thickness) can denote volume or the amount of interaction between locations. Furthermore, assorted colors and shapes can be used to reflect different values of interest.

Complementing the graphical elements described previously is annotation or text. Annotation is used to identify geographic features, such as cities, states, bodies of water, or other points of interest. Like the graphical elements, text can be varied according to size, orientation, or color. There are also numerous text fonts and styles that are incorporated into maps. For example, bodies of water are often labeled in italics.

Another map element that deserves to be mentioned and combines graphics and text is the map legend or map key. A map legend provides users with information about how geographic information is represented graphically. Legends usually consist of a title that describes the map and the various symbols, colors, and patterns used on the map. Such information is often vital to the proper interpretation of a map.

As more features and graphical elements are put on a given map, the need to generalize such features arises. Map generalization involves resolving conflicts associated with too much detail, too many features, or too much information and data to map. Generalization can take several forms:

- The simplification or symbolization of features for emphasis
- The masking or displacement of detail to increase clarity or legibility
- The selection of detail for inclusion or omission from the map
- The exaggeration of features for emphasis

Determining which aspects of generalization to use is mostly a matter of personal preference, experience, map purpose, and trial and error. Though there are general guidelines about map generalization, there are no universal standards or requirements concerning the generalization of maps and mapping. At this point that cartographic and artistic license, prejudices and biases, and creativity and design sense—or lack thereof—emerge to shape the map.

Making a map and, more generally, mapping involves a range of decisions and choices. From selecting the appropriate map scale and projection to deciding which features to map and omit, mapping is a complex blend of art and science. Many historical maps are indeed viewed as works of art, and rightly so. Learning about maps' scale, shape, and content increases our understanding of maps and deepens our appreciation of maps and map-making. This increased geographical awareness and appreciation of maps promote the sound and effective use and application of a GIS.

Image Maps

Image maps, in large part derived from satellites, are ubiquitous. Such maps can be found on the news, on the Internet, in your car, and on your mobile phone. Moreover, such images are in living color and exceedingly high resolution. Not long ago, such image maps from satellites were the sole domain of meteorologists, local weather forecasters, and various government agencies. Public access to such images was limited to the evening news.

In conjunction with the commercialization of space flight, technological advances in imaging technology opened the door for companies like Mexar (which used to be called DigitalGlobe) to provide satellite imagery and maps to the masses at the turn of the twenty-first century. With online mapping services such as Google Earth providing free and user-friendly access to such images, a revolution in maps and mapping was born.

Image maps now provide a geographic context for nightly news stories worldwide, serve as a backdrop to local real estate searches and driving directions, and are also used for research purposes. The popularity and widespread use of such images speak not only to recent technological advances and innovations but also, more important, to the geographer in us all.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

3: Spatial and Nonspatial Data

3.1: Introduction

Maps are shared, available, and distributed, unlike at any other time in history. The mapping process has also been decentralized and democratized so that many more people have access to maps and are enabled and empowered to create their maps. This democratization of maps and mapping is mainly attributable to a digital map production and consumption shift. Unlike analog or hardcopy maps that are static or fixed once printed onto paper, digital maps are highly changeable, exchangeable, and dynamic in scale, form, and content.

To understand digital maps and mapping, it is necessary to put them into the context of computing and information technology. First, this chapter introduces the building blocks of digital maps and geographic information systems (GIS), emphasizing how data and information are stored as files on a computer. Second, key issues and considerations related to data acquisition and standards are presented. Finally, the chapter concludes with a discussion of where data for use with a GIS can be found.

Learning Outcomes

- Define and describe data and information and how it is organized into files for use in a computing and geographic information systems environment.
- Highlight the difference between primary and secondary data sources and understand the importance of metadata and data standards.
- Identify and evaluate key considerations when searching for data.

GTCM Alignment

Chapter Sections

- 3.1 Data and Information
- 3.2 Metadata
- 3.3 Researching Data
- 3.4 References

3.2: Data and Information

To understand how we get from analog to digital maps, let us begin with the building blocks and foundations of the geographic information system (GIS) – namely, data and information. Geographic information systems store, edit, process, and present data and information. However, what exactly is data? Moreover, what exactly is information? The terms “data” and “information” refer to the same thing for many. For our purposes, it is helpful to distinguish between the two. Data refer to facts, measurements, characteristics, or traits of an object of interest. For your grammar sticklers out there, note that “data” is the plural form of “datum.” For example, we can collect data about all kinds of things, like the length of rainbow trout in a Colorado stream, the number of vegetarians in Alaska, the diameter of mahogany tree trunks in the Brazilian rainforest, student scores on the last GIS midterm, the altitude of mountain peaks in Nepal, the depth of snow in the Austrian Alps, or the number of people who use public transportation to get to work in London.

Once data are put into context, used to answer questions, situated within analytical frameworks, or obtain insights, they become information. Information refers to the knowledge of value obtained through collecting, interpreting, and analyzing data. Though a computer is not necessary to collect, record, manipulate, process, or visualize data or process it into information, information technology can significantly help. For instance, computers can automate repetitive tasks, store data efficiently in terms of space and cost, and provide tools for analyzing data from spreadsheets to GIS. In addition, an incredible amount of data is collected daily by satellites, grocery store product scanners, traffic sensors, temperature gauges, smartphone apps, and endlessly more. Again, this data would not be possible without the aid and innovation of information technology.

Geographic or spatial data refer to geographic facts, measurements, or characteristics of an object that permit us to define its location on the earth’s surface. Such data include, but are not restricted to, the latitude and longitude coordinates of points of interest, street addresses, postal codes, political boundaries, and even the names of places of interest. It is also important to note and reemphasize the difference between geographic and attribute data. Geographic data defines the location of an object of interest; attribute data is concerned with its nongeographic traits and characteristics.

“Spatial data is information about the locations and shapes of geographic features and the relationship between them, usually stored as coordinates and topology.” – Esri

To illustrate the distinction between geographic and attribute data, think about your home, where you grew up, or where you currently live. Within the context of this discussion, we can associate both geographic and attribute data to it. We can define the location of your home in many ways, such as with a street address, the street names of the nearest intersection, the zip code or Census block your home is located in, or latitude and longitude coordinates. What is essential is that geographic data permit us to define the location of an object (i.e., your home) on the surface of the earth.

In addition to the geographic data that defines the location of your home are the attribute data that describes the various qualities of your home. Such data could include the number of bedrooms and bathrooms in your home, whether your home has central air, the year your home was built, the number of occupants, or whether there is a swimming pool. These attribute data tell us a lot about your home but little about where it is.

It is beneficial to recognize and understand how geographic and attribute data differ and complement each other, but it is also vital when learning about and using GIS. Because a GIS requires and integrates these two distinct types of data, being able to differentiate between geographic and attribute data is the first step in organizing your GIS. Furthermore, determining which kinds of data you need will aid in your implementation and use of a GIS. Often, and in the age and context of information technology, the data and information discussed thus far is the stuff of computer files, which are the focus of the next section.

Files and Formats

When we collect data about your home, rainforests, or anything, we usually need to put them somewhere. Though we may scribble numbers and measures on the back of an envelope or write them down on a pad of paper, if we want to update, share, analyze, or map them in the future, it is often helpful to record them in digital form so a computer can read them. So, though we will not bother ourselves with the bits and bytes of computing, it is necessary to discuss some fundamental elements of computing that are both relevant and required when learning and working with a GIS.

One of the most common elements of working with computers and computing is the file. Files in a computer can contain any number of things, from a complex set of instructions (e.g., a computer program) to a list of numbers and letters (e.g., an address book). Furthermore, computer files come in all varied sizes and types. One of the clues we can use to distinguish one file from another is the file extension. A file extension refers to the letters that follow the period (".") after the file's name. The table below contains some of the most common file extensions and the types of files with which they are associated.

filename.txt Simple text file

filename.doc Microsoft Word document

filename.pdf Adobe portable document format

filename.jpg Compressed image file

filename.tif Tagged image format

filename.html Hypertext markup language (used to create websites)

filename.xml Extensible markup language

filename.zip Zipped/compressed archive

Some computer programs may be able to read or work with only specific file types, while others are more adept at reading multiple file formats. As you begin to work more with information technology and GIS, you will realize that familiarity with different file types is essential. In addition, learning how to convert or export one file type to another is also a beneficial and valuable skill to obtain. In this regard, recognizing and knowing how to identify different and unfamiliar file types will undoubtedly increase your proficiency with computers and GIS.

Of the numerous file types, one of the most common and widely accessed files is simple text, plain text, or just text file. Simple text files can be read widely by word processing programs, spreadsheet and database programs, and web browsers. Often ending with the extension ".txt" (i.e., filename.txt), text files contain no special formatting (e.g., bold, italic, underlining) and contain only alphanumeric characters. In other words, images or sophisticated graphics are not well suited for text files. Text files, however, are ideal for recording, sharing, and exchanging data because most computers and operating systems can recognize and read simple text files with programs called text editors.

When a text file contains organized or structured data in some fashion, it is sometimes called a flat file (but the file extension remains the same, i.e., .txt). Flat files are organized in a tabular format or line by line. In other words, each line or row of the file contains one and only one record. So, if we collected height measurements on three people, Tim, Jake, and Harry, the file might look something like this:

Name Height

Tim 6'1″

Sarah 5'7″

Maria 5'5″

Each row corresponds to one and only one record, observation, or case. There are two other essential elements to know about this file. First, note that the first row does not contain any data; instead, it describes the data contained in each column. When the first row of a file contains such descriptors, it is referred to as a header row or just a header. Columns in a flat-file are also called fields, variables, or attributes. For example, "Height" is the attribute, field, or variable that we are interested in, and the observations or cases in our data set are "Tim," "Jake," and "Harry." In short, rows are for records; columns are for fields.

The second unseen but critical element of the file is the spaces between each column or field. For example, a space separates the "name" column from the "height" column in the example. Upon closer inspection, however, note how the initial values of the "height" column are aligned. If a single space were used to separate each column, the height column would not be aligned. In this case, a tab is being used to separate the columns of each row. The delimiter or separator is the character used to separate columns within a flat file. Though any character can be used as a delimiter, the most common delimiters are the tab, the comma, and a single space. The following are examples of each.

Tab-Delimited Single-Space-Delimited Comma-Delimited

Name Height Name Hight Name, Height

Tim 6'1″ Tim 6'1″ Tim, 6'1″

Sarah 5'7″ Sarah 5'7″ Sarah, 5'7″

Maria 5’5″ Maria 5’5″ Maria, 5’5″

Knowing the delimiter to a flat-file is essential because it enables us to distinguish and separate the columns efficiently and without error. Sometimes such files are referred to by their delimiters, such as a “comma-separated values” file or a “tab-delimited” file.

The same general format is applied when recording and working with geographic data. Rows are reserved for records, or in the case of geographic data, locations and columns or fields are used for the attributes or variables associated with each location. For example, the following tab-delimited flat file contains data for three places (i.e., countries) and three attributes or characteristics of each country (i.e., population, language, continent), as noted by the header.

Country Population Languages Continent

France 65,000,000 French Europe

Brazil 192,000,000 Portuguese South American

Jordan 9,531,712 Arabic Southwest Asia

Files like those presented here are the building blocks of the various tables, charts, reports, graphs, and other visualizations that we see online, in print, and on television every day. They are also vital components of GIS maps and geographic representations. Rarely if ever, however, will you work with one and only one file or file type. Often, especially when working with GIS, you will work with multiple files. Such a grouping of multiple files is called a database. Since the files within a database may be of varied sizes, shapes, and even formats, we need to devise a system that will allow us to work, update, edit, integrate, share, and display the various data within the database. Such a system is referred to as a database management system (DBMS). So databases and DBMSs are crucial to GIS, and a later chapter is dedicated to them. Geodatabases are a collection of geographic data contained within a standard file system.

3.3: Metadata

Looking at the contents of the file, we can see that it contains data about the cities of Los Angeles, London, and Singapore. A comma separates each field or attribute, and the file also contains a header row that tells us about the data contained in each column. Or does it? What does the column "sun" refer to? Is it the number of sunny days this year, last year, annually, or when? What about "temp"? Does this refer to the average daytime, evening, or annual temperature? For that matter, how is temperature measured? In Celsius? Fahrenheit? Kelvin? The column "precip" refers to precipitation, but again, what are the units or time frames for such measures and data? Finally, where did these data come from? Who collected them, when were they collected, and for what purpose?

Incredibly, such a small text file can lead to so many questions. Let us extend the example to a file with one hundred records on ten variables, one thousand records on one hundred variables, or ten thousand records on one thousand variables. Through this simple example, several general but central issues related to data emerge. Such issues range from the relatively mundane naming conventions that are used to identify individual records (i.e., rows) and distinguish one field (i.e., column) from another to the issue of providing documentation about what data are included in a given file; when the data were collected; for what purpose are the data to be used; who collected them; and, of course, where did the data come from?

The previous simple text file illustrates how we cannot and should not take data and information for granted. It also highlights two important concepts regarding the source of data and the contents of data files. First, data can be put into two distinct categories regarding data sources. The first category is called primary data. Primary data refer to data collected directly or on a firsthand basis. For example, if you wanted to examine the variability of local temperatures in May, and you recorded the temperature at noon every day in May, you would be constructing a primary data set. Conversely, secondary data refer to data collected by someone else or another party. For instance, we use secondary data when working with census or economic data collected and distributed by the government.

Several factors influence the decision behind the construction and use of primary data sets versus secondary data sets. Data acquisition costs in terms of money, availability, and time are essential factors. The data acquisition and integration phase of most geographic information system (GIS) projects are often the most time-consuming. In other words, locating, obtaining, and putting together the data for a GIS project, whether you collect the data yourself or use secondary data, may take up most of your time. Of course, depending on the purpose, availability, and need, it may not be necessary to construct an entirely new data set (i.e., primary data set). However, considering the vast amounts of data and publicly available information, for example, via the Internet, the cost and time savings of using secondary data often offset any benefits associated with primary data collection.

Now that we understand the difference between primary and secondary data and the rationale, how do we find the data and information we need? As noted earlier, there is an incredibly vast and growing amount of data and information available to us and performing an online search for "deforestation data" will return hundreds—if not thousands—of results. We need to turn to even more data to overcome this data and information overload. We are looking for a special kind of data called metadata. So defined, metadata is data about data. At one level, a header row in a simple text file like those discussed in the previous section is analogous to metadata. The header row provides data (e.g., names and labels) about the subsequent rows of data.

However, header rows may need an additional explanation, as previously illustrated. Furthermore, when working with or searching through several data sets, it can be pretty tedious or impossible to open every file to determine its contents and usability. Enter metadata. Many files, particularly secondary data sets, come with a metadata file. These metadata files contain items such as general descriptions of the contents of the file, definitions for the various terms used to identify records (rows) and fields (fields), the range of values for fields, the quality or reliability of the data, and measurements, how the data were collected, when the data were collected, and who collected the data. Though not all data are accompanied by metadata, it is easy to see and understand why metadata is essential and valuable when searching for secondary data and when constructing primary data that may be shared in the future.

Just as simple files come in all shapes, sizes, and formats, so do metadata. As the amount and availability of data and information increase each day, metadata plays a critical role in making sense. The metadata class we are most concerned with when working with a GIS is called geospatial metadata. As the name suggests, geospatial metadata is data about geographical and spatial data. According to the Federal Geographic Data Committee (FGDC) in the United States (see http://www.fgdc.gov), "Geospatial metadata are used to document digital geographic resources such as GIS files, geospatial databases, and earth imagery. A geospatial metadata record includes core library catalog elements such as Title, Abstract, and Publication Data; geographic elements such as Geographic Extent and Projection Information; and database elements such as Attribute Label Definitions and Attribute Domain

Values.” The definition of geospatial metadata is about improving transparency regarding data and promoting standards. Take a few moments to explore and examine the contents of a geospatial metadata file that conforms to the FGDC here.

Standards refer to widely promoted, accepted, and followed the rules and practices. Given the range and variability of data and data sources, identifying a common thread to locate and understand the contents of any given file can be a challenge. However, just as the rules of grammar and mathematics provide the foundations for communication and numeric calculations, metadata provides similar frameworks for working with and sharing data and information from various sources.

The central point behind metadata is that it facilitates data and information sharing. Within the context of large organizations such as governments, data and information sharing can eliminate redundancies and increase efficiencies. Moreover, access to data and information promotes the integration of different data to improve analyses, inform decisions, and shape policy. The role that metadata, and geospatial metadata, play in GIS is critical and offers enormous benefits in terms of cost and time savings. The sharing, widespread distribution, and integration of various geographic and nongeographic data and information enabled by metadata drive some of the most exciting and compelling innovations in GIS and the broader geospatial information technology community. More important, widespread access, distribution, and sharing of geographic data and information have essential social costs and benefits and yield better analyses and more informed decisions.

Files and Formats

When we collect data about your home, rainforests, or anything, we usually need to put them somewhere. Though we may scribble numbers and measures on the back of an envelope or write them down on a pad of paper, if we want to update, share, analyze, or map them in the future, it is often helpful to record them in digital form so a computer can read them. Though we will not bother ourselves with the bits and bytes of computing, it is necessary to discuss some fundamental elements of computing that are both relevant and required when learning and working with a GIS.

One of the most common elements of working with computers and computing is the file. Files in a computer can contain any number of things, from a complex set of instructions (e.g., a computer program) to a list of numbers and letters (e.g., an address book). Furthermore, computer files come in all varied sizes and types. One of the clues we can use to distinguish one file from another is the file extension. A file extension refers to the letters that follow the period (“.”) after the file’s name. The table below contains some of the most common file extensions and the types of files with which they are associated.

filename.txt Simple text file

filename.doc Microsoft Word document

filename.pdf Adobe portable document format

filename.jpg Compressed image file

filename.tif Tagged image format

filename.html Hypertext markup language (used to create websites)

filename.xml Extensible markup language

filename.zip Zipped/compressed archive

Some computer programs may be able to read or work with only specific file types, while others are more adept at reading multiple file formats. As you begin to work more with information technology and GIS, you will realize that familiarity with different file types is essential. In addition, learning how to convert or export one file type to another is also a beneficial and valuable skill to obtain. In this regard, recognizing and knowing how to identify different and unfamiliar file types will undoubtedly increase your proficiency with computers and GIS.

Of the numerous file types, one of the most common and widely accessed files is simple text, plain text, or just text file. Simple text files can be read widely by word processing programs, spreadsheet and database programs, and web browsers. Often ending with the extension “.txt” (i.e., filename.txt), text files contain no special formatting (e.g., bold, italic, underlining) and contain only alphanumeric characters. In other words, images or sophisticated graphics are not well suited for text files. Text files, however, are ideal for recording, sharing, and exchanging data because most computers and operating systems can recognize and read simple text files with programs called text editors.

When a text file contains organized or structured data in some fashion, it is sometimes called a flat file (but the file extension remains the same, i.e., .txt). Flat files are organized in a tabular format or line by line. In other words, each line or row of the file contains one and only one record. So, if we collected height measurements on three people, Tim, Jake, and Harry, the file might look something like this:

Name Height

Tim 6’1″

Sarah 5’7″

Maria 5’5″

Each row corresponds to one and only one record, observation, or case. There are two other essential elements to know about this file. First, note that the first row does not contain any data; instead, it describes the data contained in each column. When the first row of a file contains such descriptors, it is referred to as a header row or just a header. Columns in a flat-file are also called fields, variables, or attributes. For example, “Height” is the attribute, field, or variable that we are interested in, and the observations or cases in our data set are “Tim,” “Jake,” and “Harry.” In short, rows are for records; columns are for fields.

The second unseen but critical element of the file is the spaces between each column or field. For example, a space separates the “name” column from the “height” column in the example. Upon closer inspection, however, note how the initial values of the “height” column are aligned. If a single space were used to separate each column, the height column would not be aligned. In this case, a tab is being used to separate the columns of each row. The delimiter or separator is the character used to separate columns within a flat file. Though any character can be used as a delimiter, the most common delimiters are the tab, the comma, and a single space. The following are examples of each.

Tab-Delimited Single-Space-Delimited Comma-Delimited

Name Height Name Hight Name, Height

Tim 6’1″ Tim 6’1″ Tim, 6’1″

Sarah 5’7″ Sarah 5’7″ Sarah, 5’7″

Maria 5’5″ Maria 5’5″ Maria, 5’5″

Knowing the delimiter to a flat-file is essential because it enables us to distinguish and separate the columns efficiently and without error. Sometimes such files are referred to by their delimiters, such as a “comma-separated values” file or a “tab-delimited” file.

The same general format is applied when recording and working with geographic data. Rows are reserved for records, or in the case of geographic data, locations and columns or fields are used for the attributes or variables associated with each location. For example, the following tab-delimited flat file contains data for three places (i.e., countries) and three attributes or characteristics of each country (i.e., population, language, continent), as noted by the header.

Country Population Languages Continent

France 65,000,000 French Europe

Brazil 192,000,000 Portuguese South American

Jordan 9,531,712 Arabic Southwest Asia

Files like those presented here are the building blocks of the various tables, charts, reports, graphs, and other visualizations that we see online, in print, and on television every day. They are also vital components of GIS maps and geographic representations. Rarely if ever, however, will you work with one and only one file or file type. Often, especially when working with GIS, you will work with multiple files. Such a grouping of multiple files is called a database. Since the files within a database may be of different sizes, shapes, and even formats, we need to devise a system that will allow us to work, update, edit, integrate, share, and display the various data within the database. Such a system is referred to as a database management system (DBMS). So databases and DBMSs are crucial to GIS, and a later chapter is dedicated to them. Geodatabases are a collection of geographic data contained within a standard file system.

3.4: Researching Data

Now that we have a basic understanding of data and information, where can we find such data and information? Though an Internet search will undoubtedly produce myriad sources and types of data, the hunt for relevant and valuable data is often a challenging and iterative process. Therefore, before hopping online and downloading the first thing that appears from a web search, it is helpful to frame our search for data with the following questions and considerations:

What exactly is the purpose of the data?

Given that the world is drowning in vast amounts of data, articulating why we need or do not need a given data set will streamline the search for valuable and relevant data. To this end, the more specific we can be about the purpose of the needed data, the more efficient our search for data will be. For example, if we are interested in understanding and studying economic growth, it is helpful to determine both temporal and geographic scales. In other words, for what periods (e.g., 1850–1900) and intervals (e.g., quarterly, annually) are we interested, and at what level of analysis (e.g., national, regional, state)? Frequently, data availability, or the lack of relevant data, will force us to change the purpose or scope of our original question. A clear purpose will yield a more efficient search for data and enables us to accept or quickly discard the various data sets that we may come across.

What data already exists and is available?

Before searching for new data, it is always a good idea to the inventory we already have data. Such data may be from previous projects or analyses or colleagues and classmates, but the key point here is that we can save a lot of time and effort using data we already possess. Furthermore, we better understand what we need by identifying what we have. For instance, though we may already have census data (i.e., attribute data), we may need updated geographic data that contains the boundaries of US states or counties.

What are the costs associated with data acquisition?

Data acquisition costs go beyond financial costs. Just as important as the financial costs to data are those that involve your time. Time is money. The time and energy you spend collecting, finding, cleaning, and formatting data are time and energy taken away from data analysis. Therefore, depending on deadlines, time constraints, and deliverables, it is critical to learn to manage your time when looking for data.

What format does the data need to be in?

Though many programs can read many formats of data, some data types can only be read by some programs, and some programs require data formats—understanding what data formats you can use and those that you cannot aid in your search for data. For instance, one of the most common forms of geographic information system (GIS) data is called the shapefile. Not all GIS programs can read or use shapefiles, but it may be necessary to convert to or from a shapefile or another format. The more data formats we are familiar with, the better off we will be in our search for data because we will understand what we can use and what format conversions will need to be made if necessary.

All these questions are of equal importance, and being able to answer them will assist in a more efficient and effective search for data. Several other considerations behind the search for data, particularly GIS data, but those listed here provide an initial pathway to a successful search for data.

As information technology evolves and more data are collected and distributed, the various forms of data that can be used with GIS increase. GIS uses and integrates two types of data: geographic and attribute data. Sometimes the source of both geographic and attribute data is the same. For instance, the United States Census Bureau distributes geographic boundary files (e.g., census tract level, county level, state-level) and the associated attribute data (e.g., population, race/ethnicity, income). What is more, such data are freely available at no charge. US census data are exceptional in many respects: free and comprehensive.

Every search for data will vary according to the purpose. However, government data tend to have good coverage and provide a point of reference from which other data can be added, compared, and evaluated. Whether you need satellite imagery data from the National Aeronautics and Space Administration (NASA) or land use data from the United States Geological Survey (USGS), such government sources tend to be dependable, reputable, and consistent. Another critical element of most government data is that they are freely accessible to the public. In other words, there is no charge to use or to acquire the data. Data that are free to use are called public data.

Unlike publicly available data, there are numerous private or proprietary data sources. The main difference between public and private data is that the former tends to be free, and the latter must be acquired at a cost. Furthermore, there are often restrictions on distributing and disseminating proprietary data sets (i.e., sharing the purchased data is not allowed). Again, proprietary data may be the only option depending on the subject matter. Another reason for using proprietary data is that the data may be formatted and cleaned according to your needs. When working with deadlines, the trade-off between financial cost and time saved must be seriously considered and evaluated.

The search for data, particularly the data you need, is often the most time-consuming aspect of any GIS-related project. Therefore, it is essential to define and clarify your data requirements and needs, from the temporal and geographic scales of data to the formats required, as clearly as possible and as early as possible. Such definition and clarity will pay dividends in your search for the correct data, better analyses, and well-informed decisions.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

4: Geospatial Data Models

4.1: Introduction

To visualize natural phenomena, one must first determine how to represent geographic space best. Data models are rules and constructs used to describe and represent aspects of the world on a computer. Two primary data models are available to complete this task: raster data models and vector data models.

Learning Outcomes

- Explain how raster data models are implemented in geographic information systems applications.
- Describe how vector data models are implemented in geographic information systems applications.
- Determine how satellite imagery, aerial photography, and unmanned aerial systems are implemented in geographic information systems applications.

GTCM Alignment

Chapter Sections

- 4.1 Introduction
- 4.2 Raster Data Models
- 4.3 Vector Data Models
- 4.4 Satellite Imagery, Aerial Photography, and Drones
- 4.5 References

4.2: Raster Data Models

The raster data model is widely used in applications beyond geographic information systems (GIS). You are already familiar with this data model if you have any experience with digital photographs. The ubiquitous JPEG, BMP, and TIFF file formats (among others) are based on the raster data model (see Chapter 5 “Geospatial Data Management,” Section 5.3 “File Formats”). Take a moment to view your favorite digital image. If you zoom deeply into the image, you will notice that it is composed of an array of tiny square pixels (or picture elements). These uniquely colored pixels combine to form a coherent image when viewed as a whole.

Furthermore, all liquid crystal display (LCD) computer monitors are based on raster technology as they are composed of a set number of rows and columns of pixels. Notably, the foundation of this technology predates computers and digital cameras by a century. The neo-impressionist artist, Georges Seurat, developed a painting technique called “pointillism” in the 1880s, which similarly relies on the amassing of small, monochromatic “dots” of ink that combine to form a larger image. If you are as generous as the author, you may think of your raster dataset creations as sublime works of art.

The raster data model consists of rows and columns of equally sized pixels interconnected to form a planar surface. These pixels are used as building blocks for creating points, lines, areas, networks, and surfaces. Although pixels may be triangles, hexagons, or even octagons, square pixels represent the simplest geometric form with which to work. Accordingly, most available raster GIS data are built on the square pixel. These squares are typically reformed into rectangles of various dimensions if the data model is transformed from one projection to another (e.g., from State Plane coordinates to UTM [Universal Transverse Mercator] coordinates).

The raster data model is a grid-based system because of the reliance on a uniform series of square pixels. Typically, a single data value will be assigned to each grid locale. Each cell in a raster carries a single value, representing the characteristic of the spatial phenomenon at a location denoted by its row and column. The data type for that cell value can be either integer or floating-point (Chapter 5 “Geospatial Data Management,” Section 5.1 “Geographic Data Acquisition”). Alternatively, the raster graphic can reference a database management system wherein open-ended attribute tables can be used to associate multiple data values to each pixel. The advance in computer technology has made this second methodology increasingly feasible as significant datasets are no longer constrained by computer storage issues as they were previously.

The raster model will average all values within a given pixel to yield a single value. Therefore, the more area covered per pixel, the less accurate the associated data values are. The area covered by each pixel determines the spatial resolution of the raster model from which it is derived. Specifically, the resolution is determined by measuring one side of the square pixel. For example, a raster model with pixels representing 10 m by 10 m (or 100 square meters) in the real world would be said to have a spatial resolution of 10 m; a raster model with pixels measuring 1 km by 1 km (1 square kilometer) in the real world would be said to have a spatial resolution of 1 km; and so forth.

Care must be taken when determining the resolution of a raster because using an overly coarse pixel resolution will cause a loss of information, whereas using an overly fine pixel resolution will result in significant increases in file size and computer processing requirements during display and analysis. An effective pixel resolution will consider both the map scale and the minimum mapping unit of the other GIS data. In the case of raster graphics with coarse spatial resolution, the data values associated with specific locations are not necessarily explicit in the raster data model. For example, if the location of telephone poles were mapped on a coarse raster graphic, it would be clear that the pole would not fill the entire cell. Instead, the pole would be assumed to be located somewhere within that cell (typically at the center).

Imagery employing the raster data model must exhibit several properties. First, each pixel must hold at least one value, even if that data value is zero. Furthermore, if no data are present for a given pixel, a data value placeholder must be assigned to this grid cell. An arbitrary, readily identifiable value (e.g., −9999) will often be assigned to pixels with no data value. Second, a cell can hold any alphanumeric index that represents an attribute. In the case of quantitative datasets, attribute assignation is straightforward. For example, if a raster image denotes elevation, the data values for each pixel would be some indication of elevation, usually in feet or meters. In the case of qualitative datasets, data values are indices that necessarily refer to some predetermined translational rule. For example, in the case of a land-use/land-cover raster graphic, the following rule may be applied: 1 = grassland, 2 = agricultural, 3 = disturbed, and so forth. The third property of the raster data model is that points and lines “move” to the center of the cell. As one might expect, if a 1 km resolution raster image contains a river or stream, the location of the actual waterway within the “river” pixel will be unclear. Therefore, there is a general assumption that all zero-dimensional (point) and one-dimensional (line) features will be located toward the center of the cell. As a corollary, the minimum width for any line feature must necessarily be one cell regardless of the actual width of the feature. If it is not, the feature will not be represented in the image and will therefore be assumed to be absent.

Several methods exist for encoding raster data from scratch. Three of these models are as follows:

- Cell-by-cell raster encoding. This minimally intensive method encodes a raster by creating records for each cell value by row and column. This method could be considered a large spreadsheet wherein each spreadsheet cell represents a pixel in the raster image. This method is also referred to as “exhaustive enumeration.”
- Run-length raster encoding. This method encodes cell values in runs of similarly valued pixels and can result in a highly compressed image file. The run-length encoding method is helpful in situations where large groups of neighboring pixels have similar values (e.g., discrete datasets such as land use/land cover or habitat suitability) and are less valuable where neighboring pixel values vary widely (e.g., continuous datasets such as elevation or sea-surface temperatures).
- Quad-tree raster encoding. This method divides a raster into a hierarchy of subdivided quadrants based on similarly valued pixels. The division of the raster stops when a quadrant is made entirely from cells of the same value. A quadrant that cannot be subdivided is called a “leaf node.”

Advantages and Disadvantages of the Raster Model

The use of a raster data model confers many advantages. First, the technology required to create raster graphics is inexpensive and ubiquitous. Nearly everyone currently owns a raster image generator, namely a digital camera, and few cellular phones are sold today that do not include such functionality. Similarly, a plethora of satellites are constantly beaming up-to-the-minute raster graphics to scientific facilities across the globe (Chapter 5 “Geospatial Data Management,” Section 5.3 “File Formats”). These graphics are often posted online for private and public use, occasionally at no cost to the user.

Additional advantages of raster graphics are the relative simplicity of the underlying data structure. Each grid location in the raster image correlates to a single value (or series of values if attributes tables are included). This simple data structure may also help explain why it is easy to perform overlay analyses on raster data (for more on overlay analyses, see Chapter 7 “Geospatial Analysis I: Vector Operations,” Section 7.1 “Single Layer Analysis”). This simplicity also lends itself to straightforward interpretation and maintenance of the graphics relative to their vector counterpart.

There are also several disadvantages to using the raster data model despite the advantages. The first disadvantage is that raster files are typically large. Particularly in the case of raster images built from the cell-by-cell encoding methodology, the sheer number of values stored for given dataset results in potentially enormous files. Any raster file that covers a large area and has finely resolved pixels will quickly reach hundreds of megabytes in size or more. Moreover, these large files are only getting more significant as the quantity and quality of raster datasets continue to keep pace with the quantity and quality of computer resources and raster data collectors (e.g., digital cameras, satellites).

A second disadvantage of the raster model is that the output images are less “pretty” than their vector counterparts. This is particularly noticeable when the raster images are enlarged or zoomed. Depending on how far one zooms into a raster image, the details and coherence of that image will quickly be lost amid a pixilated sea of randomly colored grid cells.

The geometric transformations that arise during map reprojection efforts can cause problems for raster graphics and represent the third disadvantage to using the raster data model. As described in Chapter 2, “Map Anatomy,” Section 2.2 “Map Scale, Coordinate Systems, and Map Projections,” changing map projections will alter the size and shape of the original input layer and frequently result in the loss or addition of pixels (White, 2006). [1] These alterations will result in the perfect square pixels of the input layer taking on some alternate rhomboidal dimensions. However, the problem is more significant than a simple reformation of the square pixel. Indeed, the reprojection of a raster image dataset from one projection to another brings change to pixel values that may, in turn, significantly alter the output information (Seong, 2003).

The final disadvantage of using the raster data model is that it is not suitable for some types of spatial analyses. For example, difficulties arise when attempting to overlay and analyze multiple raster graphics produced at differing scales and pixel resolutions. For example, combining information from a raster image with 10 m spatial resolution with a raster image with 1 km spatial resolution will produce nonsensical output information as the scales of analysis are far too disparate to result in meaningful and interpretable conclusions. In addition, some network and spatial analyses (i.e., determining directionality or geocoding) can be problematic to perform on raster data.

The final disadvantage of using the raster data model is that it is not suitable for some types of spatial analyses. For example, difficulties arise when attempting to overlay and analyze multiple raster graphics produced at differing scales and pixel resolutions.

For example, combining information from a raster image with 10 m spatial resolution with a raster image with 1 km spatial resolution will produce nonsensical output information as the scales of analysis are far too disparate to result in meaningful and interpretable conclusions. In addition, some network and spatial analyses (i.e., determining directionality or geocoding) can be problematic to perform on raster data.

4.3: Vector Data Models

In contrast to the raster data model is the vector data model. In this model, space is not quantized into discrete grid cells like in the raster model. Vector data models use points and their associated X, Y coordinate pairs to represent the vertices of spatial features, much as if they were being drawn on a map by hand (Aronoff, 1989). [1] The data attributes of these features are then stored in a separate database management system. The spatial information and the attribute information for these models are linked via a simple identification number given to each feature in a map.

Three fundamental vector types exist in geographic information systems (GIS): points, lines, and polygons. Points are zero-dimensional objects that contain only a single coordinate pair. Points are typically used to model singular, discrete features such as buildings, wells, power poles, sample locations, etc. Points have only the property of location. Other types of point features include the node and the vertex. Specifically, a point is a stand-alone feature, while a node is a topological junction representing a common X, Y coordinate pair between intersecting lines and polygons. Vertices are defined as each bend along with a line or polygon feature that is not the intersection of lines or polygons.

Points can be spatially linked to form more complex features. Lines are one-dimensional features composed of multiple, explicitly connected points. Lines represent linear features such as roads, streams, faults, and boundaries. In addition, lines have the property of length. Lines that directly connect two nodes are sometimes referred to as chains, edges, segments, or arcs.

Polygons are two-dimensional features created by multiple lines that loop back to create a “closed” feature. In the case of polygons, the first coordinate pair (point) on the first line segment is the same as the last coordinate pair on the last line segment. Polygons represent features such as city boundaries, geologic formations, lakes, soil associations, vegetation communities, and more. In addition, polygons have the properties of area and perimeter. Therefore, polygons are also called areas.

Vector Data Models Structures

Vector data models can be structured in many different ways. We will examine two of the more common data structures here. The simplest vector data structure is called the spaghetti data model (Dangermond, 1982). [2] In the spaghetti model, each point, line, and polygon feature is represented as a string of X, Y coordinate pairs (or as a single X, Y coordinate pair in the case of a vector image with a single point) no inherent structure. One could envision each line in this model as a single strand of spaghetti formed into complex shapes by adding more strands of spaghetti. In this model, any polygons that lie adjacent to each other must be made up of their lines or strands of spaghetti. In other words, each polygon must be uniquely defined by its own set of X, Y coordinate pairs, even if the adjacent polygons share the same boundary information. This creates some redundancies within the data model and therefore reduces efficiency.

Despite the location designations associated with each line, or strand of spaghetti, spatial relationships are not explicitly encoded within the spaghetti model; instead, they are implied by their location. This results in a lack of topological information, which is problematic if the user attempts to make measurements or analyses. Therefore, the computational requirements are steep if any advanced analytical techniques are employed on vector files. Nevertheless, the simple structure of the spaghetti data model allows for efficient reproduction of maps and graphics as this topological information is unnecessary for plotting and printing.

In contrast to the spaghetti data model, the topological data model is characterized by including topological information within the dataset, as the name implies. Topology is a set of rules that model the relationships between neighboring points, lines, and polygons and determines how they share geometry. For example, consider two adjacent polygons. In the spaghetti model, the shared boundary of two neighboring polygons is defined as two separate, identical lines. The inclusion of topology into the data model allows for a single line to represent this shared boundary with an explicit reference to denote which side of the line belongs to which polygon. Topology is also concerned with preserving spatial properties when the forms are bent, stretched, or placed under similar geometric transformations, which allows for more efficient projection and reprojection of map files.

Three basic topological precepts are necessary to understand the topological data model are outlined here. First, connectivity describes the arc-node topology for the feature dataset. As discussed previously, nodes are more than simple points. In the topological data model, nodes are the intersection points where two or more arcs meet. In the case of arc-node topology, arcs have both a from-node (i.e., starting node) indicating where the arc begins and a to-node (i.e., ending node) indicating where the arc ends. In addition, between each node pair is a line segment, sometimes called a link, which has its identification number and references both its from-node and to-node. For example, in this figure, “Arc-Node Topology,” arcs 1, 2, and 3 intersect because they share node 11. Therefore, the computer can determine that it is possible to move along arc 1 and turn onto arc 3, while it is impossible to move from arc 1 to arc 5, as they do not share a common node.

The second fundamental topological precept is area definition. Area definition states that an arc that connects to surround an area defines a polygon, also called polygon-arc topology. In the case of polygon-arc topology, arcs are used to construct polygons, and each arc is stored only once. This reduces the amount of data stored and ensures that adjacent polygon boundaries do not overlap. For example, in the figure on "Polygon-Arc Topology," the polygon-arc topology clarifies that polygon F comprises arcs 8, 9, and 10.

Contiguity, the third topological precept, is based on the concept that polygons that share a boundary are deemed adjacent. Specifically, polygon topology requires that all arcs in a polygon have a direction (a from-node and a to-node), which allows adjacency information to be determined. Polygons that share an arc are deemed adjacent or contiguous, and therefore the "left" and "right" sides of each arc can be defined. This left and right polygon information is stored explicitly within the attribute information of the topological data model. The "universe polygon" is an essential component of polygon topology that represents the external area located outside of the study area. The figure "Polygon Topology" shows that arc 6 is bound by polygon B and to the right by polygon C. Polygon A, the universe polygon, is to the left of arcs 1, 2, and 3.

Topology allows the computer to rapidly determine and analyze the spatial relationships of all its included features. In addition, topological information is essential because it allows for efficient error detection within a vector dataset. In the case of polygon features, open or unclosed polygons, which occur when an arc does not completely loop back upon itself, and unlabeled polygons, which occur when an area does not contain any attribute information, violate polygon-arc topology rules. Another topological error found with polygon features is the sliver. Slivers occur when the shared boundary of two polygons does not meet precisely.

In the case of line features, topological errors occur when two lines do not meet perfectly at a node. This error is called an "undershoot" when the lines do not extend far enough to meet each other and an "overshoot" when the line extends beyond the feature it should connect to. The result of overshoots and undershoots is a "dangling node" at the end of the line. Dangling nodes are not always an error, however, as they occur in the case of dead-end streets on a road map.

Many types of spatial analysis require the degree of organization offered by topologically explicit data models. For example, network analysis (e.g., finding the best route from one location to another) and measurement (e.g., finding the length of a river segment) rely heavily on the concept of to-and-from nodes and use this information, along with attribute information, to calculate distances, shortest routes, or the quickest route. Topology also allows for sophisticated neighborhood analysis such as determining adjacency, clustering, or nearest neighbors.

Now that the basics of the concepts of topology have been outlined, we can begin to understand the topological data model better. In this model, the node acts as more than just a simple point along a line or polygon. Instead, the node represents the point of intersection for two or more arcs. Arcs may or may not be looped into polygons. Regardless, all nodes, arcs, and polygons are individually numbered. This numbering allows for quick and easy reference within the data model.

Advantages and Disadvantages of the Vector Model

In comparison with the raster data model, vector data models tend to be better representations of reality due to the accuracy and precision of points, lines, and polygons over the regularly spaced grid cells of the raster model. This results in vector data tending to be more aesthetically pleasing than raster data.

Vector data also provides an increased ability to alter the scale of observation and analysis. However, as each coordinate pair associated with a point, line, and polygon represents an infinitesimally exact location (albeit limited by the number of significant digits and data acquisition methodologies), zooming deep into a vector image does not change the view of a vector graphic in the way that it does a raster graphic.

Vector data tend to be more compact in the data structure, so file sizes are typically much smaller than their raster counterparts. Although the ability of modern computers has minimized the importance of maintaining small file sizes, vector data often require a fraction of the computer storage space compared to raster data.

The final advantage of vector data is that topology is inherent in the vector model. Using a vector model, this topological information results in simplified spatial analysis (e.g., error detection, network analysis, proximity analysis, and spatial transformation).

Alternatively, there are two primary disadvantages of the vector data model. First, the data structure tends to be more complex than the simple raster data model. As the location of each vertex must be stored explicitly in the model, there are no shortcuts for storing data like there are for raster models (e.g., the run-length and quad-tree encoding methodologies).

Second, the implementation of spatial analysis can also be complicated due to minor differences in accuracy and precision between the input datasets. Similarly, the algorithms for manipulating and analyzing vector data are complex and can lead to intensive processing requirements, mainly when dealing with large datasets.

4.4: Remote Imagery Acquisition

A wide variety of satellite imagery and aerial photography is available in geographic information systems (GIS). Although these products are raster graphics, they are substantively different in their usage within a GIS. Satellite imagery and aerial photography provide essential contextual information for a GIS and are often used to conduct heads-up digitizing (Chapter 5 "Geospatial Data Management," Section 5.1.4 "Secondary Data Capture"), whereby features from the image are converted into vector datasets.

Satellite Imagery

Remotely sensed satellite imagery is becoming increasingly common as satellites equipped with technologically advanced sensors are continually being sent into space by public agencies and private companies around the globe. Satellites are used for military and civilian Earth observation, communication, navigation, weather, research, etc. More than 3,000 satellites have been sent to space, with over 2,500 from Russia and the United States. These satellites maintain different altitudes, inclinations, eccentricities, synchronizes, and orbital centers, allowing them to image various surface features and processes.

Satellites can be active or passive. Active satellites use remote sensors that detect reflected responses from objects irradiated from artificially generated energy sources. For example, active sensors such as radars emit radio waves, laser sensors emit light waves, and sonar sensors emit sound waves. In all cases, the sensor emits the signal and then calculates the time it takes for the returned signal to "bounce" back from some remote feature. Knowing the speed of the emitted signal, the time delay from the original emission to the return can be used to calculate the distance to the feature.

Passive satellites, alternatively, make use of sensors that detect the reflected or emitted electromagnetic radiation from natural sources. This natural source is typically the energy from the sun, but other sources can be imaged as well, such as magnetism and geothermal activity. Using an example we have all experienced, taking a picture with a flash-enabled camera would be active remote sensing, while using a camera without a flash (i.e., relying on ambient light to illuminate the scene) would be passive remote sensing.

Their resolution determines the quality and quantity of satellite imagery. There are four types of resolution that characterize any remote sensor (Campbell, 2002). As described previously in the raster data model section (Section 4.1 "Raster Data Models"), the spatial resolution of a satellite image is a direct representation of the ground coverage for each pixel shown in the image. If a satellite produces imagery with a 10 m resolution, the corresponding ground coverage for each pixel is 10 m by 10 m, or 100 square meters on the ground. Spatial resolution is determined by the sensors' instantaneous field of view (IFOV). The IFOV is the ground area through which the sensor receives the electromagnetic radiation signal and is determined by the height and angle of the imaging platform.

Spectral resolution denotes the ability of the sensor to resolve wavelength intervals, also called bands, within the electromagnetic spectrum. The spectral resolution is determined by the interval size of the wavelengths and the number of intervals being scanned. Multispectral and hyperspectral sensors are those sensors that can resolve a multitude of wavelength intervals within the spectrum. For example, the IKONOS satellite resolves images for bands at the blue (445–516 nm), green (506–95 nm), red (632–98 nm), and near-infrared (757–853 nm) wavelength intervals on its 4-meter multispectral sensor.

Temporal resolution is the time between each image collection period and is determined by the repeat cycle of the satellite's orbit. Temporal resolution can be thought of as true-nadir or off-nadir. Areas considered true-nadir are located directly beneath the sensor, while off-nadir areas are imaged obliquely. In the case of the IKONOS satellite, the temporal resolution is 3 to 5 days for off-nadir imaging and 144 days for true-nadir imaging.

The fourth and final type of resolution, radiometric resolution, refers to the sensor's sensitivity to variations in brightness and specifically denotes the number of grayscale levels that the sensor can image. Typically, the available radiometric values for a sensor are 8-bit (yielding values that range from 0–255 as 256 unique values or as 28 values); 11-bit (0–2,047); 12-bit (0–4,095); or 16-bit (0–63,535) (see Chapter 5 "Geospatial Data Management", Section 5.1.1 "Data Types" for more on bits). Landsat-7, for example, maintains 8-bit resolution for its bands and can therefore record values for each pixel that range from 0 to 255.

Because of the technical constraints associated with satellite remote sensing systems, there is a trade-off between these diverse types of resolution. Improving one type of resolution often necessitates reducing one of the other types of resolution. For example, an increase in spatial resolution is typically associated with a decrease in spectral resolution and vice versa. Similarly, geostationary satellites (those that circle the earth proximal to the equator once each day) yield high temporal resolution but low spatial resolution, while sun-synchronous satellites (those that synchronize a near-polar orbit of the sensor with the sun's illumination) yield low temporal resolution while providing high spatial resolution. Although technological advances can improve the various resolutions of an image, care must always be taken to ensure that the imagery you have chosen is adequate to represent or model the geospatial features that are most important to your study.

Aerial Photography

Aerial photography, like satellite imagery, represents a vast source of information for use in any GIS. Platforms for the hardware used to take aerial photographs include airplanes, helicopters, balloons, rockets, etc. While aerial photography connotes visible spectrum images, sensors to measure bands within the nonvisible spectrum (e.g., ultraviolet, infrared, near-infrared) can also be fixed to aerial sources. Similarly, aerial photography can be active or passive and from vertical or oblique angles. However, care must be taken with aerial photographs as the sensors used to take the images are like cameras in their use of lenses. These lenses add curvature to the images, which becomes more pronounced as one moves away from the center of the photo.

Another source of potential error in an aerial photograph is relief displacement. This error arises from the three-dimensional aspect of terrain features and is seen as apparent leaning away of vertical objects from the center point of an aerial photograph. To imagine this type of error, consider that a smokestack would look like a doughnut if the viewing camera were directly above the feature. However, if this same smokestack were observed near the edge of the camera's view, one could observe the sides of the smokestack. This error is frequently seen with trees and multistory buildings and worsens with increasingly taller features.

Orthophotos are vertical photographs that have been geometrically "corrected" to remove the curvature and terrain-induced error from images. The most common orthophoto product is the digital ortho quarter quadrangle (DOQQ). DOQQs are available through the US Geological Survey (USGS), which began producing these images from their library of 1:40,000-scale National Aerial Photography Program photos. These images can be obtained either grayscale or color with 1-meter spatial resolution and 8-bit radiometric resolution. As the name suggests, these images cover a quarter of a USGS 7.5-minute quadrangle, which equals an approximately 25 square mile area. In addition, these photos include an additional 50 to 300-meter edge around the photo that allows users to many mosaic DOQQs into a single, continuous image. These DOQQs are ideal for use in a GIS as background display information, for data editing, and heads-up digitizing.

Small Unmanned Aerial Systems (sUAS)

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

5: Geospatial Data Management

5.1: Introduction

Every user of geospatial data has experienced the challenge of obtaining, organizing, storing, sharing, and visualizing their data. The variety of formats and data structures and the disparate quality of geospatial data can result in a dizzying accumulation of useful and useless pieces of spatially explicit information that must be poked, prodded, and wrangled into a single, unified dataset. This chapter addresses the fundamental concerns related to data acquisition and management of geospatial data's various formats and qualities currently available for use in modern geographic information system (GIS) projects.

Learning Objectives

- Introduce different data types, measurement scales, and data capture methods.
- Understand the basic properties of a relational database management system.
- Overview of a sample of the most common types of vectors, raster, and hybrid file formats.
- Ascertain the diverse types of error inherent in geospatial datasets.

GTCM Alignment

Chapter Sections

5.2: Geographic Data Acquisition

Acquiring geographic data is crucial in any geographic information system (GIS) effort. It has been estimated that data acquisition typically consumes 60 to 80 percent of the time and money spent on any given project. Therefore, care must be taken to ensure that GIS projects remain mindful of their stated goals so spatial data collection proceeds efficiently and effectively. This chapter outlines the many forms and sources of geospatial data available in a GIS.

Data Types

The type of data that we employ to help us understand a given entity is determined by (1) what we are examining, (2) what we want to know about that entity, and (3) our ability to measure that entity at the desired scale. The most common data types available in a GIS are alphanumeric strings, numbers, Boolean values, dates, and binaries.

An alphanumeric string, or text, the data type is any simple combination of letters and numbers that may or may not form coherent words. The number data type can be subcategorized as either floating-point or integer. A floating-point is any data value that contains decimal digits, while an integer is any data value that does not contain decimal digits. Integers can be short or long, depending on the number of significant digits. Also, they are based on the concept of the "bit" in a computer. As you may recall, a bit is the most basic unit of information in a computer and stores values in one of two states: 1 or 0.

Therefore, an 8-bit attribute would consist of eight 1s or 0s in any combination (e.g., 10010011, 00011011, 11100111).

Short integers are 16-bit values and, therefore, can characterize numbers ranging from −32,768 to 32,767 or from 0 to 65,535 depending on whether the number is signed or unsigned (i.e., contains a + or − sign).Long integers, alternatively, are 32-bit values and therefore can characterize numbers ranging either from −2,147,483,648 to 2,147,483,647 or from 0 to 4,294,967,295.

A single-precision floating-point value occupies 32 bits, like the long integer. However, this data type provides a value of up to 7 bits to the left of the decimal (a maximum value of 128, or 127 if signed) and up to 23-bit values to the right of the decimal point (approximately seven decimal digits). A double-precision floating-point value stores two 32-bit values as a single value. Then, double precision floats can represent a value with up to 11 bits to the left of the decimal point and values with up to 52 bits to the right of the decimal (approximately 16 decimal digits).

Boolean, date, and binary values are less complex. Boolean values are simply those values deemed true or false based on the application of a Boolean operator such as AND, OR, and NOT. The date data type is self-explanatory, while the binary data type represents attributes whose values are either 1 or 0.

Measurement Scale

In addition to defining data by type, a measurement scale acts to group data according to the level of complexity (Stevens, 1946). For GIS analyses, measurement scales can be grouped into two broad categories. Nominal and ordinal data represent categorical data; interval and ratio data represent numeric data.

The most straightforward data measurement scale is the nominal, or named scale. The nominal scale makes statements about what to call data points but does not allow for scalar comparisons between one object and another. For example, the attribution of nominal information to points representing cities will describe whether the given locale is "Los Angeles" or "New York." However, no further denotations can be made about those locales, such as population or voting history. Other examples of nominal data include last name, eye color, land-use type, ethnicity, and gender.

Ordinal data places attribute information into ranks and yields more precisely scaled information than nominal data. Ordinal data describes the position in which data occur, such as first, second, third, etc. These scales may also take on names such as "very unsatisfied," "unsatisfied," "satisfied," and "very satisfied." Although this measurement scale indicates the ranking of each data point relative to other data points, the ordinal scale does not explicitly denote the exact quantitative difference between these rankings. For example, if an ordinal attribute represents which runner came in first, second, or third place, it does not state how much time the winning runner beat the second-place runner. Therefore, one cannot undertake arithmetic operations with ordinal data. The only sequence is explicit.

A measurement scale that does allow precise quantitative statements to be made about attributes is interval data. Interval data are measured along a scale in which each position is equidistant. Elevation and temperature readings are typical representations of interval data. For example, this scale can determine that 30 degrees Fahrenheit is 5 degrees Fahrenheit warmer than 25 degrees Fahrenheit. A notable property of the interval scale is that zero is not a meaningful value because zero does not represent nothingness or the absence of a value. Indeed, 0 degrees Fahrenheit does not indicate that no temperature exists. Similarly, an elevation of 0 feet does not indicate a lack of elevation; rather, it indicates the mean sea level.

Ratio data are like the interval measurement scale; however, it is based on a meaningful zero value. Population density is an example of ratio data whereby a 0-population density indicates that no people live in interest. Similarly, the Kelvin temperature scale is a ratio scale as 0 K does imply that no heat (temperature) is measurable within the given attribute.

Specific to numeric datasets, data values also can be discrete or continuous. Discrete data maintain a finite number of values, while an infinite number of values can represent continuous data. For example, the number of mature trees on a small property will necessarily be between one and one hundred (for argument's sake). However, the height of those trees represents a constant data value as there are an infinite number of potential values (e.g., one tree may be 20 feet tall, 20.1 feet, or 20.15 feet, 20.157 feet, etc.).

Primary Data Capture

Now that we have a sense of the different data types and measurement scales available for use in a GIS, we must direct our thoughts to how this data can be acquired. Primary data capture is a direct data acquisition methodology usually associated with in-the-field effort. In the case of vector data, directly captured data commonly comes from a global positioning system (GPS) or other types of surveying equipment such as a total station. Total stations are specialized, primary data capture instruments that combine a theodolite (or transit), which measures horizontal and vertical angles, with a tool to measure the slope distance from the unit to an observed point. A total station allows field crews to derive the topography quickly and accurately for a particular landscape.

In the case of GPS, handheld units access positional data from satellites and log the information for subsequent retrieval. A network of twenty-four navigation satellites is situated around the globe and provides precise coordinate information for any point on the earth's surface. Maintaining a line of sight to four or more of these satellites provides the user with reasonably accurate location information. Depending on user preference, these locations can be collected as individual points or linked together to form lines or polygons. In addition, the user can enter attribute data such as land-use type, telephone pole number, and river name. This location and attribute data can then be uploaded to the GIS for visualization. Depending on the GPS make and model, this upload often requires some type of intermediate file conversion via software provided by the manufacturer of the GPS unit. However, some free online resources can convert GPS data from one format to another. GPS Babel is an example of such an online resource (http://www.gpsvisualizer.com/gpsbabel).

In addition to the typical GPS unit, GPS is becoming increasingly incorporated into other innovative technologies. For example, smartphones now embed GPS capabilities as a standard technological component. These phone/GPS units maintain comparable accuracy to similarly priced stand-alone GPS units and are responsible for a renaissance in facilitating portable, real-time data capture and sharing to the masses. Furthermore, the ubiquity of this technology led to a proliferation of crowdsourced data acquisition alternatives. Crowdsourcing is a data collection method whereby users contribute freely to building spatial databases. This rapidly expanding methodology is utilized in such applications as TomTom's MapShare application, Google Earth, Bing Maps, and ArcGIS.

Raster data obtained via direct capture comes more commonly from remotely sensed sources. Remotely sensed data can obviate the need for physical access to the imaged area. In addition, the researcher can characterize vast tracts of land with little to no additional time and labor. On the other hand, validation is required for remotely sensed data to ensure that the sensor is operating correctly and adequately calibrated to collect the desired information. Satellites and aerial cameras provide the most ubiquitous sources of direct-capture raster data (Chapter 4 "Data Models for GIS," Section 4.3.1 "Satellite Imagery").

Secondary Data Capture

Secondary data capture is an indirect methodology that utilizes the vast amount of existing geospatial data in digital and hard-copy formats. Before initiating any GIS effort, it is always wise to mine online resources for existing GIS data that may fulfill your mapping needs without the potentially intensive step of creating the data from scratch. Such digital GIS data are available from a variety of sources, including international agencies (CGIAR, CIESIN, United Nations, World Bank, etc.); federal governments (USGS, USDA, NOAA, USFWS, NASA, EPA, US Census, etc.); state governments (CDFG, Teale Data Center, INGIS, MARIS, NH GIS Resources, etc.); local governments (SANDAG, RCLIS, etc.); university websites (UCLA, Duke, Stanford, University of Chicago, Indiana Spatial Data Portal, etc.); and commercial websites (ESRI, GeoEye, Geocomm, etc.). These secondary data are available in various file types, extents, and sizes but are ready-made in most GIS software packages. Often these data are free, but many sites will charge a fee for access to the proprietary information they have developed.

Although these data sources are all cases where the information has been converted to digital format and projected adequately for use in a GIS, much spatial information can be gleaned from existing, non-digital sources. Paper maps, for example, may contain current or historical information on a locale that cannot be found in digital format. In this case, the process of digitization can be used to create digital files from the original paper copy. Three primary methods exist for digitizing spatial information: two are manual, and one is automated.

Tablet digitizing is a manual data capture method whereby a user enters coordinate information into a computer using a digitizing tablet and a digitizing puck. To begin, a paper map is secured to a backlit digitizing tablet. The backlight allows all features on the map to be easily observed, which reduces eyestrain. The coordinates of the point, line, and polygon features on the paper map are then entered into a digital file as the user employs a puck, which is like a multi-buttoned mouse with a crosshair, to "click" their way around the vertices of each desired feature. The resulting digital file will need to be correctly georeferenced following completion of the digitization task to ensure that this information will adequately align with existing datasets.

The second manual data capture method, heads-up digitizing, is called "on-screen" digitizing. Heads-up digitizing can be used on either paper maps or existing digital files. In the case of a paper map, the map must first be scanned into the computer at a high enough resolution to resolve all pertinent features. Second, the now-digital image must be registered so the map will conform to an existing coordinate system. Third, the user can enter control points on the screen and transform, or "rubber-sheet," the scanned image into real-world coordinates. Finally, the user simply zooms to specific areas on the map and traces the points, lines, and polygons, like the tablet digitization example.

Heads-up digitizing is particularly simple when existing GIS files, satellite images, or aerial photographs are used as a baseline. For example, if a user plans to digitize the boundary of a lake as seen from a georeferenced satellite image, the steps of scanning and registering can be skipped, and projection information from the originating image can simply be copied over to the digitized file.

The third automated method of secondary data capture requires the user to scan a paper map and vectorize the information therein. This vectorization method typically requires a specific software package to convert a raster scan to vector lines. This requires a very high-resolution, clean scan. If the image is not clean, all the imperfections on the map will be converted to false points/lines/polygons in the digital version. If a clean scan is not available, it is often faster to use a manual digitization methodology. This method is much quicker than the aforementioned manual methods and may be the best option if multiple maps must be digitized and time is a limiting factor. Often, a semiautomatic approach is employed whereby a map is scanned and vectorized, followed by a heads-up digitizing session to edit and repair any errors that occurred during automation.

The final secondary data capture method worth noting is reports and documents. Via this method, one enters reports and documents into the attribute table of an existing digital GIS file containing all the pertinent points, lines, and polygons. For example, new information specific to census tracts may become available following a scientific study. The GIS user simply needs to download the existing GIS file of census tracts and begin entering the study's report/document information directly into the attribute table. If the data tables are available digitally, the use of the "join" and "relate" functions in a GIS (Section 5.2.2 "Joins and Relates") are often extremely helpful as they will automate much of the data entry effort.

5.3: Geospatial Database Management

A database is a structured collection of data files. A database management system (DBMS) is a software package that allows creating, storing, maintaining, manipulating, and retrieving large datasets distributed over one or more files. A DBMS and its associated functions are usually accessed through commercial software packages such as Microsoft Access, Oracle, FileMaker Pro, or Avanquest MyDataBase. Database management usually refers to storing tabular data in row and column format and is frequently used for personal, business, government, and scientific endeavors. Alternatively, geospatial database management systems include the functionality of a DBMS and contain specific geographic information about each data point, such as identity, location, shape, and orientation. Integrating this geographic information with the tabular attribute data of a classical DBMS provide users with powerful tools to visualize and answer the spatially explicit questions that arise in an increasingly technological society.

Several database models exist, such as the flat, hierarchical, network, and relational models (Worboys, 1995; Jackson, 1999). A flat database is a spreadsheet storing all data in a single, large table. A hierarchical database is also a simple model that organizes data into a "one-to-many" association across levels

. Common examples of this model include phylogenetic trees for classifying plants and animals and familial genealogical trees showing parent-child relationships. However, network databases are like hierarchical databases because they also support "many-to-many" relationships. This expanded capability allows greater search flexibility within the dataset and reduces the potential redundancy of information. Alternatively, both the hierarchical and network models can become incredibly complex depending on the size of the databases and the number of interactions between the data points. Modern geographic information system (GIS) software typically employs a fourth model referred to as a relational database (Codd, 1970).

Relational Database Management Systems

A relational database management system (RDBMS) is a collection of connected tables so that data can be accessed without reorganization of the tables. The tables are created such that each column represents a particular attribute (e.g., soil type, PIN number, last name, acreage), and each row contains a unique instance of data for that columnar attribute (e.g., Delhi Sands Soils, 5555, Smith, 412.3 acres)

In the relational model, each table (not surprisingly called a relation) is linked to each other table via predetermined keys (Date 1995). [4] The primary key represents the attribute (column) whose value uniquely identifies a particular record (row) in the relation (table). The primary key may not contain missing values as multiple missing values would represent nonunique entities that violate the basic rule of the primary key. The primary key corresponds to an identical attribute in a secondary table (and possibly third, fourth, fifth, etc.) called a foreign key. This results in all the information in the first table being related to the information in the second table via the primary and foreign keys, hence the term "relational" DBMS. With these links in place, tables within the database can be kept relatively simple, resulting in minimal computation time and file complexity. This process can be repeated over many tables if each contains a foreign key corresponding to another table's primary key.

The relational model has two primary advantages over the other database models described earlier. First, each table can now be separately prepared, maintained, and edited. This is particularly useful when considering the potentially massive size of many of today's modern databases. Second, the tables may be maintained separately until the need for a particular query or analysis calls for the tables to be related. This creates a significant degree of efficiency for processing information within a given database.

It may become apparent to the reader that there is immense potential for redundancy in this model as each table must contain an attribute that corresponds to an attribute in every other related table. Therefore, redundancy must actively be monitored and managed in an RDBMS. A set of rules called normal forms was developed (Codd, 1970). There are three primary normal forms. The first normal form refers to five conditions that must be met (Date 1995). [6] They are as follows:

- There is no sequence to the ordering of the rows.
- There is no sequence to the ordering of the columns.
- Each row is unique.
- Every cell contains one and only one value.
- All values in a column pertain to the same subject.

The second normal form states that any column, not a primary key, must depend on the primary key. This reduces redundancy by eliminating the potential for multiple primary keys throughout multiple tables. However, this step often involves the creation of new tables to maintain normalization.

The third normal form states that all nonprimary keys must depend on the primary key, while the primary key remains independent of all nonprimary keys. This form was wittily summed up by Kent (1983) [7], who quipped that all nonprimary keys "must provide a fact about the key, the whole key, and nothing but the key." Echoing this quote is the rejoinder: "so help me, Codd" (personal communication with Foresman 1989).

Joins and Relations

An additional advantage of an RDBMS is that it allows attribute data in separate tables to be linked together. The two operations commonly used to accomplish this are the *join* and *relate*.

The **join operation** appends the fields of one table into a second table using an attribute or field that is common to both tables. This is commonly utilized to combine attribute information from one or more nonspatial data tables (i.e., information taken from reports or documents) with a spatially explicit GIS feature layer. The second type of join combines feature information based on spatial location and association rather than common attributes. In ArcGIS, three types of spatial joins are available. Users may (1) match each feature to the closest feature, (2) match each feature to the feature that it is part of, or (3) match each feature to the feature that it intersects.

Alternatively, the **related operation** temporarily associates two map layers or tables while keeping them physically separate. Relates are bidirectional, so data can be accessed from one of the tables by selecting records in the other table. The related operation also allows for the association of three or more tables.

Sometimes it can be unclear as to which operation one should use. As a general rule, joins are most suitable for instances involving one-to-one or many-to-one relationships. Joins are also advantageous since the data from the two tables are readily observable in the single output table. On the other hand, the use of relates is suitable for all table relationships (one-to-one, one-to-many, many-to-one, and many-to-many); however, they can slow down computer access time if the tables are extensive or spread out over remote locations.

5.4: File Formats

Geospatial data are stored in many different file formats. Each geographic information system (GIS) software package, and each version of these software packages, supports different formats. This is true for both vector and raster data. Although several more common file formats are summarized here, many other formats exist for use in various GIS programs.

Vector File Formats

The most common vector file format is the shapefile. Shapefiles, developed by ESRI in the early 1990s for use with the dBASE III database management software package in ArcView 2, are simple, nontopological files developed to store the geometric location and attribute information of geographic features.

Shapefiles are incapable of storing null values and annotations or network features. Field names within the attribute table are limited to ten characters, and each shapefile can represent an only point, line, or polygon feature sets. Supported data types are limited to floating-point, integer, date, and text. All commercial and open-source GIS software supports Shapefiles.

Despite being called a "shapefile," this format is a compilation of many different files. Table 5.1 "Shapefile File Types" lists and describes the different file formats associated with the shapefile. Only the SHP, SHX, and DBF file formats are mandatory to create a functioning shapefile, while all others are conditionally required. As a rule, the names for each file should conform to the MS-DOS 8.3 convention when using older versions of GIS software packages. According to this convention, the filename prefix can contain eight characters, and the filename suffix contains three characters. However, the more recent GIS software packages have relaxed this requirement and will accept longer filename prefixes.

The ArcInfo coverage is the earliest vector format file for use in GIS software packages, which is still in use today. This georelational file format supports multiple feature types (e.g., points, lines, polygons, annotations) while storing the topological information associated with those features. Attribute data are stored as multiple files in a separate directory labeled "Info." Due to its creation in an MS-DOS environment, these files maintain strict naming conventions. File names cannot be longer than thirteen characters, cannot contain spaces, cannot start with a number, and must be entirely in lowercase. Coverages cannot be edited in ArcGIS 9.x or later versions of ESRI's software package.

The US Census Bureau maintains a specific type of shapefile referred to as TIGER or TIGER/Line (Topologically Integrated Geographic Encoding and Referencing system). Although these open-source files do not contain actual census information, they map features such as census tracts, roads, railroads, buildings, rivers, and other features that support and improve the Bureau and improve the Bureau's ability to collect census information. TIGER/Line shapefiles, first released in 1990, are topologically explicit and linked to the Census Bureau's Master Address File (MAF), enabling the geocoding of street addresses. These files are accessible to the public and can be freely downloaded from private vendors that support the format.

The AutoCAD DXF (Drawing Interchange Format or Drawing Exchange Format) is a proprietary vector file format developed by Autodesk to allow interchange between engineering-based CAD (computer-aided design) software and other mapping software packages. DXF files were initially released in 1982 to provide an exact representation of AutoCAD's native DWG format. Although the DXF is still commonly used, newer versions of AutoCAD have incorporated more complex data types (e.g., regions, dynamic blocks) that are not supported in the DXF format. Therefore, it may be presumed that the DXF format may become less popular in geospatial analysis over time.

Finally, the US Geological Survey (USGS) maintains an open-source vector file format that details physical and cultural features across the United States. These topologically explicit DLGs (Digital Line Graphics) come in large-, intermediate-, and small-scale depending on whether they are derived from 1:24,000-; 1:100,000-; or 1:2,000,000-scale USGS topographic quadrangle maps. The features available in the different DLG types depend on the scale of the DLG but include data such as administrative and political boundaries, hydrography, transportation systems, hypsography, and land cover.

Vector data files can also be structured to represent surface elevation information. A TIN (Triangulated Irregular Network) is an open-source vector data structure that uses contiguous, nonoverlapping triangles to represent geographic surfaces. In comparison, the raster depiction of a surface represents elevation as an average value over the spatial extent of the individual pixel (see Section 5.3.2 "Raster File Formats"), and the TIN data structure models each vertex of the triangle as an exact elevation value at a specific point on the earth. The arcs between each vertex approximate the elevation between two vertices. These arcs are then aggregated into triangles from which information on elevation, slope, aspect, and surface area can be derived across the entire extent of the model's space. Note that the term "irregular" in the name of the data model refers to the fact that the vertices are typically laid out in a scattered fashion.

The use of TINs confers certain advantages over raster-based elevation models (see Section 5.3.2 "Raster File Formats"). First, linear topographic features are accurately represented relative to their raster counterpart. Second, a comparatively small number of data points are needed to represent a surface, so file sizes are typically much smaller. This is particularly true as vertices can be clustered in areas where relief is complex and sparse in areas where relief is simple. Third, specific elevation data can be incorporated into the data model post hoc via the placement of additional vertices if the original is deemed insufficient or inadequate. Finally, specific spatial statistics can be calculated that cannot be obtained when using a raster-based elevation model, such as flood plain delineation, storage capacity curves for reservoirs, and time-area curves for hydrographs.

Raster File Formats

A multitude of raster file format types is available for use in GIS. The selection of raster formats has dramatically increased with the widespread availability of imagery from digital cameras, video recorders, satellites, etc. Raster imagery is typically 8-bit (256 colors) or 24-bit (16 million colors). Due to ongoing technological advancements, raster image file sizes have been getting larger and larger. Two types of file compression are commonly used to deal with this potential constraint: lossless and lossy. Lossless compression reduces file size without decreasing image quality. Lossy compression exploits the human eye's limitations by removing information from the image that cannot be sensed, resulting in smaller file sizes than lossless compression.

Among the most common raster files used on the web are the JPEG, TIFF, and PNG formats, all of which are open source and can be used with most GIS software packages. The JPEG (Joint Photographic Experts Group) and TIFF (Tagged Image File Format) raster formats are most frequently used by digital cameras to store 8-bit values for each of the red, blue, and green colors spaces (and sometimes 16-bit colors, in the case of TIFF images). JPEGs support lossy compression, while TIFFs can be either lossy or lossless. Unlike JPEG, TIFF images can be saved in either RGB or CMYK color spaces.PNG (Portable Network Graphics) files are 24-bit images that support either lossy or lossless compression. PNG files are designed for efficient viewing in web-based browsers such as Internet Explorer, Mozilla Firefox, Netscape, and Safari.

Native JPEG, TIFF, and PNG files do not have georeferenced information associated with them and, therefore, cannot be used in any geospatial mapping efforts. A world file must first be created to employ these files in a GIS. A world file is a separate, plaintext data file that specifies the locations and transformations that allow the image to be projected into a standard coordinate system (e.g., Universal Transverse Mercator [UTM] or State Plane). The filename of the world file is based on the name of the raster file, while a w is typically added to the file extension. For example, the world file extension name for a JPEG is JPW; for a TIFF, it is TFW; and for a PNG, PGW.

An example of a raster file format with explicit georeferencing information is the proprietary MrSID (Multiresolution Seamless Image Database) format. LizardTech, Inc. developed this lossless compression format for large aerial photographs or satellite images, whereby portions of a compressed image can be viewed quickly without decompressing the entire file. In addition, the MrSID format is frequently used for visualizing orthophotos.

Like MrSID, the proprietary ECW (Enhanced Compression Wavelet) format also includes georeferencing information within the file structure. This lossy compression format was developed by Earth Resource Mapping and supports up to 255 layers of image information. Due to the potentially substantial file sizes associated with an image that supports so many layers, ECW files represent an excellent option for performing rapid analysis on large images while using a small amount of the computer's RAM (Random Access Memory), thus accelerating computation speed.

Like the open-source, vector-based DLG, DRGs (Digital Raster Graphics) are scanned versions of USGS topographic maps and include all the collar material from the originals. In addition, the geospatial information found within the image's neat line is georeferenced, specifically to the UTM coordinate system. These graphics are scanned at a minimum of 250 dpi (dots per inch) and therefore have a spatial resolution of approximately 2.4 meters. DRGs contain thirteen colors and may look slightly different from the originals. In addition, they include all the collar material from the original print version, are georeferenced to the surface of the earth, fit the Universal Transverse Mercator (UTM) projection, and are based on the NAD27 data points (NAD stands for North American Datum).

Some raster file formats are developed explicitly for modeling elevation, like the TIN vector format. These include the USGS DEM, USGS SDTS, and DTED file formats. The USGS DEM (US Geological Survey Digital Elevation Model) is a popular file format due to widespread availability, the simplicity of the model, and the extensive software support for the format. Each pixel value in these grid-based DEMs denotes spot elevations on the ground, usually in feet or meters. Care must be taken when using grid-based DEMs due to the enormous volume of data accompanying these files as the spatial extent covered in the image increases. DEMs are referred to as digital terrain models (DTMs) when they represent a simple, bare-earth model and digital surface models (DSMs)when they include the heights of landscape features such as buildings and trees.

USGS DEMs can be classified into one of four levels of quality (labeled 1 to 4) depending on their source data and resolution. This source data can be 1:24,000-; 1:63,360-; or 1:250,000-scale topographic quadrangles. The DEM format is a single file of ASCII text comprised of three data blocks: A, B, and C. The A block contains header information such as data origin, type, and measurement systems. The B block contains contiguous elevation data described as a six-character integer. The C block contains trailer information such as the scene's root-mean-square (RMS) error. The USGS DEM format has recently succeeded the USGS SDTS (Spatial Data Transfer Standard) DEM format. The SDTS format [1] was specifically developed as a distribution for transferring data from one computer to another with zero data loss.

The DTED (Digital Terrain Elevation Data) format is another elevation-specific raster file format. It was developed in the 1970s for military purposes such as line of sight analysis, 3-D visualization, and mission planning. The DTED format maintains three levels of data over five different latitudinal zones. Level 0 data has a resolution of approximately 900 meters; Level 1 data has a resolution of approximately 90 meters, and Level 2 data has a resolution of approximately 30 meters.

Hybrid File Formats

A geodatabase is a recently developed, proprietary ESRI file format that supports vector and raster feature datasets (e.g., points, lines, polygons, annotation, JPEG, TIFF) within a single file. In addition, this format maintains topological relationships and is stored as an MDB file. The geodatabase was developed to be a comprehensive model for representing and modeling geospatial information.

There are three diverse types of geodatabases. The personal geodatabase was developed for single-user editing, whereby two editors cannot work on the same geodatabase at a given time. The personal geodatabase employs the Microsoft Access DBMS file format and maintains a size limit of 2 gigabytes per file, although it has been noted that performance begins to degrade after file size approaches 250 megabytes. The personal geodatabase is currently being phased out by ESRI and is therefore not used for new data creation.

The file geodatabase similarly allows only single-user editing, but this restriction applies only to unique feature datasets within a geodatabase. The file geodatabase incorporates new tools such as domains (rules applied to attributes), subtypes (groups of objects with a feature class or table), and split/merge policies (rules to control and define the output of split and merge operations). This format stores information as binary files with a size limit of 1 terabyte and has been noted to perform and scale much more efficiently than the personal geodatabase (approximately one-third of the feature geometry storage required by shapefiles and personal geodatabases). File databases are not tied to any specific relational database management system and can be employed on Windows and UNIX platforms. Finally, file geodatabases can be compressed to read-only formats that further reduce file size without subsequently reducing performance.

The third hybrid ESRI format is the ArcSDE geodatabase, which allows multiple editors to simultaneously work on feature datasets within a single geodatabase (a.k.a. versioning). This format can be employed on Windows and UNIX platforms like the file geodatabase. However, the file size is limited to 4 gigabytes, and its proprietary nature requires an ArcInfo or ArcEditor license for use. The ArcSDE geodatabase is implemented on the SQL Server Express software package, a free DBMS platform developed by Microsoft.

In addition to the geodatabase, Adobe Systems Incorporated's geospatial PDF (Portable Document Format) is an open-source format that allows for representing geometric entities such as points, lines, and polygons. In addition, geospatial PDFs can be used to find and mark coordinate pairs, measure distances, reproject files, and georegister raster images. This format is handy as the PDF is widely accepted as the preferred standard for printable web documents. Although functionally similar, the geospatial PDF should not be confused with the GeoPDF format developed by TerraGo Technologies. Instead, the GeoPDF is a branded version of the geospatial PDF.

Finally, Google Earth supports a new, open-source, hybrid file format called a KML (Keyhole Markup Language). KML files associate points, lines, polygons, images, 3-D models, longitude and latitude values, and other view information such as tilt, heading, altitude, etc. KMZ files are commonly encountered, and they are zipped versions of KML files.

5.5: Data Quality

Not all geospatial data are created equally. Data quality refers to the ability of a given dataset to satisfy the objective for which it was created. With the voluminous amounts of geospatial data being created and served to the cartographic community, care must be taken by individual geographic information system (GIS) users to ensure that the data employed for their project is suitable for the task at hand.

Two primary attributes characterize data quality. Accuracy describes how close a measurement is to its actual value and is often expressed as a probability (e.g., 80 percent of all points are within +/− 5 meters of their true locations). Precision refers to the variance of a value when repeated measurements are taken. For example, a watch may be correct to 1/1000th of a second (precise) but maybe 30 minutes slow (not accurate). The blue darts are precise and accurate, while the red darts are precise but inaccurate.

Several errors can arise when accuracy and precision requirements are not met during data capture and creation. Positional accuracy is the probability of a feature within +/− units of its true location on earth (absolute positional accuracy) or its location concerning other mapped features (relative positional accuracy). For example, it could be said that a particular mapping effort may result in 95 percent of trees being mapped to within +/− 5 feet for their true location (absolute), or 95 percent of trees are mapped to within +/− 5 feet of their location as observed on a digital ortho quarter quadrangle (relative).

Speaking about absolute positional error does beg the question of what exactly is the true location of an object? As discussed in Chapter 2, “Map Anatomy,” differing conceptions of the earth’s shape have led to many projections, data points, and spheroids, each attempting to clarify positional errors for locations on the earth. To begin addressing this unanswerable question, the US National Map Accuracy Standard (or NMAS) suggests that a paper map is expected to have no more than 10 percent of measurable points fall outside the accuracy values to meet the horizontal accuracy requirements range. Similarly, the vertical accuracy of no more than 10 percent of elevations on a contour map shall be in error of more than one-half the contour interval. Any map that does not meet these horizontal and vertical accuracy standards will be deemed unacceptable for publication.

Positional errors arise via multiple sources. First, the process of digitizing paper maps commonly introduces such inaccuracies. Second, errors can arise while registering the map on the digitizing board. Third, a paper map can shrink, stretch, or tear over time, changing the scene’s dimensions. Fourth, input errors created from hastily digitized points are standard. Finally, converting between coordinate systems and transforming data points may also introduce errors to the dataset.

The root-mean-square (RMS) error is frequently used to evaluate the degree of inaccuracy in a digitized map. This statistic measures the deviation between the control points’ actual (true) and estimated (digitized) locations. For example, the inaccuracies of lines representing soil types result from input control point location errors. By applying an RMS error calculation to the dataset, one could determine the accuracy of the digitized map and thus determine its suitability for inclusion in each study.

Positional errors can also arise when features to be mapped are inherently vague. Take the example of a wetland. What defines a wetland boundary? Wetlands are determined by hydrologic, vegetative, and edaphic factors. Although the US Army Corps of Engineers is currently responsible for defining the boundary of wetlands throughout the country, this task is not as simple as it may seem.

Moreover, regional differences in the characteristics of a wetland make delineating these features particularly troublesome. For example, the definition of a wetland boundary for the riverine wetlands in the eastern United States, where water is abundant, is often useless when delineating similar types of wetlands in the desert southwest United States. Indeed, the complexity and confusion associated with the conception of what a “wetland” is may result in difficulties defining the feature in the field, which subsequently leads to positional accuracy errors in the GIS database.

In addition to positional accuracy, attribute accuracy is a common source of error in a GIS. Attribute errors can occur when an incorrect value is recorded within the attribute field or when a field is missing a value. Misspelled words and other typographical errors are common as well. Similarly, a common inaccuracy occurs when developers enter “0” in an attribute field when the value is “null.” This is common in count data where “0” would represent zero findings, while a “null” would represent a locale where no data collection effort was undertaken. Finally, in the case of categorical values, inaccuracies occasionally occur when attributes are mislabeled. For example, a land-use/land-cover map may list a polygon as “agricultural” when it is, in fact, “residential.” This is particularly true if the dataset is out of date, which leads us to our next source of error.

Temporal accuracy addresses the age or timeliness of a dataset. No dataset is ever wholly current. It has already become outdated in the time it takes to create the dataset. Regardless, there are several dates to be aware of while using a dataset. These dates should be found within the metadata. The publication date will tell you when the dataset was created and released. The field date relates to the date and time the data was collected. If the dataset contains any future predictions, there should also be a forecast period and date. To address temporal accuracy, many datasets undergo a regular data update regimen. For example, the California Department of Fish and Game updates its sensitive species databases on a near-monthly basis as new findings are continually being made. It is essential to ensure that, as an end-user, you are constantly using the most up-to-date data for your GIS application.

The fourth type of accuracy in a GIS is logical consistency. Logical consistency requires that the data be topologically correct. For example, does a stream segment of a line shapefile fall within the floodplain of the corresponding polygon shapefile? Do roadways connect at nodes? Do all the connections and flows point in the correct direction in a network? Regarding the last question, the author recently used an unnamed smartphone application to navigate a busy city roadway and was twice told to turn in the wrong direction down one-way streets. So beware, errors in logical consistency may lead to traffic violations, or worse!

The final type of accuracy is data completeness. Comprehensive inclusion of all features within the GIS database is required to ensure accurate mapping results. All the data must be present for a dataset to be accurate. Are all the counties in the state represented? Are all the stream segments included in the river network? Is every convenience store listed in the database? Are only certain types of convenience stores listed within the database? Indeed, incomplete data will inevitably lead to incomplete or insufficient analysis.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

6: Geospatial Data Characteristics and Visualization

6.1: Introduction

In previous chapters, we learned how geographic information system (GIS) software packages use databases to store extensive attribute information for geospatial features within a map. However, the usefulness of this information is not realized until similarly powerful analytical tools are employed to access, process, and simplify the data. To accomplish this, GIS typically provides comprehensive tools for searching, querying, describing, summarizing, and classifying datasets. With these data exploration tools, even the most expansive datasets can be mined to allow users to make meaningful insights into and statements about that information.

Learning Objectives

- Review the most frequently used distribution, central tendency, and dispersion measures.
- Outline the basics of the SQL language and understand the various query techniques available in a geographic information system.
- Describe the methodologies available to parse data into various classes for visual representation in a map.

GTCM Alignment

Chapter Sections

- 6.1 Introduction
- 6.2 Descriptions and Summaries
- 6.3 Searches and Queries
- 6.4 Data Classification
- 6.5 References

6.2: Descriptions and Summaries

No discussion of geospatial analysis would be complete without a brief overview of basic statistical concepts. The basic statistics outlined here represent a starting point for any attempt to describe, summarize, and analyze geospatial datasets. An example of a common geospatial statistical endeavor is analyzing point data obtained by a series of rainfall gauges patterned throughout a particular region. Given these rain gauges, one could determine the typical amount and variability of rainfall at each station and typical rainfall throughout the region. In addition, you could interpolate the amount of rainfall that falls between each station or the location where the most (or least) rainfall occurs. Furthermore, you could predict the expected amount of rainfall in the future at each station, between each station, or within the region

Over the past few decades, the increase in computational power has given rise to vast datasets that cannot be summarized easily. Descriptive statistics provide simple numeric descriptions of these large datasets. Descriptive statistics tend to be univariate analyses, meaning they examine one variable at a time. There are three families of descriptive statistics that we will discuss here: measures of distribution, measures of central tendency, and measures of dispersion. However, before we delve too deeply into various statistical techniques, we must first define a few terms.

- Variable: a symbol used to represent any given value or set of values
- Value: an individual observation of a variable (in a geographic information system [GIS], this is also called a record)
- Population: the universe of all possible values for a variable
- Sample: a subset of the population
- n: the number of observations for a variable
- Array: a sequence of observed measures (in a GIS, this is also called a field and is represented in an attribute table as a column)
- Sorted Array: an ordered, quantitative array

Measures of Distribution

The measure of the distribution of a variable is merely a summary of the frequency of values over the range of the dataset (hence, this is often called a frequency distribution). Typically, the values for the given variable will be grouped into a predetermined series of classes (also called intervals, bins, or categories), and the number of data values that fall into each class will be summarized. A graph showing the number of data values within each class range is called a histogram. For example, the percentage of grades received by a class on an exam may result in the following array (n = 30):

Array of Exam Scores: {87, 76, 89, 90, 64, 67, 59, 79, 88, 74, 72, 99, 81, 77, 75, 86, 94, 66, 75, 74, 83, 100, 92, 75, 73, 70, 60, 80, 85, 57}

The following general guidelines should be observed when placing this array into a frequency distribution. First, between five and fifteen different classes should be employed, although the exact number of classes depends on the number of observations. Second, each observation goes into one and only one class. Third, when possible, use classes that cover an equal range of values (Freund and Perles

2006). With these guidelines in mind, the exam score array shown earlier can be visualized with the following histogram.

As you can see from the histogram, specific descriptive observations can be readily made. Most students received a C on the exam (70–79). Two students failed the exam (50–59). Five students received an A (90–99). Note that this histogram does violate the third basic rule that each class covers an equal range because an F grade ranges from 0–59, whereas the other grades have ranges of equal size. Regardless, in this case, we are most concerned with describing the distribution of grades received during the exam. Therefore, creating class ranges that best suit our individual needs makes perfect sense.

Measures of Central Tendency

We can further explore the exam score array by applying measures of central tendency. There are three primary measures of central tendency: the mean, mode, and median. The mean, more commonly referred to as the average, is the most often used measure of central tendency. To calculate the mean, simply add all the values in the array and divide that sum by the number of observations. To return to the exam score example from earlier, the sum of that array is 2,340, and there are 30 observations (n = 30). So, the mean is 2,340 / 30 = 78.

The mode measures central tendency representing the most frequently occurring value in the array. For example, in the case of the exam scores, the mode of the array is 75 as this was received by the most number of students (three in total). Finally, the median is the observation that, when the array is ordered from lowest to highest, it falls precisely in the center of the sorted array. More specifically, the median is the value in the middle of the sorted array when there are an odd number of observations. Alternatively, when there is an even number of observations, the median is calculated by finding the mean of the two central values. For example, if the array of exam scores were reordered into a sorted array, the scores would be listed this way:

Sorted Array of Exam Scores: {57, 59, 60, 64, 66, 67, 70, 72, 73, 74, 74, 75, 75, 75, 76, 77, 79, 80, 81, 83, 85, 86, 87, 88, 89, 90, 92, 93, 94, 99}

Since n = 30 in this example, there are an even number of observations. Therefore, the mean of the two central values (15th = 76 and 16th = 77) is used to calculate the median as described earlier, resulting in (76 + 77) / 2 = 76.5. The mean, mode, and median represent the most basic ways to examine trends in a dataset.

Measures of Dispersion

The third type of descriptive statistics is measures of dispersion (also referred to as measures of variability). These measures describe the spread of data around the mean. The most straightforward measure of dispersion is the range. The range equals the most significant value minus in the dataset the smallest. In our case, the range is 99 − 57 = 42.

The interquartile range represents a slightly more sophisticated measure of dispersion. This method divides the data into quartiles. The median is used to divide the sorted array into two halves. These halves are again divided into halves by their median. The first quartile (Q1) is the median of the lower half of the sorted array and is also referred to as the lower quartile. Q2 represents the median. Q3 is the median of the upper half of the sorted array and is referred to as the upper quartile. The difference between the upper and lower quartile is the interquartile range. In the exam score example, Q1 = 72.25 and Q3 = 86.75. Therefore, the interquartile range for this dataset is 86.75 − 72.25 = 14.50.

The third measure of dispersion is the variance (s2). To calculate the variance, subtract the raw value of each exam score from the mean of the exam scores. As you may guess, some of the differences will be positive, and some will be negative, resulting in the sum of differences equaling zero. As we are more interested in the magnitude of differences (or deviations) from the mean, one method to overcome this “zeroing” property is to square each deviation, thus removing the negative values from the output. This results in the following:

We then divide the sum of squares by either n − 1 (in the case of working with a sample) or n (in the case of working with a population). As the exam scores given here represent the entire class population, we will employ this figure on variance, which results in a variance of s2 = 116.4. However, we would be working with a population sample if we wanted to use these exam scores to extrapolate information about the larger student body. In that case, we would divide the sum of squares by n − 1.

Standard deviation, the final measure of dispersion discussed here, is the most commonly used measure of dispersion. To compensate for the squaring of each different from the mean performed during the variance calculation, standard deviation takes the square root of the variance. As determined from the figure on standard deviation, our exam score example shows a standard deviation of s = SQRT(116.4) = 10.8.

Calculating the standard deviation allows us to make some notable inferences about the dispersion of our dataset. A slight standard deviation suggests the values in the dataset are clustered around the mean, while a significant standard deviation suggests the values are scattered widely around the mean. Additional inferences about the standard deviation may be made if the dataset conforms to a normal distribution. When placed into a frequency distribution (histogram), a normal distribution implies that the data looks symmetrical or “bell-shaped.” When not “normal,” the frequency distribution of the dataset is said to be positively or negatively “skewed..” Skewed data maintain values that are not symmetrical around the mean. Regardless, normally distributed data maintains the property of having approximately 68 percent of the data values fall within ± 1 standard deviation of the mean, and 95 percent of the data value fall within ± 2 standard deviations of the mean. In our example, the mean is 78, and the standard deviation is 10.8. It can therefore be stated that 68 percent of the scores fall between 67.2 and 88.8 (i.e., 78 ± 10.8), while 95 percent of the scores fall between 56.4 and 99.6 (i.e., 78 ± [10.8 * 2]). For datasets that do not conform to the standard curve, it can be assumed that 75 percent of the data values fall within ± 2 standard deviations of the mean.

6.3: Searches and Queries

Access to robust search and query tools is essential to examining the general trends of a dataset. Queries are essentially questions posed to a database. The selective display and retrieval of information based on these queries are essential components of any geographic information system (GIS). There are three basic methods for searching and querying attribute data: (1) selection, (2) query by attribute, and (3) query by geography.

Selection

Selection represents the easiest way to search and query spatial data in a GIS. Selecting features highlights those attributes of interest, both on-screen and in the attribute table, for subsequent display or analysis. For example, one selects points, lines, and polygons simply by using the cursor to "point-and-click" the feature of interest or by dragging a box around those features. Alternatively, one can select features by using a graphic object, such as a circle, line, or polygon, to highlight all of those features that fall within the object. Advanced options for selecting subsets of data from the larger dataset include creating a new selection, selecting from the currently selected features, adding to the current selection, and removing from the current selection.

Query by Attribute

Map features and their associated data can be retrieved via the query of attribute information within the data tables. For example, search and query tools allow a user to show all the census tracts with a population density of 500 or greater, show all counties that are less than or equal to 100 square kilometers, or show all convenience stores within 1 mile of an interstate highway.

Specifically, SQL (Structured Query Language) is a computer language developed to query attribute data within a relational database management system. Created by IBM in the 1970s, SQL allows retrieving a subset of attribute information based on specific, user-defined criteria via implementing particular language elements. More recently, the use of SQL has been extended for use in a GIS (Shekhar and Chawla 2003).[1] One important note related to the use of SQL is that the exact expression used to query a dataset depends on the GIS file format being examined. For example, ANSI SQL is a particular version used to query ArcSDE geodatabases, while Jet SQL is used to access personal geodatabases. Similarly, shapefiles, coverages, and dBASE tables use a restricted version of SQL that does not support all the features of ANSI SQL or Jet SQL.

As discussed in Chapter 5, "Geospatial Data Management," Section 5.2 "Geospatial Database Management," all attribute tables in a relational database management system (RDBMS) used for an SQL query must contain primary and foreign keys for proper use. In addition to these keys, SQL implements clauses to structure database queries. A clause is a language element that includes the SELECT, FROM, WHERE, ORDER BY, and HAVING query statements.

- SELECT denotes what attribute table fields you wish to view.
- FROM denotes the attribute table in which the information resides.
- WHERE denotes the user-defined criteria for the attribute information that must be met in order for it to be included in the output set.
- ORDER BY denotes the sequence in which the output set will be displayed.
- HAVING denotes the predicate used to filter output from the ORDER BY clause.

While the SELECT and FROM clauses are mandatory statements in an SQL query, the WHERE is an optional clause to limit the output set. Likewise, the ORDER BY and HAVING are optional clauses used to present the information in an interpretable manner.

The following is a series of SQL expressions and results applied to "Personal Addresses in "ExampleTable" Attribute Table." The title of the attribute table is "ExampleTable." Note that the asterisk (*) denotes a special case of SELECT whereby all columns for a given record are selected: SELECT * FROM ExampleTable WHERE City = "Upland"

This statement returns the following:

Consider the following statement:

SELECT LastName FROM ExampleTable WHERE State = "CA" ORDER BY FirstName

This statement results in the following table sorted in ascending order by the FirstName column (not included in the output table as directed by the SELECT clause):

In addition to clauses, SQL allows for the inclusion of specific operators to delimit the query's result further. These operators can be relational, arithmetic, or Boolean and will typically appear inside conditional statements in the WHERE clause. A relational operator employs the statements equal to (=), less than (<), less than or equal to (<=), greater than (>), or greater than or equal to (>=). Arithmetic operators are those mathematical functions that include addition (+), subtraction (−), multiplication (*), and division (/). Boolean operators (also called Boolean connectors) include the statements AND, OR, XOR, and NOT. The AND connector selects records from the attribute table that satisfies both expressions. The OR connector selects records that satisfy either one or both expressions. The XOR connector selects records that satisfy one and only one of the expressions (the functional opposite of the AND connector). Lastly, the NOT connector is used to negate (or unselect) an expression that would otherwise be true. Put into the language of probability, the AND connector is used to represent an intersection OR represents a union, and NOT represents a complement. "Venn Diagram of SQL Operators" illustrates the logic of these connectors, where circles A and B represent two sets of intersecting data. Remember that SQL is a very exacting language, and minor inconsistencies in the statement, such as additional spaces, can result in a failed query.

These operators combine to provide the GIS user with powerful and flexible search and query options. With this in mind, can you determine the output set of the following SQL query as it is applied to "Histogram Showing the Frequency Distribution of Exam Scores"?

SELECT LastName, FirstName, StreetNumber FROM ExampleTable WHERE StreetNumber >= 10000 AND StreetNumber < 100 ORDER BY LastName

The following are the results:

Query by Geography

Query by geography, also known as a "spatial query," allows one to highlight particular features by examining their position relative to other features. For example, a GIS provides robust tools that determine the number of schools within 10 miles of a home. Several spatial query options are available, as outlined here. Throughout this discussion, the "target layer" refers to the feature dataset whose attributes are selected, while the "source layer" refers to the feature dataset on which the spatial query is applied. For example, if we used a state boundary polygon feature dataset to select highways from a line feature dataset (e.g., select all the highways that run through Arkansas), the state layer is the source, while the highway layer is the target.

Intersect

This often used spatial query technique selects all features in the target layer that share a common locale with the source layer. For example, the "intersect" query allows points, lines, or polygon layers as both the source and target layers.

The highlighted blue and yellow features are selected because they intersect the red features.

Are Within a Distance Of

This technique requires the user to specify some distance value, which is then used to buffer the source layer. All features that intersect this buffer are highlighted in the target layer. The "are within a distance of" query allows points, lines, or polygon layers for the source and target layers.

The highlighted blue and yellow features are selected because they are within the selected distance of the red features; tan areas represent buffers around the various features.

Completely Contain

This spatial query technique returns those features that are entirely within the source layer. This query type does not select features with coincident boundaries. For example, the "completely contain" query allows for points, lines, or polygons as the source layer, but only polygons can be used as a target layer.

The highlighted blue and yellow features are selected because they contain the red features entirely.

Are Completely Within

This query selects those features in the target layer whose entire spatial extent occurs within the geometry of the source layer. For example, the "are completely within" query allows for points, lines, or polygons as the target layer, but only polygons can be used as a source layer.

The highlighted blue and yellow features are selected because they are entirely within the red features.

Have Their Center In

This technique selects target features whose center, or centroid, is located within the boundary of the source feature dataset. The "have their center in" query allows points, lines, or polygon layers to be used as the source and target layers.

The highlighted blue and yellow features are selected because they have their centers in the red features.

Share a Line Segment

This spatial query selects target features whose boundary geometries share a minimum of two adjacent vertices with the source layer. The "share a line segment" query allows for line or polygon layers for either source and target layers.

The highlighted blue and yellow features are selected because they share a line segment with the red features.

Touch the Boundary Of

This methodology is similar to the INTERSECT spatial query; however, it selects line and polygon features that share a common boundary with the target layer. The "touch the boundary of" query allows line or polygon layers to be used as both the source and target layers.

The highlighted blue and yellow features are selected because they touch the boundary of the red features.

Are Identical To

This spatial query returns features that have the exact geographic location. The "are identical to" query can be used on points, lines, or polygons, but the target layer type must be the same source layer type.

The highlighted blue and yellow features are selected because they are identical to the red features.

Are Crossed by the Outline Of

This selection criterion returns features that share a single vertex but not an entire line segment. The "are crossed by the outline of" query allows line or polygon layers to be used as source and target layers.

The highlighted blue and yellow features are selected because the outline of the red features crosses them.

Contain

This method is similar to the COMPLETELY CONTAIN spatial query; however, features in the target layer will be selected even if the boundaries overlap. For example, the "contain" query allows for point, line, or polygon features in the target layer when points are used as a source, when line and polygon target layers with a line source, and when only polygon target layers with a polygon source.

The highlighted blue and yellow features are selected because they contain the red features.

Are Contained By

This method is similar to the ARE COMPLETELY WITHIN spatial query; however, features in the target layer will be selected even if the boundaries overlap. For example, the "is contained by" query allows for point, line, or polygon features in the target layer when polygons are used as a source when pointing and line target layers with a line source, and when only point target layers with a point source.

The highlighted blue and yellow features are selected because the red features contain them.

6.4: Data Classification

The process of data classification combines raw data into predefined classes or bins. These classes may be represented in a map by some unique symbols or, in the case of choropleth maps, by a unique color or hue (for more on color and hue, see Chapter 8 “Geospatial Analysis II: Raster Data,” Section 8.1 “Basic Geoprocessing with Rasters”). Choropleth maps are thematic maps shaded with graduated colors to represent some statistical variable of interest. Although seemingly straightforward, several different classification methodologies are available to a cartographer. These methodologies break the attribute values down along various interval patterns. Monmonier (1991) noted that different classification methodologies could significantly impact the interpretability of a given map as the visual pattern presented is easily distorted by manipulating the specific interval breaks of the classification. In addition to the methodology employed, the number of classes chosen to represent the feature of interest will also significantly affect the ability of the viewer to interpret the mapped information. Including too many classes can make a map look overly complex and confusing. On the other hand, too few classes can oversimplify the map and hide important data trends. Most effective classification attempts utilize approximately four to six distinct classes.

While problems potentially exist with any classification technique, a well-constructed choropleth increases the interpretability of any given map. The following discussion outlines the classification methods commonly available in geographic information system (GIS) software packages. In these examples, we will use the US Census Bureau’s population statistics for US counties in 1997. These data are freely available on the US Census website (http://www.census.gov).

Equal Interval Classification Method

The equal interval (or equal step) classification method divides the range of attribute values into equally sized classes. The user determines the number of classes. The equal interval classification method is best used for continuous datasets such as precipitation or temperature. For example, in the case of the 1997 Census Bureau data, county population values across the United States range from 40 (Yellowstone National Park County, MO) to 9,184,770 (Los Angeles County, CA) for a total range of 9,184,770 − 40 = 9,184,730. If we decide to classify this data into five equal interval classes, the range of each class will cover a population spread of 9,184,730 / 5 = 1,836,946. The advantage of the equal interval classification method is that it creates a legend that is easy to interpret and present to a non-technical audience. The primary disadvantage is that specific datasets will end up with most data values falling into only one or two classes, while few to no values will occupy the other classes. For example, as shown in the figure “Equal Interval Classification for 1997 US County Population Data”, almost all the counties are assigned to the first (yellow) bin.

Quantile Classification Method

The quantile classification method places equal numbers of observations into each class. This method is best for data that is evenly distributed across its range. Figure “Quantiles” shows the quantile classification method with five total classes. As there are 3,140 counties in the United States, each class in the quantile classification methodology will contain 3,140 / 5 = 628 different counties. The advantage of this method is that it often excels at emphasizing the relative position of the data values (i.e., which counties contain the top 20 percent of the US population). The primary disadvantage of the quantile classification methodology is that features placed within the same class can have wildly differing values, mainly if the data are not evenly distributed across its range. In addition, the opposite can also happen, whereby values with small range differences can be placed into different classes, suggesting a broader difference in the dataset than exists.

Natural Breaks (or Jenks) Classification Method

The natural breaks (or Jenks) classification method utilizes an algorithm to group values in classes that are separated by distinct breakpoints. This method is best used with unevenly distributed data but not skewed toward either end of the distribution. The “Natural Breaks” figure shows the natural breaks classification for the 1997 US county population density data. One potential disadvantage is that this method can create classes containing widely varying number ranges.

Accordingly, class 1 is characterized by a range of just over 150,000, while class 5 is characterized by over 6,000,000. It is often helpful to “tweak” the classes following the classification effort or change the labels to some ordinal scale such as “small, medium, or large.” The latter example, in particular, can result in a map that is more comprehensible to the viewer. A second disadvantage is that comparing two or more maps created with the natural breaks classification method can be challenging because the class ranges are particular to each dataset. In these cases, datasets that may not be overly disparate may appear in the output graphic.

Standard Deviation Classification Method

Finally, the standard deviation classification method forms each class by adding and subtracting the standard deviation from the mean of the dataset. The method is best suited for data that conforms to a normal distribution. In the county population example, the mean is 85,108, and the standard deviation is 277,080. Therefore, as shown in the figure on “Standard Deviation,” the central class contains values within a 0.5 standard deviation of the mean, while the upper and lower classes contain values of 0.5 or more standard deviations above or above the mean.

In conclusion, several viable data classification methodologies can be applied to choropleth maps. Although other methods are available (e.g., equal area, optimal), those outlined here represent the most commonly used and widely available. Each of these methods presents the data differently and highlights different aspects of the trends in the dataset. Indeed, the classification methodology and the number of classes utilized can result in wildly varying interpretations of the dataset. Therefore, it is incumbent upon you, the cartographer, to select the method that best suits the needs of the study and presents the data in as meaningful and transparent a way as possible.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

7: Vector Data Analysis

7.1: Introduction

In Chapter 6 “Data Characteristics and Visualization,” we discussed different ways to query, classify, and summarize information in attribute tables. These methods are indispensable for understanding the fundamental quantitative and qualitative trends of a dataset. However, they do not take particular advantage of the greatest strength of a geographic information system (GIS), notably the direct spatial relationships. Spatial analysis is a fundamental component of a GIS that allows for an in-depth study of a dataset or datasets’ topological and geometric properties. This chapter discusses the basic spatial analysis techniques for vector datasets.

Learning Objectives

- Familiarize yourself with concepts and terms related to the variety of single overlay analysis techniques available to analyze and manipulate the spatial attributes of a vector feature dataset.
- Explain the concepts and terms related to implementing basic multiple-layer operations and methodologies used on vector feature datasets.

GTCM Alignment

Chapter Sections

- 7.1 Introduction
- 7.2 Single Layer Analysis
- 7.3 Multiple Layer Analysis
- 7.4 References

7.2: Single Layer Analysis

As the name suggests, single-layer analyses are undertaken on an individual feature dataset. Buffering is creating an output polygon layer containing a zone (or zones) of a specified width around an input point, line, or polygon feature. Buffers are particularly suited for determining the area of influence around features of interest. Geoprocessing is a suite of tools provided by much geographic information system (GIS) software packages that allow the user to automate many mundane tasks associated with manipulating GIS data. Geoprocessing usually involves the input of one or more feature datasets, followed by a spatially explicit analysis, resulting in an output feature dataset.

Buffering

Buffers are standard vector analysis tools used to address questions of proximity in a GIS and can be used on points, lines, or polygons (Figure 7.1 “Buffers around Red Point, Line, and Polygon Features”). For instance, suppose that a natural resource manager wants to ensure that no areas are disturbed within 1,000 feet of the breeding habitat for the federally endangered Delhi Sands flower-loving fly. Unfortunately, this species is found only in the few remaining Delhi Sands soil formations of the western United States. To accomplish this task, a 1,000-foot protection zone (buffer) could be created around all the observed point locations of the species. Alternatively, the manager may decide that there is insufficient point-specific location information related to this rare species and decide to protect all Delhi Sands soil formations. In this case, he or she could create a 1,000-foot buffer around all polygons labeled as “Delhi Sands” on a soil formations dataset. In either case, buffers provide a quick-and-easy tool for determining which areas are to be maintained as preserved habitats for the endangered fly.

Several buffering options are available to refine the output. For example, the buffer tool will typically buffer only selected features. If no features are selected, all features will be buffered. Two primary buffers are available to GIS users: constant width and variable width. Constant width buffers require users to input a value by which features are buffered (Figure 7.1 “Buffers around Red Point, Line, and Polygon Features”), as seen in the preceding paragraph examples. Variable width buffers, on the other hand, call on a premade buffer field within the attribute table to determine the buffer width for each specific feature in the dataset (Figure 7.2 “Additional Buffer Options around Red Features: (a) Variable Width Buffers, (b) Multiple Ring Buffers, (c) Doughnut Buffer, (d) Setback Buffer, (e) Nondissolved Buffer, (f) Dissolved Buffer”).

In addition, users can choose to dissolve or not dissolve the boundaries between overlapping, coincident buffer areas. Multiple ring buffers can be made such that a series of concentric buffer zones (much like an archery target) is created around the originating feature at user-specified distances (Figure 7.2 “Additional Buffer Options around Red Features: (a) Variable Width Buffers, (b) Multiple Ring Buffers, (c) Doughnut Buffer, (d) Setback Buffer, (e) Nondissolved Buffer, (f) Dissolved Buffer”). In the case of polygon layers, buffers can be created that include the originating polygon feature as part of the buffer, or they are created as a doughnut buffer that excludes the input polygon area. Setback buffers are similar to doughnut buffers; however, they only buffer the area inside of the polygon boundary. For example, linear features can be buffered on both sides of the line, only on the left or right. Linear features can also be buffered so that the endpoints of the line are rounded (ending in a half-circle) or flat (ending in a rectangle).

“Geoprocessing” is a loaded term in the field of GIS. The term can (and should) be widely applied to any attempt to manipulate GIS data. However, the term came into common usage due to its application to a somewhat arbitrary suite of single layer and multiple layer analytical techniques in the Geoprocessing Wizard of ESRI’s ArcView software package in the mid-1990s. Regardless, the suite of geoprocessing tools available in a GIS greatly expands and simplifies many of the management and manipulation processes associated with vector feature datasets. The primary use of these tools is to automate the repetitive preprocessing needs of typical spatial analyses and assemble exact graphical representations for subsequent analysis and inclusion in presentations and final mapping products. The union, intersect, symmetrical difference, and identity overlay methods discussed in Section 7.2.2 “Other Multilayer Geoprocessing Options” are often used with these geoprocessing tools. The following represents the most common geoprocessing tools.

The dissolve operation combines adjacent polygon features in a single feature dataset based on a predetermined attribute. For example, part (a) of figure 7.3, “Single Layer Geoprocessing Functions,” shows the boundaries of seven different parcels of land owned by four different families (labeled 1 through 4). The dissolve tool automatically combines all adjacent features with the same attribute values. The result is an output layer with the same extent as the original but without all of the unnecessary, intervening line segments. The dissolved output layer is easier to visually interpret when the map is classified according to the dissolved field.

The append operation creates an output polygon layer by combining the spatial extent of two or more layers (part (d) of Figure 7.3 “Single Layer Geoprocessing Functions”). For use with point, line, and polygon datasets, the output layer will be the same feature type as the input layers (which must each be the same feature type as well). Unlike the dissolve tool, the append does not remove the boundary lines between appended layers (in the case of lines and polygons). Therefore, it is often helpful to perform a dissolve after using the append tool to remove these potentially unnecessary dividing lines. Append is frequently used to mosaic data layers, such as digital US Geological Survey (USGS) 7.5-minute topographic maps, to create a single map for analysis and display.

The select operation creates an output layer based on a user-defined query that selects particular features from the input layer (part (f) of Figure 7.3 “Single Layer Geoprocessing Functions”). The output layer contains only those features that are selected during the query. For example, a city planner may choose to select all areas that are zoned “residential” so he or she can quickly assess which areas in town are suitable for a proposed housing development.

Finally, the merge operation combines features within a point, line, or polygon layer into a single feature with identical attribute information. Often, the original features will have different values for a given attribute. The first attribute encountered is carried over into the attribute table, and the remaining attributes are lost. This operation is beneficial when polygons are found to be unintentionally overlapping. Merge will conveniently combine these features into a single entity.

7.3: Multiple Layer Analysis

Among the most powerful and commonly used tools in a geographic information system (GIS) is the overlay of cartographic information. In a GIS, an overlay is a process of taking two or more different thematic maps of the same area and placing them on top of one another to form a new map. Inherent in this process, the overlay function combines the spatial features of the dataset and the attribute information.

A typical example of the overlay process is, "Where is the best place to put a mall?" Imagine you are a corporate bigwig and are tasked with determining where your company's next shopping mall will be placed. How would you attack this problem? With a GIS at your command, answering such spatial questions begins with amassing and overlaying pertinent spatial data layers. For example, you may first want to determine what areas can support the mall by accumulating information on which land parcels are for sale and zoned for commercial development.

After collecting and overlaying the baseline information on available development zones, you can determine which areas offer the most economic opportunity by collecting regional information on average household income, population density, location of proximal shopping centers, local buying habits, and more. Next, you may want to collect information on restrictions or roadblocks to development, such as the cost of land, cost to develop the land, community response to the development, adequacy of transportation corridors to and from the proposed mall, tax rates, and so forth. Indeed, simply collecting and overlaying spatial datasets provides a valuable tool for visualizing and selecting the optimal site for such a business endeavor.

Overlay Operations

Several basic overlay processes are available in a GIS for vector datasets: point-in-polygon, polygon-on-point, line-on-line, line-in-polygon, polygon-on-line, and polygon-on-polygon. As you may be able to divine from the names, one of the overlay datasets must always be a line or polygon layer, while the second may be point, line, or polygon. The new layer produced following the overlay operation is termed the "output" layer.

Point-in-Polygon Overlay Operation

The point-on-polygon overlay operation requires a point input layer and a polygon overlay layer. Upon performing this operation, a new output point layer is returned that includes all the points that occur within the spatial extent of the overlay (Figure 7.4 "A Map Overlay Combining Information from Point, Line, and Polygon Vector Layers, as Well as Raster Layers"). In addition, all the points in the output layer contain their original attribute information and the attribute information from the overlay. For example, suppose you were tasked with determining if an endangered species residing in a national park was found primarily in a particular vegetation community. The first step would be to acquire the point occurrence locales for the species in question, plus a polygon overlay layer showing the vegetation communities within the national park boundary. Upon performing the point-in-polygon overlay operation, a new point file contains all the points that occur within the national park. The attribute table of this output point file would also contain information about the vegetation communities being utilized by the species at the time of observation. A quick scan of this output layer and its attribute table would allow you to determine where the species was found in the park and review the vegetation communities in which it occurred. This process would enable park employees to make informed management decisions regarding which onsite habitats to protect to ensure continued site utilization by the species.

Polygon-on-Point Overlay Operation

As its name suggests, the polygon-on-point overlay operation is the opposite of the point-in-polygon operation. In this case, the polygon layer is the input, while the point layer is the overlay. The polygon features that overlay these points are selected and preserved in the output layer. For example, given a point dataset containing the locales of some crime and a polygon dataset representing city blocks, a polygon-on-point overlay operation would allow police to select the city blocks in which crimes have been known to occur and hence determine those locations where an increased police presence may be warranted.

Line-on-Line Overlay Operation

A line-on-line overlay operation requires line features for the input and overlay layer. The output from this operation is a point or points located precisely at the intersection(s) of the two linear datasets (Figure 7.7 "Line-on-Line Overlay"). For example, a linear feature dataset containing railroad tracks may be overlaid on a linear road network. The resulting point dataset contains all the locales of the railroad crossings over a town's road network. The attribute table for this railroad crossing point dataset would contain information on the railroad and the road it passed.

Line-in-Polygon Overlay Operation

The line-in-polygon overlay operation is similar to the point-in-polygon overlay, with the obvious exception that a line input layer is used instead of a point input layer. In this case, each line that has any part of its extent within the overlay polygon layer will be included in the output line layer, although these lines will be truncated at the boundary of the overlay (Figure 7.9 "Polygon-on-Line Overlay"). For example, a line-in-polygon overlay can take an input layer of interstate line segments and a polygon overlay representing city boundaries and produce a linear output layer of highway segments within the city boundary. The attribute table for the output interstate line segment will contain information on the interstate name and the city through which they pass.

Polygon-on-Line Overlay Operation

The polygon-on-line overlay operation is the opposite of the line-in-polygon operation. In this case, the polygon layer is the input, while the line layer is the overlay. The polygon features that overlay these lines are selected and subsequently preserved in the output layer. For example, given a layer containing the path of a series of telephone poles/wires and a polygon map containing city parcels, a polygon-on-line overlay operation would allow a land assessor to select those parcels containing overhead telephone wires.

Polygon-in-Polygon Overlay Operation

Finally, the polygon-in-polygon overlay operation employs a polygon input and a polygon overlay. This is the most commonly used overlay operation. Using this method, the polygon input and overlay layers are combined to create an output polygon layer with the extent of the overlay. The attribute table will contain spatial data and attribute information from the input and overlay layers (Figure 7.10 "Polygon-in- Polygon Overlay"). For example, you may choose an input polygon layer of soil types with an overlay of agricultural fields within a given county. The output polygon layer would contain information on the location of agricultural fields and soil types throughout the county.

The overlay operations discussed previously assume that the user desires to combine the overlain layers. This is not always the case. Overlay methods can be more complex and employ the basic Boolean operators: AND, OR, and XOR (see Section 6.1.2 "Measures of Central Tendency"). Depending on which operator(s) are utilized, the overlay method will result in an intersection, union, symmetrical difference, or identity.

Specifically, the union overlay method employs the OR operator. A union can be used only in the case of two polygon input layers. It preserves all features, attribute information, and spatial extents from both input layers (part (a) of Figure 7.11 "Vector Overlay Methods "). This overlay method is based on the polygon-in-polygon operation described in Section 7.1.1 "Buffering."

Alternatively, the intersection overlay method employs the AND operator. An intersection requires a polygon overlay but can accept a point, line, or polygon input. The output layer covers the spatial extent of the overlay and contains features and attributes from both the input and overlay (part (b) of Figure 7.11 "Vector Overlay Methods ").

The symmetrical difference overlay method employs the XOR operator, which results in the opposite output as an intersection. This method requires both input layers to be polygons. The output polygon layer produced by the symmetrical difference method represents those areas common to only one of the feature datasets (part (c) of Figure 7.11 "Vector Overlay Methods ").

In addition to these simple operations, the identity (also referred to as "minus") overlay method creates an output layer with the spatial extent of the input layer (part (d) of Figure 7.11 "Vector Overlay Methods ") but includes attribute information from the overlay (referred to as the "identity" layer, in this case). The input layer can be points, lines, or polygons. The identity layer must be a polygon dataset.

Other Multilayer Geoprocessing Options

In addition to the vector mentioned above overlay methods, other standard multiple-layer geoprocessing options are available to the user. These included the clip, erase, and split tools. The clip geoprocessing operation is used to extract those features from an input point, line, or polygon layer that falls within the spatial extent of the clip layer (part (e) of Figure 7.11 "Vector Overlay Methods "). Following the clip, all attributes from the preserved portion of the input layer are included in the output. If any features are selected during this process, only those selected features within the clip boundary will be included in the output. For example, the clip tool could clip the extent of a river floodplain by the extent of a county boundary. This would provide county managers with insight into which portions of the floodplain they are responsible for maintaining. This is similar to the intersect overlay method; however, the attribute information associated with the clip layer is not carried into the output layer following the overlay.

The erase geoprocessing operation is essentially the opposite of a clip. Whereas the clip tool preserves areas within an input layer, the erase tool preserves only those areas outside the extent of the analogous erase layer (part (f) of Figure 7.11 “Vector Overlay Methods “). While the input layer can be a point, line, or polygon dataset, the erase layer must be a polygon dataset. Continuing with our clip example, county managers could then use the erase tool to erase the areas of private ownership within the county floodplain area. Officials could then focus specifically on public reaches of the countywide floodplain for their upkeep and maintenance responsibilities.

The split geoprocessing operation is used to divide an input layer into two or more layers based on a split layer (part (g) of Figure 7.11 “Vector Overlay Methods “). The split layer must be a polygon, while the input layers can be a point, line, or polygon. For example, a homeowner’s association may choose to split up a countywide soil series map by parcel boundaries, so each homeowner has a specific soil map for their parcel.

Spatial Join

A spatial join is a hybrid between an attribute operation and a vector overlay operation. Like the “join” attribute operation described in Section 5.2.2, “Joins and Relates,” a spatial join results in combining two feature dataset tables by a common attribute field. However, unlike the attribute operation, a spatial join determines which fields from a source layer’s attribute table are appended to the destination layer’s attribute table based on the relative locations of selected features. This relationship is explicitly based on the property of proximity or containment between the source and destination layers rather than the primary or secondary keys. The proximity option is used when the source layer is a point or line feature dataset, while the containment option is used when the source layer is a polygon feature dataset.

When employing the proximity (or “nearest”) option, a record for each feature in the source layer’s attribute table is appended to the closest given feature in the destination layer’s attribute table. In addition, the proximity option will typically add a numerical field to the destination layer attribute table, called “Distance,” within which the measured distance between the source and destination feature is placed. For example, suppose a city agency had a point dataset showing all known polluters in town and a line dataset of all the river segments within the municipal boundary. This agency could then perform a proximity-based spatial join to determine the nearest river segment that each polluter would most likely affect.

When using the containment (or “inside”) option, a record for each feature in the polygon source layer’s attribute table is appended to the record in the destination layer’s attribute table. No value will be appended if a destination layer feature (point, line, or polygon) is not entirely contained within a source polygon. For example, suppose a pool cleaning business wanted to hone its marketing services by providing flyers only to homes that owned a pool. They could obtain a point dataset containing the location of every pool in the county and a polygon parcel map for that same area. That business could then conduct a spatial join to append the parcel information to the pool locales. This would provide them with information on each land parcel that contained a pool, and they could subsequently send their mailers only to those homes.

Overlay Errors

Although overlays are one of the essential tools in a GIS analyst’s toolbox, some problems can arise when using this methodology. In particular, slivers are a standard error produced when two slightly misaligned vector layers overlap (Figure 7.12 “Slivers”). This misalignment can come from several sources, including digitization errors, interpretation errors, or source map errors (Chang, 2008). [1] For example, most vegetation and soil maps are created from field survey data, satellite images, and aerial photography. While you can imagine that the boundaries of soils and vegetation frequently coincide, the fact that different researchers most likely created them at different times suggests that their boundaries will not overlap perfectly. GIS software incorporates a cluster tolerance option that forces nearby lines to be snapped together if they fall within a user-specified distance to ameliorate this problem. Care must be taken when assigning cluster tolerance. Too strict a setting will not snap shared boundaries, while too lenient a setting will snap unintended, neighboring boundaries together (Wang and Donaghy 1995).

A second potential source of error associated with the overlay process is error propagation. Error propagation arises when inaccuracies are present in the original input and overlay layers and are propagated through to the output layer (MacDougall, 1975). For example, these errors can be related to positional inaccuracies of the points, lines, or polygons. Alternatively, they can arise from attribute errors in the original data table(s). Regardless of the source, error propagation represents a common problem in overlay analysis, which depends mainly on the accuracy and precision requirements of the project at hand.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

8: Raster Data and Imagery Analysis

8.1: Introduction

Following our attribute and vector data analysis discussion, raster data analysis presents the final powerful data mining tool available to geographers. Raster data are particularly suited to specific analyses, such as basic geoprocessing, surface analysis, and terrain mapping. While not always true, raster data can simplify many spatial analyses that would otherwise be overly cumbersome to perform on vector datasets. Some of the most common of these techniques are presented in this chapter.

Learning Objectives

- Explain the single and multiple raster geoprocessing techniques
- Describe how local, neighborhood, zonal, and global analysis can be applied to raster datasets.
- Describe the concepts and terms related to GIS surfaces, how to create them, and how they are used to answer specific spatial and temporal questions.
- Explain how to apply fundamental raster surface analyses to terrain mapping applications.

GTCM Alignment

Chapter Sections

8.2: Geoprocessing with Raster Imagery

Raster data can undergo similar spatial operations to the geoprocessing tools available for vector datasets. Although the actual computation of these operations is significantly different from their vector counterparts, their conceptual underpinning is similar. The geoprocessing techniques covered here include both single layer (Section 8.1.1 “Single Layer Analysis”) and multiple-layer (Section 8.1.2 “Multiple Layer Analysis”) operations.

Single Layer Analysis

Reclassifying or recoding a dataset is commonly one of the first steps undertaken during raster analysis. Reclassification is the single layer process of assigning a new class or range value to all pixels in the dataset based on their original values (Figure 8.1 “Raster Reclassification.” For example, an elevation grid commonly contains a different value for nearly every cell within its extent. These values could be simplified by aggregating each pixel value in a few discrete classes (i.e., 0–100 = “1,” 101–200 = “2,” 201– 300 = “3,” etc.). This simplification allows for fewer unique values and cheaper storage requirements. In addition, these reclassified layers are often used as inputs in secondary analyses, such as those discussed later in this section.

As described in Chapter 7, “Geospatial Analysis I: Vector Operations,” buffering is creating an output dataset that contains a zone (or zones) of a specified width around an input feature. In the case of raster datasets, these input features are given as a grid cell or a group of grid cells containing a uniform value (e.g., buffer all cells whose value = 1). Buffers are particularly suited for determining the area of influence around features of interest. Whereas buffering vector data results in a precise area of influence at a specified distance from the target feature, raster buffers tend to be approximations representing those cells that are within the specified distance range of the target (Figure 8.2 “Raster Buffer around a Target Cell(s)”). Most geographic information system (GIS) programs calculate raster buffers by creating a grid of distance values from the center of the target cell(s) to the center of the neighboring cells and then reclassifying those distances such that a “1” represents those cells composing the original target, a “2” represents those cells within the user-defined buffer area, and a “0” represents those cells outside of the target and buffer areas. These cells could also be further classified to represent multiple ring buffers by including values of “3,” “4,” “5,” and so forth to represent concentric distances around the target cell(s).

A raster dataset can also be clipped, similar to a vector dataset (Figure 8.3 “Clipping a Raster to a Vector Polygon Layer”). Here, the input raster is overlain by a vector polygon clip layer. The raster clip process results in a single raster that is identical to the input raster but shares the extent of the polygon clip layer.

Raster overlays are relatively simple compared to their vector counterparts and require less computational power (Burroughs, 1983). However, despite their simplicity, it is essential to ensure that all overlain rasters are coregistered (i.e., spatially aligned), cover identical areas, and maintain equal resolution (i.e., cell size). If these assumptions are violated, the analysis will fail, or the resulting output layer will be flawed. With this in mind, there are several different methodologies for performing a raster overlay (Chrisman, 2002).

The mathematical raster overlay is the most common overlay method. The numbers within the aligned cells of the input grids can undergo any user-specified mathematical transformation. Following the calculation, an output raster contains a new value for each cell (Figure 8.4 “Mathematical Raster Overlay”). As you can imagine, there are many uses for such functionality. In particular, raster overlay is often used in risk assessment studies where various layers are combined to produce an outcome map showing high risk/reward areas.

The Boolean raster overlay method represents a second powerful technique. As discussed in Chapter 6, “Data Characteristics and Visualization,” the Boolean connectors AND, OR, and XOR can be employed to combine the information of two overlying input raster datasets into a single output raster. Similarly, the relational raster overlay method utilizes relational operators (<, <=, =, <>, >, and =>) to evaluate the conditions of the input raster datasets. In both the Boolean and relational overlay methods, cells that meet the evaluation criteria are coded in the output raster layer with a 1. At the same time, those evaluated as false receive a value of 0.

However, the simplicity of this methodology can also lead to easily overlooked errors in interpretation if the overlay is not appropriately designed. For example, assume that a natural resource manager has two input raster datasets she plans to overlay; one showing the location of trees (“0” = no tree; “1” = tree) and one showing the location of urban areas (“0” = not urban; “1” = urban). Suppose she hopes to find the location of trees in urban areas. In that case, a simple mathematical sum of these datasets will yield a “2” in all pixels containing a tree in an urban area. Similarly, if she hopes to find the location of all treeless (or “non-tree,” nonurban areas, she can examine the summed output raster for all “0” entries, and finally, suppose she hopes to locate urban, treeless areas. In that case, she will look for all cells containing a “1.” Unfortunately, the cell value “1” also is coded into each pixel for nonurban tree cells. Indeed, the input pixel values and overlay equation in this example will yield confounding results due to the poorly devised overlay scheme.

8.3: Scale of Raster Analysis

Raster analyses can be undertaken on four different scales of operation: local, neighborhood, zonal, and global. Each of these presents unique options to the GIS analyst and is presented here in this section.

Local Operations

Local operations can be performed on single or multiple rasters. When used on a single raster, a local operation usually applies some mathematical transformation to each cell in the grid. For example, a researcher may obtain a digital elevation model (DEM) with each cell value representing elevation in feet. However, suppose it is preferred to represent those elevations in meters. In that case, a simple arithmetic transformation (original elevation in feet * 0.3048 = new elevation in meters) of each cell value can be performed locally to accomplish this task.

When applied to multiple rasters, it becomes possible to perform such analyses as changes over time. For example, given two rasters containing information on groundwater depth on a parcel of land in the years 2000 and 2010, it is simple to subtract these values and place the difference in an output raster that will note the change in groundwater between those two times (Figure 8.5 "Local Operation on a Raster Dataset"). However, these local analyses can become somewhat more complicated as the number of input rasters increases. For example, the Universal Soil Loss Equation (USLE) applies a local mathematical formula to several overlying rasters, including rainfall intensity, soil erodibility, slope, cultivation type, and vegetation type, to determine the average soil loss (in tons) in a grid cell.

Tobler's first law of geography states that "everything is related to everything else, but near things are more than distant things." Neighborhood operations represent a group of frequently used spatial analysis techniques that rely heavily on this concept. Neighborhood functions examine the relationship of an object with similar surrounding objects. They can be performed on point, line, or polygon vector datasets and raster datasets. In the case of vector datasets, neighborhood analysis is most frequently used to perform basic searches. For example, given a point dataset containing the location of convenience stores, a GIS could be employed to determine the number of stores within 5 miles of a linear feature (i.e., Interstate 10 in California).

Neighborhood analyses are often more sophisticated when used with raster datasets. Raster analyses employ moving windows, also called filters or kernels, to calculate new cell values for every location throughout the raster layer's extent. These moving windows can take many different forms depending on the desired output type and the phenomena being examined. For example, a rectangular, 3-by-3 moving window is commonly used to calculate the mean, standard deviation, sum, minimum, maximum, or range of values immediately surrounding a given "target" cell (Figure 8.6 "Common Neighborhood Types around Target Cell "x": (a) 3 by 3, (b) Circle, (c) Annulus, (d) Wedge"). The target cell is that cell found in the center of the 3-by-3 moving window. The moving window passes over every cell in the raster. As it passes each central target cell, the nine values in the 3-by-3 window are used to calculate a new value for that target cell. This new value is placed in the exact location in the output raster. If one wanted to examine a larger sphere of influence around the target cells, the moving window could be expanded to 5 by 5, 7 by 7, etc. Additionally, the moving window need not be a simple rectangle. Other shapes used to calculate neighborhood statistics include the annulus, wedge, and circle (Figure 8.6 "Common Neighborhood Types around Target Cell "x": (a) 3 by 3, (b) Circle, (c) Annulus, (d) Wedge").

Neighborhood operations are commonly used for data simplification on raster datasets. For example, an analysis that averages neighborhood values would result in a smoothed output raster with dampened highs and lows as the influence of the outlying data values is reduced by the averaging process. Alternatively, neighborhood analyses can be used to exaggerate differences in a dataset. For example, edge enhancement is a type of neighborhood analysis that examines the range of values in the moving window. A significant range value would indicate that an edge occurs within the extent of the window, while a small range indicates the lack of an edge.

Zonal Operations

A zonal operation is employed on groups of cells of similar value or like features, not surprisingly called zones (e.g., land parcels, political/municipal units, water bodies, soil/vegetation types). These zones could be conceptualized as raster versions of polygons. Zonal rasters are often created by reclassifying an input raster into a few categories (see Section 8.2.2 "Neighborhood Operations"). Zonal operations may be applied to a single raster or two overlaying rasters. Given a single input raster, zonal operations measure the geometry of each zone in the raster, such as area, perimeter, thickness, and centroid. Given two rasters in a zonal operation, one input raster, and one zonal raster, a zonal operation produces an output raster, which summarizes the cell values in the input raster for each zone in the zonal raster (Figure 8.7 “Zonal Operation on a Raster Dataset”).

Zonal operations and analyses are valuable in fields of study such as landscape ecology, where the geometry and spatial arrangement of habitat patches can significantly affect the type and number of species that can reside in them. Similarly, zonal analyses can effectively quantify the narrow habitat corridors important for the regional movement of flightless, migratory animal species moving through otherwise densely urbanized areas.

Global Operations

Global operations are similar to zonal operations, whereby the entire raster dataset’s extent represents a single zone. Typical global operations include determining fundamental statistical values for the raster. For example, the minimum, maximum, average, range, and so forth can be quickly calculated over the entire extent of the input raster and subsequently be output to a raster. Every cell contains that calculated value (Figure 8.8 “Global Operation on a Raster Dataset”).

8.4: Spatial Interpolation for Spatial Analysis

A surface is a vector or raster dataset that contains an attribute value for every locale throughout its extent. In a sense, all raster datasets are surfaces, but not all vector datasets are surfaces. Surfaces are commonly used in a geographic information system (GIS) to visualize phenomena such as elevation, temperature, slope, aspect, rainfall, and more. In a GIS, surface analyses are usually carried out on either raster datasets or TINs (Triangular Irregular Network; Chapter 5 "Geospatial Data Management," Section 5.3.1 "Vector File Formats"), but isolines or point arrays can also be used. Interpolation is used to estimate the value of a variable at an unsampled location from measurements made at nearby or neighboring locales. Spatial interpolation methods draw on the theoretical creed of Tobler's first law of geography, which states that "everything is related to everything else, but near things are more related than distant things." Indeed, this basic tenet of positive spatial autocorrelation forms the backbone of many spatial analyses (Figure 8.9 "Positive and Negative Spatial Autocorrelation").

Creating Surfaces

The ability to create a surface is a valuable tool in a GIS. However, the creation of raster surfaces often starts with the creation of a vector surface. One standard method to create such a vector surface from point data is the generation of Thiessen (or Voronoi) polygons. Thiessen polygons are mathematically generated areas that define the sphere of influence around each point in the dataset relative to all other points (Figure 8.10 "A Vector Surface Created Using Thiessen Polygons"). Polygon boundaries are calculated as the perpendicular bisectors of the lines between each pair of neighboring points. The derived Thiessen polygons can then be used as crude vector surfaces that provide attribute information across the entire area of interest. A typical example of Thiessen polygons is the creation of a rainfall surface from an array of rain gauge point locations. These Thiessen polygons can be easily converted to equivalent raster representations by some basic reclassification techniques.

While creating Thiessen polygons results in a polygon layer whereby each polygon, or raster zone, maintains a single value, interpolation is a potentially complex statistical technique that estimates the value of all unknown points between the known points. The three primary methods used to create interpolated surfaces are spline, inverse distance weighting (IDW), and trend surface. The spline interpolation method forces a smoothed curve through the set of known input points to estimate the unknown, intervening values. IDW interpolation estimates the values of unknown locations using the distance to proximal, known values. The weight placed on the value of each proximal value is in inverse proportion to its spatial distance from the target locale. Therefore, the farther the proximal point, the less weight it carries in defining the target point's value. Finally, trend surface interpolation is the most complex method. It fits a multivariate statistical regression model to the known points, assigning a value to each unknown location based on that model.

Other highly complex interpolation methods exist, such as kriging. Kriging is a complex geostatistical technique, similar to IDW, that employs semivariograms to interpolate the values of an input point layer and is more akin to regression analysis (Krige, 1951). The specifics of the kriging methodology will not be covered here as this is beyond the scope of this text. For more information on kriging, consult review texts such as Stein (1999).

Inversely, raster data can also be used to create vector surfaces. For instance, isoline maps are made up of continuous, nonoverlapping lines that connect points of equal value. Isolines have specific monikers depending on the information they model (e.g., elevation = contour lines, temperature = isotherms, barometric pressure = isobars, wind speed = isotachs). Figure 8.11 "Contour Lines Derived from a DEM" shows an isoline elevation map. As the elevation values of this digital elevation model (DEM) range from 450 to 950 feet, the contour lines are placed at 500, 600, 700, 800, and 900 feet elevations throughout the extent of the image. In this example, the contour interval, defined as the vertical distance between each contour line, is 100 feet. The user determines the contour interval during the creation of the surface.

8.5: Terrain Mapping for Spatial Analysis

Surface analysis is often referred to as terrain (elevation) analysis when information related to slope, aspect, viewshed, hydrology, volume, and so forth are calculated on raster surfaces such as DEMs (digital elevation models; Chapter 5 "Geospatial Data Management," Section 5.3.1 "Vector File Formats"). In addition, surface analysis techniques can also be applied to more esoteric mapping efforts, such as the probability of tornados or the concentration of infant mortalities in a given region. This section discusses a few methods for creating surfaces and standard surface analysis techniques related to terrain datasets.

Several standard raster-based neighborhood analyses provide valuable insights into the surface properties of the terrain. For example, slope maps (part (a) of Figure 8.12 "(a) Slope, (b) Aspect, and (c and d) Hillshade Maps") are excellent for analyzing and visualizing landform characteristics and are frequently used in conjunction with aspect maps (defined later) to assess watershed units, inventory forest resources, determine habitat suitability, estimate slope erosion potential, and more. They are typically created by fitting a planar surface to a 3-by-3 moving window around each target cell. When dividing the horizontal distance across the moving window (which is determined via the spatial resolution of the raster image) by the vertical distance within the window (measured as the difference between the most significant cell value and the central cell value), the slope is relatively easily obtained. The output raster of slope values can be calculated as either percent slope or degree of slope.

Any cell that exhibits a slope must, by definition, be oriented in a known direction. This orientation is referred to as aspect. Aspect maps (part (b) figure 8.12 "(a) Slope, (b) Aspect, and (c and d) Hillshade Maps") use slope information to produce output raster images whereby the value of each cell denotes the direction it faces. This is usually coded as either one of the eight ordinal directions (north, south, east, west, northwest, northeast, southwest, southeast) or in degrees from 1° (nearly due north) to 360° (back to due north). Flat surfaces have no aspect and are given a value of −1. A 3-by-3 moving window is used to find the highest and lowest elevations around the target cell to calculate the aspect. For example, if the highest cell value is located at the top-left of the window ("top" being due north) and the lowest value is at the bottom-right, it can be assumed that the aspect is southeast. The combination of slope and aspect information is of great value to researchers such as botanists and soil scientists because sunlight availability varies widely between north-facing and south-facing slopes. Indeed, the various light and moisture regimes resulting from aspect changes encourage vegetative and edaphic differences.

A hillshade map (part (c) of Figure 8.12 "(a) Slope, (b) Aspect, and (c and d) Hillshade Maps") represents the illumination of a surface from some hypothetical, user-defined light source (presumably, the sun). Indeed, the slope of a hill is relatively brightly lit when facing the sun and dark when facing away. Using the surface slope, aspect, angle of incoming light, and solar altitude as inputs, the hillshade process codes each cell in the output raster with an 8-bit value (0–255) increasing from black to white. As you can see in part (c) of Figure 8.12, "(a) Slope, (b) Aspect, and (c and d) Hillshade Maps," hillshade representations are an effective way to visualize the three-dimensional nature of land elevations on a two-dimensional monitor or paper map. Hillshade maps can also be used effectively as a baseline map when overlaid with a semitransparent layer, such as a false-color digital elevation model (DEM; part (d) of Figure 8.12 "(a) Slope, (b) Aspect, and (c and d) Hillshade Maps").

Viewshed analysis is a valuable visualization technique that uses the elevation value of cells in a DEM or TIN (Triangulated Irregular Network) to determine those areas that can be seen from one or more specific location(s) (part (a) of Figure 8.13 "(a) Viewshed and (b) Watershed Maps"). The viewing location can be either a point or line layer and placed at any desired elevation. The output of the viewshed analysis is a binary raster that classifies cells as either 1 (visible) or 0 (not visible). For example, in the case of two viewing locations, the output raster values would be 2 (visible from both points), 1 (visible from one point), or 0 (not visible from either point).

Additional parameters influencing the resultant viewshed map are the viewing azimuth (horizontal and vertical) and viewing radius. The horizontal viewing azimuth is the horizontal angle of the view area and is set to a default value of 360°. However, the user may want to change this value to 90° if, for example, the desired viewshed included only the area seen from an office window.

Similarly, the vertical viewing angle can be set from 0° to 180°. Finally, the viewing radius determines the distance from the viewing location to be included in the output. This parameter is typically set to infinity (functionally, this includes all areas within the DEM or TIN under examination). However, it may be decreased if, for instance, you only wanted to include the area within the 100 km broadcast range of a radio station.

Similarly, watershed analyses are a series of surface analysis techniques that define the topographic divides that drain surface water for stream networks (part (b) of Figure 8.13 "(a) Viewshed and (b) Watershed Maps"). In geographic information systems (GISs), a watershed analysis is based on a "filled" DEM input. A filled DEM contains no internal depressions (such as would be seen in a pothole, sink wetland, or quarry). A flow direction raster is created to model the direction of water movement across the surface from these inputs. A flow accumulation raster calculates the number of cells that contribute flow to each cell from the flow direction information. Generally speaking, cells with a high flow accumulation value represent stream channels, while cells with low flow accumulation represent uplands. With this in mind, a network of rasterized stream segments is created. These stream networks are based on some user-defined minimum threshold of flow accumulation. For example, it may be decided that a cell needs at least one thousand contributing cells to be considered a stream segment. Altering this threshold value will change the density of the stream network. Following the creation of the stream network, a stream link raster is calculated whereby each stream segment (line) is topologically connected to stream intersections (nodes). Finally, the flow direction and stream link raster datasets are combined to determine the output watershed raster as seen in part (b) of Figure 8.13 "(a) Viewshed and (b) Watershed Maps" (Chang, 2008). [1] Such analyses are invaluable for watershed management and hydrologic modeling.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

9: Cartographic Principles

9.1: Introduction

From projections to data management to spatial analysis, we have, up to now, focused on the more technical points of a geographic information system (GIS). This chapter is concerned less with the computational options available to the GIS user and more with the artistic options. In essence, this chapter shifts the focus away from GIS tools and toward cartographic tools, although the two are becoming more and more inextricably bound. Unfortunately, many GIS users are never exposed to the field of cartography. In these cases, the hard work of creating, maintaining, aligning, and analyzing complex spatial datasets is not truly appreciated as the final mapping product may not adequately communicate this information to the consumer. In addition, maps, like statistics, can distort information, as illustrated by Mark Monmonier’s (1996) book titled How to Lie with Maps. Indeed, a strong working knowledge of cartographic rules will assist in avoiding the potential misrepresentation of spatial information and enhance one’s ability to identify these indiscretions in other cartographers’ creations. The cartographic principles discussed herein are laid out to guide GIS users through transforming accumulated bits of GIS data into attractive, useful maps for print and display. This discussion explicitly addresses the intricacies of practical color usage (Section 9.1 “Color”), symbol selection (Section 9.2 “Symbology”), and map layout and design (Section 9.3 “Cartographic Design”).

Learning Objectives

- Describe the properties of color and how best to utilize them in your cartographic projects.
- Explain how best to utilize point, line, and polygon symbols to assist in interpreting your map and its features.
- Familiarize yourself with the properties of cartographic principles that contribute to effective map design.

GTCM Alignment

-

Chapter Sections

- 9.1 Introduction
- 9.2 Color Theory
- 9.3 Symbology
- 9.4 Cartographic Design
- 9.5 References

9.2: Color Theory

Although a high-quality map is composed of many different elements, color is one of the first components noticed by end-users. This is partly because we each have an intuitive understanding of how colors should be used to create a compelling and pleasing visual experience. Nevertheless, it is not always clear to the map-maker which colors should be used to convey the purpose of the product best. This intuition is much like listening to our favorite music. We know when a note is in tune or out of tune, but we would not necessarily have any idea of how to fix a wrong note. Color is indeed a tricky piece of the cartographic puzzle and is, not surprisingly, the most frequently criticized variable on computer-generated maps (Monmonier, 1996). [1] This section attempts to outline the essential components of color and the guidelines to most effectively employ this important map attribute.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=188#oembed-1

Color Theory

As electromagnetic radiation (ER) travels via waves from the sun (or a lightbulb) to objects on the earth, portions of the ER spectrum are absorbed, scattered, or reflected by various objects. The resulting property of the absorbed, scattered, and reflected ER is termed “color.” White is the color resulting from the full range of the visual spectrum and is therefore considered the benchmark color by which all others are measured. Black is the absence of ER. All other colors result from a partial interaction with the ER spectrum.

The three primary aspects of color that must be addressed in map making are hue, value, and saturation. Hue is the dominant wavelength or color associated with a reflecting object. Hue is the most fundamental component of color and includes red, blue, yellow, purple, etc. Value is the amount of white or black. Value is often synonymous with contrast. Variations in the amount of value for a given hue result in varying degrees of lightness or darkness for that color. Lighter colors are highly valued, while dark colors possess low value. Monochrome colors are groups of colors with the same hue but with incremental variations in value. As seen in, variations in value will typically lead the viewer’s eye from dark areas to light areas.

Saturation describes the intensity of color. Full saturation results in pure colors, while low saturation colors approach gray—variations in saturation yield different shades and tints. Shades are produced by blocking light, such as an umbrella, tree, curtain, etc. Increasing the number of shading results in grays and blacks. The tint is the opposite of shade and is produced by adding white to the color. Tints and shades are germane when using additive color models (see more on additive color models). For example, to maximize the interpretability of a map, use saturated colors to represent hierarchically prominent features and washed-out colors to represent background features.

If used properly, color can greatly enhance and support map design. Likewise, color can detract from a mapping product if abused. To use color properly, one must first consider the map’s purpose. In some cases, the use of color is not warranted. Grayscale maps can be as effective as color maps if the subject matter merits it. Regardless, there are many reasons to use color. The five primary reasons are outlined here.

Color is particularly suited to convey meaning (). For example, red is an intense color that evokes a passionate response in humans. Red has been shown to evoke physiological responses such as increasing the rate of respiration and raising blood pressure. Red is frequently associated with blood, war, violence, and love. On the other hand, blue is a color associated with calming effects. Associated with the sky or ocean, blue colors can assist in sleep and are recommended for bedrooms. Too much blue, however, can result in a lapse from calming effects into feelings of depression (i.e., having the “blues”). Green is most commonly associated with life or nature (plants). Green is undoubtedly one of the most topical colors in today’s society, with commonplace references to green construction, the Green party, going green, etc. Green, however, can also represent envy and inexperience (e.g., the green-eyed monster, greenhorn). Brown is also a natural color but more represents earth and stone. Brown can also imply dullness. Yellow is most commonly associated with sunshine and warmth, similar to red. However, yellow can also represent cowardice (e.g., yellow-bellied). Black, the absence of color, is possibly the most meaning-laden color in modern parlance. The color black purports surprisingly strong positive and negative connotations, even more than the others. Black conveys mystery, elegance, and sophistication (e.g., a black-tie affair, in the black), while also conveying loss, evil, and negativity (e.g., blackout, black-hearted, black cloud, blacklist).

The second reason to use color is for clarification and emphasis (). Warm colors, such as reds and yellows, are notable for emphasizing spatial features. These colors will often jump off the page and are usually the first to attract the reader’s eye, particularly if they are counterbalanced with cool colors, such as blues and greens (see more on warm and cool colors). In addition, the use of a hue with high saturation will stand out starkly against similar hues of low saturation.

Color use is also essential for creating a map with pleasing aesthetics (). Indeed, one of the most challenging aspects of map creation is developing an effective color palette. When looking at maps through an aesthetic lens, we are truly starting to think of our creations as artwork. Although somewhat particular to individual viewers, we all have an innate understanding of when colors in a graphic/art are aesthetically pleasing and when they are not. For example, color use is considered harmonious when colors from opposite sides of the color wheel are used (), whereas equitable use of several significant hues can create an unbalanced image.

The fourth use of color is an abstraction. Color abstraction effectively illustrates quantitative and qualitative data, particularly for thematic products such as choropleth maps. Here, colors are used solely to denote different variables' different values and may not have any particular rhyme or reason. Shows a typical thematic map with abstract colors representing different countries.

Opposite abstraction, color can also be used to represent reality. Maps showing elevation (e.g., digital elevation models or DEMs) often give false colors that approximate reality. For example, low areas are colored in variations of green to show areas of lush vegetation growth. Mid-elevations (or low-lying desert areas) are colored brown to show sparse vegetation growth. Mountain ridges and peaks are colored white to show accumulated snowfall. Watercourses and water bodies are colored blue. Unless there is a specific reason not to, natural phenomena represented on maps should permanently be colored to approximate their actual color to increase interpretability and decrease confusion.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=188#oembed-2

Color Models

Color models are systems that allow for creating a range of colors from a short list of primary colors. Color models can be additive or subtractive. Additive color models combine emitted light to display color variations and are commonly used with computer monitors, televisions, scanners, digital cameras, and video projectors. The RGB (red-green-blue) color model is the most common additive model (part (a) of). The RGB model combines light beams of the primary hues of red, green, and blue to yield additive secondary hues of magenta, cyan, and yellow. Although there is a substantive difference between pure yellow light (~580 nm) and a mixture of green and red light, the human eye perceives these signals as the same. The RGB model typically employs three 8-bit numeric values (called an RGB triplet) ranging from 0 to 255 to model colors. For instance, the RGB triplets for the pure primary and secondary colors are as follows:

- Red = (255, 0, 0)
- Green = (0, 255, 0)
- Blue = (0, 0, 255)
- Magenta = (255, 0, 255)
- Cyan = (0, 255, 255)
- Yellow = (255, 255, 0)
- Black, the absence of additive color = (0, 0, 0)
- White, the sum of all additive colors = (255, 255, 255)

Two other standard additive color models, based on the RGB model, are the HSL (hue, saturation, lightness) and HSV (hue, saturation, value) models (b and c). These models are based on cylindrical coordinate systems. The angle around the central vertical axis corresponds to the hue; the distance from the central axis corresponds to saturation. The distance along the central axis corresponds to either saturation or lightness. Because of their basis in the RGB model, both the HSL and HSV color models can be directly transformed between the three additive models. While these relatively simple additive models provide minimal computer-processing time, they do possess the disadvantage of glossing over some of the complexities of color. For example, the RGB color model does not define "absolute" color spaces, which connotes that these hues may look differently when viewed on different displays. Also, the RGB hues are not evenly spaced along the color spectrum, meaning combinations of the hues is less than exact.

In contrast to an additive model, subtractive color models involve mixing paints, dyes, or inks to create full-color ranges. These subtractive models display color, assuming that white ambient light is scattered, absorbed, and reflected from the page by the printing inks. Subtractive models, therefore, create white by restricting ink from the print surface. In addition, these models assume that using white paper as other paper colors will result in skewed hues. CMYK (cyan, magenta, yellow, black) is the most common subtractive color model and is occasionally referred to as a “four-color process” ().

Although the CMY inks are sufficient to create all of the colors of the subtractive rainbow, black ink is included in this model as it is much cheaper than using a CMY mix for all blacks (black being the most commonly printed color) and because combining CMY often results in more of a dark brown hue. The CMYK model creates color values by entering percentages for each color ranging from 0 percent to 100 percent. For example, pure red comprises 14 percent cyan, 100 percent magenta, 99 percent yellow, and 3 percent black.

As you may guess, additive models are preferred when maps are displayed on a computer monitor, while subtractive models are preferred when printing. If in doubt, it is usually best to use the RGB model as this supports a more significant percentage of the visible spectrum than the CMYK model. Once an image is converted from RGB to CMYK, the additional RGB information is irretrievably lost. If possible, collecting both RGB and CMYK versions of an image is ideal, mainly if your graphic is to be both printed and placed online. You will also want to be selective in using file formats for these color models. The JPEG and GIF graphic file formats are the best choices for RGB images, while the EPS and TIFF graphic file formats are preferred with printed CMYK images.

Color Choices

Practical color usage requires a modicum of knowledge about the color wheel. Invented by Sir Isaac Newton in 1706, the color wheel visualizes colors arranged according to their chromatic relationships. Primary hues are equidistant, with secondary and tertiary colors intervening. The red-yellow-blue color wheel is the most frequently used (); however, the magenta-yellow- the cyan wheel is the preferred choice of printmakers (for reasons described in the previous section). Primary colors cannot be created by mixing other colors; secondary colors are created by mixing two primary hues; tertiary colors are created by mixing primary and secondary hues.

Furthermore, complementary colors are placed opposite each on the wheel, while analogous colors are located proximally. Complementary colors emphasize differences. Analogues suggest harmony.

Colors can be further referred to as warm or cool. Warm colors might be seen on a bright, sunny day. Cool colors are those associated with overcast days. Warm colors are typified by hues ranging from red to yellow, including browns and tans. Cool color hues range from blue-green through blue-violet and include most gray variants. When used in mapping, it is wise to use warm and cool colors with care. Indeed, warm colors stand out, appear active, and stimulate the viewer. On the other hand, cool colors appear small, recede, and calm the viewer. As you might guess, you must apply warm colors to the map features of primary interest while using cool colors on the secondary, background, and contextual features.

Following some basic color usage guidelines is wise in light of the plethora of color schemes and options available. For example, changes in hue are best suited to visualizing qualitative data, while changes in value and saturation are effective at visualizing quantitative data. Likewise, variations in lightness and saturation are best suited to representing ordered data since these establish a hierarchy among features. In particular, a monochromatic color scale is an effective way to represent the order of data whereby light colors represent smaller data values, and dark colors represent larger values. It is best to use more light shades than dark ones as the human eye can better discern lighter shades. Also, the number of coincident colors that humans can distinguish is around seven, so be careful not to abuse the color palette in your maps. If the data being mapped has a zero point, a dichromatic scale () provides a natural breaking point with increasing color values on each end of the scale representing increasing data values.

In addition, darker colors result in more critical or pronounced graphic features (assuming the background is not overly dark). Therefore, use dark colors on features whose visual impact you wish to magnify. Finally, do not use all the spectrum colors in a single map. It is best to leave such messy, rainbow-spectacular effects to the late Jackson Pollock and his abstract expressionist ilk.

9.3: Symbology

While color is a crucial variable when choosing how to represent spatial data best, making informed decisions on the size, shape, and type of symbols is equally essential. Although raster data are restricted to symbolizing features as a single cell or as cell groupings, vector data allows for a vast array of options to symbolize points, lines, and polygons in a map. Like color, cartographers must use symbols judiciously to most effectively communicate the meaning and purpose of the map to the viewer.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=192#oembed-1

Basic Symbol Guidelines

Vector points, lines, and polygons can be symbolized in myriad ways. The guidelines in this section will help you make informed decisions on how best to represent the features on your map. The primary visual variables associated with symbolization include size, texture, pattern, and shape (Figure 9.12 "Visual Variables"). Changes to symbol size and texture are most effectively used with ordinal, interval, and ratio data. Changes to symbol pattern and shape are preferred in conjunction with nominal data.

Variations in the size of symbols are potent indicators of feature importance. Intuitively, more prominent symbols are assumed to be more important than smaller symbols. Although symbol size is most commonly associated with point features, linear symbols can effectively be altered in size by adjusting line width.

Polygon features can also benefit from resizing. Even though the polygon area cannot be changed, a point representing the polygon's centroid can be included in the map. These polygon centroids can be resized and symbolized as desired like any other point feature. Varying symbol size is moderately effective when applied to ordinal or numerical data but is ineffective with nominal data.

Symbol texture also referred to as spacing, refers to the compactness of the marks that make up the symbol. For instance, points, lines, and polygons can be filled with horizontal hash marks. The closer these hash marks are spaced within the feature symbol, the more hierarchically important the feature will appear. Varying symbol texture is most effective when applied to ordinal or numerical data but is ineffective with nominal data.

Much like texture, symbols can be filled with different patterns. These patterns are typically some artistic abstraction that may or may not attempt to visualize real-world phenomena. For example, a land-use map may change the observed fill patterns of various land types to depict the dominant plants associated with each vegetation community. Changes to symbol patterns are most often associated with polygon features, although there is limited utility in changing the fill patterns of points and lines. Varying symbol size is moderately effective when applied to ordinal or numerical data and is ineffective when applied to nominal data.

Altering symbol shapes can have dramatic effects on the appearance of map features. Point symbols are most commonly symbolized with circles. Circles tend to be the default point symbol due to their unchanging orientation, compact shape, and viewer preference. Other geometric shapes can also constitute effective symbols due to their visual stability and conservation of map space. Unless specific conditions allow, volumetric symbols (spheres, cubes, etc.) should be used sparingly as they rarely contribute more than simple, two-dimensional symbols. In addition to geometric symbols, pictograms are good representations of point features and can help to add artistic flair to a map. Pictograms should denote features of interest and should not require interpretation by the viewer (Figure 9.13 "Pictograms"). Locales that frequently employ pictograms include picnic areas, camping sites, road signs, bathrooms, airports, etc. The varying symbol shape is most effective when applied to nominal data and is moderately effective with ordinal and nominal data.

Finally, applying variations in lightness/darkness will affect the hierarchical value of a symbol. The darker the symbol, the more it stands out among lighter features. Therefore, variations in the lightness/darkness of a symbol are most effective when applied to ordinal data, are moderately effective when applied to numerical data, and are ineffective when applied to nominal data.

Keep in mind that many other visual variables can be employed in a map, depending on the cartographic software used. Therefore, it is essential to maintain a logical relationship between the symbol and the data regardless of the chosen symbology. Also, visual contrast between different mapped variables must be preserved. Indeed, the efficacy of your map will be significantly diminished if you do not ensure that its symbols are readily identifiable and look markedly different from each other.

Proportional Symbolization

In addition to the uniform symbols presented in the previous section, symbols for a single, quantitative variable can be sized proportionally to match the data values. These proportional symbols help present a relatively clear understanding of the differences in magnitude within a dataset. As the numeric values for each class increase, so too do the size of the symbol representing that class. This allows the symbol size of features to be directly related to the attribute values, whereby small points denote small and large data values.

Like proportional symbols, range graded symbols group raw data into classes, with each class represented by a differently sized symbol. Both proportional and range graded symbols are frequently used with point data, but lines and polygons can also benefit from proportional symbolization. In the case of linear datasets, the line width is most frequently used as the proportional visual variable. Polygon datasets typically summarize a quantitative variable within each polygon, place a centroid within that polygon, and proportion that centroid point symbol. Range grading should not be used if the data range for a given variable is small. In these cases, range grading will suggest more considerable differences in the data values than is merited.

The advantage of proportional symbolization is the ease with which the viewer can discriminate symbol size and thus understand variations in the data values over a given map extent. On the other hand, viewers may misjudge the magnitude of the proportional symbols if they do not pay close attention to the legend. In addition, the human eye does not see and interpret symbol size in absolute terms. For example, when proportional circles are used in maps, it is typical that the viewer will underestimate the larger circles relative to the smaller circles. Graduated symbols can be based on mathematical or perceptual scaling to address this potential pitfall. Mathematical scaling directly relates symbol size with the data value for that locale. If one value is twice as large as another, it will be represented with a symbol twice as significant. Perceptual scaling overcomes the underestimation of prominent symbols by making these symbols much larger than their actual value would indicate (Figure 9.14 "Mathematical versus Perceptual Scaling").

A disadvantage of proportional symbolization is that the symbol size can vary depending on the surrounding symbols. This is best shown via the Ebbinghaus illusion (also known as Titchener circles). As you can see in Figure 9.15 "Ebbinghaus Illusion," the central circles are both the same size but appear different due to the visual influence of the surrounding circles. If you are creating a graphic with many different symbols, this illusion can wreak havoc on the interpretability of your map.

9.4: Cartographic Design

In addition to the effective use of colors and symbols, a map that is well designed will significantly enhance its ability to relate pertinent spatial information to the viewer. Judicious use of map elements, typography/labels, and design principles will result in maps that minimize confusion and maximize interpretability. Furthermore, these components must be guided by a keen understanding of the map's purpose, intended audience, topic, scale, and production/reproduction method.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://slcc.pressbooks.pub/maps/?p=194#oembed-1

Map Elements

Chapter 9 "Cartographic Principles," Section 9.1 "Color," and Section 9.2 "Symbology," discussed visual variables specific to the spatial features of a map. However, a map is composed of many more elements than just the spatial features, each of which contributes immensely to the interpretability and flow of the overall map. This section outlines the essential map elements that should be incorporated into a "complete" map. Following Slocum et al. (2005), [1] these elements are listed in the logical order in which they should be placed on the map (Figure 9.16 "A US Map Showing Various Map Elements").

The first feature that should be placed into the map layout is the frame line. This line is a bordering box surrounding all the map elements described hereafter. All of these map elements should be balanced within the frame line. To balance a map, ensure that neither large blank spaces nor jumbled masses of information are present within the map. Similar to frame lines are neat lines. Neat lines are border boxes that are placed around individual map elements. By definition, neat lines must occur within the frame line. Both frame lines and neat lines are typically thin, black-lined boxes, but they can be altered to match the specific aesthetics of an individual map.

The mapped area is the primary geographic component of the overall map. The mapped area contains all the features and symbols used to represent the displayed spatial phenomena. A neat line typically borders the mapped area.

Insets can be considered secondary map areas encased within neat lines. These neat lines should be of different thickness or type than other line features on the map to adequately demarcate them from other map features. Insets often display the primary mapped area concerning a larger area. For example, if the primary map shows the locales of national parks with a county, an inset displaying the location of that county within the more significant state boundary may be included. Conversely, insets are also used to display areas related to the primary map, which occurs at some far-off locale. This type of inset is often used with maps of the United States, whereby Alaska and Hawaii are placed as insets to a map of the contiguous United States. Finally, insets can clarify areas where features would otherwise be overcrowded if restricted to the primary mapping area. For example, if the county map of national parks contained four small, adjacent parks, an inset could be used to expand that jumbled portion of the map to show the exact spatial extent of each of the four parks. This type of inset is frequently seen when showing the small northeastern states on a map of the entire United States.

All maps should have a title. The title is one of the first map elements to catch the viewer's eye, so care should be taken to represent the map's intent with this leading text effectively. The title should clearly and concisely explain the map's purpose and should specifically target the intended viewing audience. When overly verbose or cryptically abbreviated, a poor title will detract immensely from the interpretability of the cartographic end-product. The title should contain the most significant type on the map and be limited to one line. It should be placed at the top-center of the map unless there is a specific reason otherwise. An alternate locale for the title is directly above the legend.

The legend provides a self-explanatory definition for all symbols used within the mapped area. When developing this map element, care must be taken, as many features within a dataset can lead to an overly complex legend. Although the placement of the legend is variable, it should be placed within the white space of the map and not in such a way that it masks any other map elements. Atop the legend box is the optional legend header. The legend header should not simply repeat the information from the title, nor should it include extraneous non-legend-specific information. The symbols representing mapped features should be to the left of the explanatory text. Placing a neat line around the legend will help bring attention to the element and is recommended but not required. Be careful not to take up too much of the map with the legend while also not making it so small that it becomes difficult to read or symbols become cluttered. Removing information related to base map features (e.g., state boundaries on a US map) or readily identifiable features (e.g., highway or interstate symbols) is one effective way to minimize legend size. If a prominent legend is unavoidable, it is acceptable to place this feature outside of the map's frame line.

Attribution of the data source within the map allows users to assess from where the data are derived. Stylistically, the data source attribution should be hierarchically minimized by using a relatively small, simple font. It is also helpful to preface this map element with “Source:” to avoid confusing other typographic elements.

An indicator of scale is invaluable to provide viewers with the means to adjudicate the map’s dimensions properly. While not as important when mapping large or widely familiar locales such as a country or continent, the scale element allows viewers to measure distances on the map. The three primary representations of scale are the representational fraction, verbal scale, and bar scale (for more, see Chapter 2 “Map Anatomy,” Section 2.1 “Maps and Map Types”). The scale indicator should not be prominently displayed within the map as this element is of secondary importance.

Finally, map orientation notifies the viewer of the direction of the map. A graticule can also be included in the mapped area to clarify orientation. Most maps are made such that the top of the page points to the north (i.e., a north-up map). If your map is not north-up, there should be a good reason. Orientation is most often indicated with a north arrow, of which there are many stylistic options available in current geographic information system (GIS) software packages. One of the most commonly encountered map errors is using an overly large or ornate north arrow. North arrows should be reasonably inconspicuous as they only need to be viewed once by the reader. Ornate north arrows can be used on small-scale maps. However, simple north arrows are preferred on medium to large-scale maps to not detract from the presumably more important information appearing elsewhere.

Taken together, these map elements should work together to achieve the goal of a clear, ordered, balanced, and unified map product. Since modern GIS packages allow users to add and remove these graphic elements with little effort, care must be taken to avoid the inclination to employ these components with as little forethought as it takes to create them. The following sections provide further guidance on composing these elements on the page to honor and balance the mapped area.

Typography and Label Placement

Type is found throughout all the elements of a map. Type is similar to map symbols in many senses. Coloring effects alter typographic hierarchy as lighter type fades into the background and dark type jumps to the fore. Using all uppercase letters and bolded letters will produce more pronounced textual effects. Larger font sizes increase the hierarchical weight of the type, so ensure that the size of the type corresponds with the importance of the map feature. Use decorative fonts, bold, and italics sparingly. These fonts, as well as overly small fonts, can be challenging to read if overused.

Most importantly, always spell check your final cartographic product. After spell-checking, spell check again. You will not regret the ecstra effort.

Other typographic options for altering text include serif, sans serif, and display fonts. While serif fonts are preferred in written documents to provide horizontal guidelines, either is acceptable in a mapping application (Slocum, 2005). On the other hand, Sans serif fonts are preferred for maps viewed over the Internet.

Kerning is a practical typographic effect that alters the space between adjacent letters in a word. Decreasing the kerning of a typeset is helpful if the text is too large for the space given. Increasing the kerning is an effective way to label large map areas, particularly with all-uppercase lettering. Like kerning, changes in leading (pronounced “led-ing”) alter the vertical distance between lines of text. However, leading should not be so cramped that lines of text begin to overwrite each other, nor should it be so vast that lines of text appear unrelated. Other common typographic effects include masks, callouts, shadows, and halos (Figure 9.17 “Typographic Effects”). These effects increase the visibility and importance of the text to which they are applied.

In addition to the general typographic guidelines discussed earlier, there are specific typographic suggestions for feature labels. Labels must be placed proximal to their symbols to be directly and readily associated with their described features. Labels should maintain a consistent orientation throughout so the reader does not have to rubberneck about to read various entries. Also, avoid overprinting labels on top of other graphics or typographic features. If that is not possible, consider using a halo, mask, callout, or shadow to help the text stand out from the background. In the case of maps with many symbols, be sure that no features intervene between a symbol and its label.

Some typographic guidelines are specific to labels for point, line, and polygon features. Point labels, for example, should not employ exaggerated kerning or leading. If leader lines are used, they should not touch the point symbol, nor should they include arrowheads. Leader lines should always be represented with consistent color and thickness throughout the map. Lastly, point labels should be placed within the larger polygon in which they reside. For example, if the cities of Illinois were being mapped as points atop a state polygon layer, the label for the Chicago point symbol should occur entirely over land and not reach into Lake Michigan. As this feature is located entirely on land, so should its label.

Line labels should be placed above their associated features but should not touch them. If the linear feature is complex and meandering, the label should follow the general trend of the feature and not attempt to match the alignment of each twist and turn. If the linear feature is particularly long, it can be labeled multiple times across its length. Line labels should always read from left to right.

Polygon labels should be placed within the center of the feature whenever possible. If the increased emphasis is desired, all-uppercase letters can be effective. If all-uppercase letters are used, exaggerated kerning and leading are also appropriate to increase the hierarchical importance of the feature. If the polygon feature is too small to include text, label the feature as if it were a point symbol. However, leader lines should just enter into the feature, unlike point labels.

Map Design

Map design is a complex process that provides many variables and choices to the cartographer. The British Cartographic Society Design Group presented five "Principles of Cartographic Design" on their listserv on November 26, 1999. These principles, and a summary of each, are as follows:

Concept Before Compilation

A basic understanding of the concept and purpose of the map must be secured before the actual mapping exercise begins. Furthermore, there is no way to determine what information to include in a map without first determining who the end-user is and how the map will be used. A map without a purpose is of no use to anyone.

Hierarchy with Harmony

Important map features must appear prominent on the map. The less important features should fade into the background. Creating harmony between the primary and secondary representations on the map will lead to a quality product that will best suit the needs for which it was developed.

Simplicity from Sacrifice

It is tempting to add as much information to the graphic view upon creating a map. However, in reality, it is best to leave some stones unturned. Just as the key to good communication is brevity, it can be said that the key to good mapping is simplicity. A map can be considered complete when no other features can be removed. Less, in this instance, is more.

Maximum Information at Minimum Cost

The purpose of a map is to convey the most significant amount of information with the least amount of interpretive effort by the user. Therefore, map design should allow complex spatial relationships to be understood at a glance.

Engage the Emotion to Engage the Understanding

Well-constructed maps are essential works of art. All of the artistic and aesthetic rules outlined in this chapter serve to engage the emotive center of the viewer. The message will be lost if the viewer does not formulate some primary emotional response to the map.

It should become increasingly clear that the cartographic choices made during the mapping process influence the interpretation of a map, as does the data being mapped. Borrowing liberally from the popularized Mark Twain quote, it could be said, "There are three kinds of lies: lies, damned lies, and maps." Map-makers, indeed, can use (or misuse) cartographic principles to represent (or misrepresent) the spatial data at their disposal. It is now up to you, the cartographer, to master the tools presented in this book to harness the power of maps to elucidate and address the spatial issues with which you are confronted.

Welcome to the Geosciences Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Index

D

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Geographic Information Systems and Cartography

Webpages: 70

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 98.6% (69 pages)
- Undeclared: 1.4% (1 page)

By Page

- Geographic Information Systems and Cartography - *CC BY-NC-SA 4.0*
  - Front Matter - *CC BY-NC-SA 4.0*
    - TitlePage - *CC BY-NC-SA 4.0*
    - InfoPage - *CC BY-NC-SA 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY-NC-SA 4.0*
    - Introduction - *CC BY-NC-SA 4.0*
  - 1: Introduction to Geographic Information Systems - *CC BY-NC-SA 4.0*
    - 1.1: Introduction - *CC BY-NC-SA 4.0*
    - 1.2: Geographic Science - *CC BY-NC-SA 4.0*
    - 1.3: Geospatial Technology - *CC BY-NC-SA 4.0*
    - 1.4: Geographic Concepts - *CC BY-NC-SA 4.0*
    - 1.5: Map Fundamentals - *CC BY-NC-SA 4.0*
    - 1.6: Cartographic Fundamentals - *CC BY-NC-SA 4.0*
    - 1.7: Future of Mapping Technology - *CC BY-NC-SA 4.0*
    - 1.8: Review - *CC BY-NC-SA 4.0*
    - 1.9: Attributions and References - *CC BY-NC-SA 4.0*
  - 2: Spatially Representing Earth - *CC BY-NC-SA 4.0*
    - 2.1: Introduction - *CC BY-NC-SA 4.0*
    - 2.2: Fundamentals of Maps - *CC BY-NC-SA 4.0*
    - 2.3: Datums, Coordinate Systems, and Map Projections - *CC BY-NC-SA 4.0*
    - 2.4: Representing Geographic Features - *CC BY-NC-SA 4.0*
    - 2.5: References - *CC BY-NC-SA 4.0*
  - 3: Spatial and Nonspatial Data - *CC BY-NC-SA 4.0*
    - 3.1: Introduction - *CC BY-NC-SA 4.0*
    - 3.2: Data and Information - *CC BY-NC-SA 4.0*
    - 3.3: Metadata - *CC BY-NC-SA 4.0*
    - 3.4: Researching Data - *CC BY-NC-SA 4.0*
    - 3.5: References - *CC BY-NC-SA 4.0*
  - 4: Geospatial Data Models - *CC BY-NC-SA 4.0*
    - 4.1: Introduction - *CC BY-NC-SA 4.0*
    - 4.2: Raster Data Models - *CC BY-NC-SA 4.0*
    - 4.3: Vector Data Models - *CC BY-NC-SA 4.0*
    - 4.4: Remote Imagery Acquisition - *CC BY-NC-SA 4.0*
    - 4.5: References - *CC BY-NC-SA 4.0*
  - 5: Geospatial Data Management - *CC BY-NC-SA 4.0*
    - 5.1: Introduction - *CC BY-NC-SA 4.0*
    - 5.2: Geographic Data Acquisition - *CC BY-NC-SA 4.0*
    - 5.3: Geospatial Database Management - *CC BY-NC-SA 4.0*
    - 5.4: File Formats - *CC BY-NC-SA 4.0*
    - 5.5: Data Quality - *CC BY-NC-SA 4.0*
    - 5.6: References - *CC BY-NC-SA 4.0*
  - 6: Geospatial Data Characteristics and Visualization - *CC BY-NC-SA 4.0*
    - 6.1: Introduction - *CC BY-NC-SA 4.0*
    - 6.2: Descriptions and Summaries - *CC BY-NC-SA 4.0*
    - 6.3: Searches and Queries - *CC BY-NC-SA 4.0*
    - 6.4: Data Classification - *CC BY-NC-SA 4.0*
    - 6.5: References - *CC BY-NC-SA 4.0*
  - 7: Vector Data Analysis - *CC BY-NC-SA 4.0*
    - 7.1: Introduction - *CC BY-NC-SA 4.0*
    - 7.2: Single Layer Analysis - *CC BY-NC-SA 4.0*
    - 7.3: Multiple Layer Analysis - *CC BY-NC-SA 4.0*
    - 7.4: References - *CC BY-NC-SA 4.0*
  - 8: Raster Data and Imagery Analysis - *CC BY-NC-SA 4.0*
    - 8.1: Introduction - *CC BY-NC-SA 4.0*
    - 8.2: Geoprocessing with Raster Imagery - *CC BY-NC-SA 4.0*
    - 8.3: Scale of Raster Analysis - *CC BY-NC-SA 4.0*
    - 8.4: Spatial Interpolation for Spatial Analysis - *CC BY-NC-SA 4.0*
    - 8.5: Terrain Mapping for Spatial Analysis - *CC BY-NC-SA 4.0*
    - 8.6: References - *CC BY-NC-SA 4.0*
  - 9: Cartographic Principles - *CC BY-NC-SA 4.0*
    - 9.1: Introduction - *CC BY-NC-SA 4.0*