Introduction to Geomatics

Introduction to Geomatics

Scott Bell

Contents

Introduction

Author letter

SCOTT BELL

Dear Student,

This OPEN textbook was developed as a supplement to Geography 222.3 (GEOG 222), Introduction to Geomatics at the University of Saskatchewan. GEOG 222 is a required course for all Geography majors (B.A., B.Sc., B.A.Sc., and Planning), as well as the gateway geomatics course for a Specialization and Certificate in Geomatics. The content of this reader is a mix of original content (95% to 100% of the text and most of the images) created by Professor Scott Bell while other material comes from attributed sources (attribution is included at the beginning of a chapter or section, or for the note taking guide, on each slide or at the culmination of a series of slides).

Lecture notes and slides will also be available via the course BBlearn page; in the future the note-taking guide will also be available with this textbook. In my version of the course for which this book is used, I reserve the right to make minor revisions to my PowerPoint slides but most lectures will be the same as what you downloaded.

I have spent a lot of time thinking about the value of non-OPEN textbooks and reading material for this class. Despite the advanced nature of some of the material, almost all of the facts, concepts, and information are well established. For almost any topic in this course, a rationale/critical/“thorough” web search should provide ample additional material to round out your knowledge. I am under no illusion that my explanation of a topic will be satisfactory for all of you; I have a lecturing and writing style that is my own. In my class I ask students to approach me after lecture (for a brief discussion) or schedule an office visit to discuss any topic or seek my assistance finding supplementary readings.

At the U of S, success in this course is based on attendance and attention during lectures, assigned readings from the textbook, and completion and comprehension of lab work. Geomatics is a subdiscipline of geography that is highly practical; it is also built on a theoretical and scientific foundation with a long intellectual history. The skills you will acquire and tools you will learn to use in any course associated with this text can be applied in almost any field and can enhance inquiry across science, social science, engineering, and the humanities. If you find any errors or have suggestions please do not hesitate to contact me at scott.bell@usask.ca.

I hope you enjoy your course, find the material here useful, and that it improves your lecture experience.

Have a great term,

Scott.

I

Introduction to Geomatics

This section of the book will introduce general themes related to working with geographic information in a digital environment. The theory, concepts, practice, and tools of geographic information systems exist to support our ability to make maps with computers and all the ancillary uses that are related (analysis, communication, storage, management, manipulation, integration, etc.)

1.

Scale in Geography and Cartography

cartographic to problem scale

Scale is an essential geographic concept. Geographers study phenomena at various scales and often use the term *scale* to help define their research interests. Scale is integral to developing a scientific and policy-oriented understanding of our environment. Furthermore, scale can be a useful tool for examining the extent to which a single policy will have its intended impact on the environment. In terms of the scale of a particular environmental issue, an important question is whether the policy is intended to affect a local community or municipality or is it intended to impact a larger jurisdiction like a province, nation, or multinational region. Many policies surrounding environmental issues such as climate change and variability are designed to change the behaviour of local communities in ways that will positively impact a problem that occurs or functions at a much larger scale; in the case of global climate variability, the problem is global and will not be effectively addressed unless collective global action is taken. An example of government attempting to address a large scale environmental issue is the Act passed by the Saskatchewan government in December 2009 titled, *An Act Respecting the Management and Reduction of Green House Gases and Adaptation to Climate Change.* This Act addresses climate change issues that can be addressed by the province, while requiring that climate change and adaptation be addressed in concert with the federal government as well as partners around the world (other countries, multi-state organizations like the UN, G8, G20, and other influential bodies). It has been suggested that global climate variability is effectively mitigated through the harmonization and coordination of all levels of government (DeMarco, Routliffe, and Ladymore, 2004).

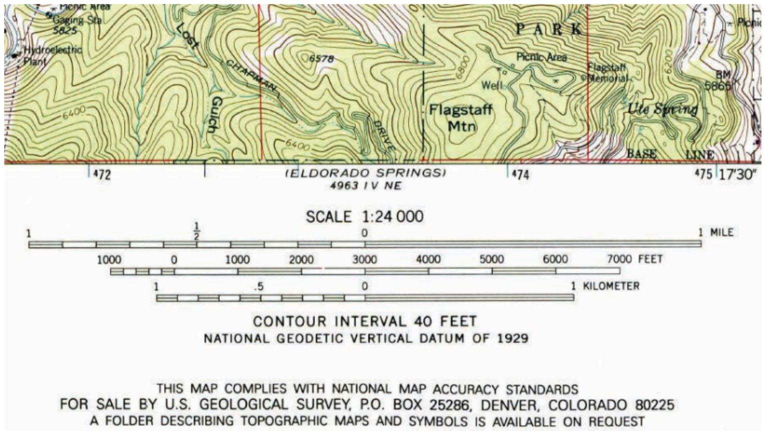
Scale is a concept as relevant to public policy as it is to natural and social science. Finding a good fit for environmental policy, and what a policy's purpose is, requires understanding the intended environmental outcome (global climate variability, water quality, biodiversity, recycling, etc.) and the extent to which a policy is implemented in a political jurisdiction (province, country, etc.) can have the intended impact. Lay definitions of scale are also employed that are related and unrelated to the more strict definitions of the term employed in research-related endeavours. While each of the preceding applications of scale as a term and concept are not patently inappropriate, they do lead to confusion within, between, and among geographers, non-geography researchers, and the public. For example, when governments of different levels (such as Canada's federal government, Saskatchewan's provincial government, or local municipal governments) develop overlapping legislation, each legislative policy has the potential to contradict that from a different jurisdiction. Certainly this is a likely outcome in a province like Saskatchewan that has distinct urban/rural communities that have quite different needs and exhibit different day-to-day behaviours. While the classification of scale based on spatial dimensions, social institutions and practices, phenomena, and mathematics exist and have flourished within and beyond geography, there have been few attempts to categorize the nature of these different definitions and provide a systematic understanding of scale across disciplines. This chapter examines the diverse ways that scale is used within and beyond geography and academia with examples from environmental policy.

When the term scale is employed, it is contingent on the existence of a range of content-specific sub-scales. If something is said to be "scale-dependent" or "scale-specific" we can assume that there are spaces in which that phenomena flourish (in the case of scale specificity). While this is true, these types of statements (scale dependency or specificity) do not suggest anything about the special nature of the scale for the phenomenon or the process. For example, Grizzly bears in the Canadian and US Rocky Mountains require vast tracts of open wilderness for survival. Habitat for the individual bears must be available at a scale appropriate to provide a source of food and security (from other bears and humans). The survival of the Rocky Mountain bear population can be considered to be at least partially scale dependant. Taking a relatively common use of scale in geography, cartographic scale, as an example, one can see that the range of scales that become possible within this single definition of scale.

As suggested above, the term scale can be used in a variety of ways, only some of which are related to spatial extent. While cartographic scale considers spatial extent, its definition is essentially concerned with the direct relationship between what is shown on a map and the physical extent of that space in reality. ***Cartographic* or *representational scale* is based on the mathematical relationship (ratio) between the extent of the representation and that which it represents. The *representative fraction* exists within the continuum of ratios (or fractions) of any two rational numbers**. When comparing the relative size of two cartographic scales we compare the size of the number calculated by the ratio of the size on the map (numerator) and the size on Earth (denominator). For instance, a 1:100,000 cartographic map, would display a distance of 1km on Earth (1000 metres, 100,000 centimetres) with a distance of 1 cm on the map (1 cm on the map reprsents 100,000 cm in reality). Some might suggest that an extension to this broad definition be that the ratio must be less than one, resulting in a representation that is smaller than what it represents. It is for this reason that *representational* and *cartographic* are used above, to allow for representations, or models, that are larger than that which they represent. Therefore, this application of scale results in an infinite number of scales (for example, 1:24,000 or 1:50,000 for standard US and Canadian topographic sheets).

Cartographic scale is convenient for many reasons, not the least of which is that for any two scales we have an immediate understanding of the quantitative spatial relationship between them. As one changes cartographic scale, a direct relationship between the representation and the actual space is explicit and known; space and scale are dependent on one another. For a given cartographic scale, very little can be said about the phenomena being represented, in essence the square, or rectangle (or other shape), that is doing the representing can represent a space on the Earth's surface, or any space, for that matter.

When cartographic scale increases (like from 1:1,000,000 to 1:50,000) the area on Earth that can be displayed decreases, while the detail of the geography being represented can increase (for the area displayed).



An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=35

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=35

Problem scale refers to the **relative *size* of the space** covered by a process or phenomena; large scale problems cover relatively large spaces, and small scale problems cover relatively small spaces. For example, climate change and variability are large scale problems because their impact is global, while developing a curbside recycling program in Saskatoon, Saskatchewan, is a small scale problem because it is an issue which affects the city of Saskatoon alone. These differences are relative because there are no finite boundaries between a large, small, or any other space of a designated size. As introduced earlier, in 2010, provincial climate policy is identified as one of Saskatchewan's most pressing environmental policy areas of the year. However, the specific implementation of a climate change action plan is multi-scaled and does not equitably administer resources. Subsets of the policy apply to homes, individuals, or families, while other climate-focused programs have already been criticized as being out of sync with actually dealing with climate change variability. By focusing on certain alternatives, a provincial policy dedicated to modifying individual behaviour (retrofitting and renovating homes) might benefit some groups more than others (urban vs. rural). On the other hand, recycling programs are considered substantially smaller scale problems that have both impacts and solutions that meet at a common scale. Reducing the amount of open land required for waste can reduce a community's infrastructure costs. Furthermore, recycling has the additional benefit of allowing waste to re-enter the supply stream in the local economy.

Many geographers would suggest this use of the term scale (problem scale) could, and perhaps should, be replaced by the word **size**. As will be made evident below, *problem scale* is likely the more often used concept of scale than *cartographic scale*. It has been used to describe differences in spatial extent as they extend to functional relationships that exist in both physical reality and policy outcomes. *Problem scale*, not unlike *cartographic scale*, takes advantage of common terms for changes in scale; these include, but are not limited to, small, medium, large, etc. While problem scale is also based on a relative relationship between the sizes of the space in which problems exist, the nature of this size (physical extent)—scale relationship is quite different than that between varying *cartographic scale* and the referent space. *Problem scale* is based on a direct relationship (positive correlation) between scale and spatial extent (larger problem, larger space; smaller problem, smaller space). Furthermore, unlike *cartographic scale*, changes in *problem scale*, or comparisons between two scales, are not quantifiable. While a local problem, as defined by problem scale, is obviously a problem occurring in a smaller space than a regional or global problem, it is difficult to calculate a ratio between the two because discrete boundaries are likely unavailable (as they are on a map sheet). In addition, the range of problems that can be described is so broad that for a single environmental issue there might be instances where the problem impacts a relatively small space and other instances where the space is vast. Water quality and watershed health is an example. Contamination of a water supply can happen locally or at a large scale; for instance, if the Mackenzie River is contaminated, the impacts will be felt across much of Canada's north and have the potential to cross international political boundaries; however, if a subsystem of a watershed is contaminated (such as Beaver Creek along the South Saskatchewan River), the impact will be felt over a much more limited and local area.

Functional scale refers to those situations where the general concept of *problem scale* (i.e., large scale refers to large spaces) can be used to describe a relationship between changing scale and function of some phenomena or process. *Problem scale* is relevant when comparing across types of phenomena or processes, *functional scale* should be applied in situations in which spatial extent is used to define a classification scheme for a process or related phenomena. In situations where a process or function varies as spatial extent (or scale) varies, and scale is used to define that variability, *functional scale* is the relevant concept for describing the spatial component of that relationship. While spatial extent may be the determining factor in the variability of some process or phenomena, the nature of the relationship with space is determined by the internal characteristics of the phenomena. It is for these relationships that the hierarchical nature of scale is most relevant. For a given phenomenon or process it is important to define its existence, or how it functions, across different scales, and whether an optimal scale (or range of scales) exists. For example, the hierarchical nature of political systems (municipalities within provinces, or provinces within countries) simply means that decisions and policies implemented at the top will in turn affect smaller and smaller units as that policy "trickles down." For instance, Canada's signing onto the Kyoto Accord results in specific federal policies regarding climate change that govern the development of provincial policies and their implementation at the local level. However, we cannot assume that a provincial policy in Canada will necessarily impact a smaller space than a federal decision elsewhere. These internally structured scaled relationships result in a subset of the functional scale that could be termed ***internal*** *functional scale*.

Psychologists, cognitive scientists, geographers, and many others have examined the cognitive dimensions of space and scale (Mandler 1983; Gärling and Golledge 1987; Montello 1993; Freundschuh and Egenhofer 1997; Egenhofer and Golledge 1998). Each of these classifications has subdivided space based on human interactions (physical, cognitive, and perceptual) with spaces of difference size. Relatively larger spaces require different mechanisms than smaller spaces. A good example would be the search for a hidden object. If that object is hidden on a desk, the individual performing the search would likely move objects around on the (cluttered) desk in order to reveal hidden objects. In a larger space, perhaps a university campus, one would be compelled to walk around the space in order to place him/herself in a position to see hidden objects that are obscured by other (likely immovable) objects such as buildings. So, a shared objective results in quite different behaviours in these two different sized spaces. We can extend this to political or policy environments; in Canada, provinces can only impact behaviours within their jurisdictions and with respect to topics covered by their mandate. Such relationships suggest that in a given political jurisdiction the government will focus only on their mandate; in reality, however, there are many instances of "overreaching" by governments (for instance, the long "states rights" in the United States).

The "size is equivalent to scale" definition is not limited to a cognitive classification of space, but can be extended to many other situations. Within the biological sciences, particularly those branches interested in the complex interactions among species, there are also a variety of classifications of space based on the physical extent (size) of space. Ecology, conservation biology, and environmental management are disciplines in which scale is a prevalent theme, and almost always refers to the physical size of the space and types of interactions that occur in such spaces. Poiani et al. (2000) classify difference sized spaces (scales) using physical extent as well as ecosystem organization. For each scale there are species, terrestrial, and aquatic targets, as well as some general expectations of what one might find at each scale. The smallest geographic scale is local (<2,000 acres), followed by intermediate (1,000–50,000 acres), coarse (20,000–1,000,000 acres), and regional (>1,000,000 acres) (Poiani, Richter et al. 2000). In addition, there are general references to the types of features that might be included in each spatial scale, including biotic and abiotic factors.

The "size is equivalent to scale" relationship works nicely for internal classification schemes, or when a common process or feature can be used to determine how size and phenomena interact. There are many situations in which the relationship between scale and the terms to which it is appended do not have easily described relationships with changes in spatial extent. Examples include the many environmental issues surrounding water and source water protection. Issues and policies surrounding water protection, as stated by organizations such as The Prairie Water Directive, can be trans-provincial and transnational in nature. This results in bodies like the Prairie Water Directive pressing governments to make policies that are appropriate not only for a specific province or municipality, but for multiple provinces and potentially multiple countries, as a lack of environmental policy in one area may, in fact, have substantial impacts on a neighbour (whether those neighbours are communities, provinces, or countries). These situations often arise when the term scale is used to differentiate between place-based differences in geographic phenomena. The spaces characterized by this type of scale generally do not have shared meaning, although geographers, and others, will often use scale to differentiate between different types of places, and the phenomena and processes unique to those places. For instance, certain problems that are endemic of built environments with high population densities are often described as occurring at the urban scale, or being urban scale issues. In order for this definition to exist (*urban scale*), non-urban scales must exist, but these are not necessarily tied to changes in spatial extent or spatial properties of any kind (location, distance, area, etc.). In individual situations the urban scale may extend over larger or smaller spaces than the rural scale. The nature of the difference between the two scales (urban and rural) is based on non-spatial properties, such as demographic character, organizational structure, culture, cultural and social landscape, transportation features, etc. These types of scales are often said to be socially constructed, and while they have important spatial characteristics, they are not defined by their spatial nature. An example of a socially constructed scale would be a distinctive community within a city that is made up of culturally defined groups, and not necessarily related to a specific boundary or physical geographic space within the citiy's limits.

The use of the term **external** is based on the relative importance of space to the variations in scale and on the fact that for each scale employing this definition (rural, urban, homeless, education, etc.) the critical processes or phenomena has an important non-spatial foundation, and the basis of a definition of that phenomena (essential *rural* or *urban* phenomena) is uniquely important to it. What is critical, however, is that within any scale a hierarchy of component scales or entities can be identified and declared (Brenner 2001; Purcell 2003). The urban scale is made up of individuals, households, neighbourhoods, precincts, etc., and the rural scale is made up of individual, households, community groups, etc., that are unique to the respective rural or urban setting and may not necessarily be unique to the physical space or place but can be culturally, socially, and economically defined. Such definitions are critical when considering environmental policy, as the community in question must be part of the decision- and policy-making process. A "one size (scale) fits all" approach will not work.

2.

Representations of Space and Spatial Representations

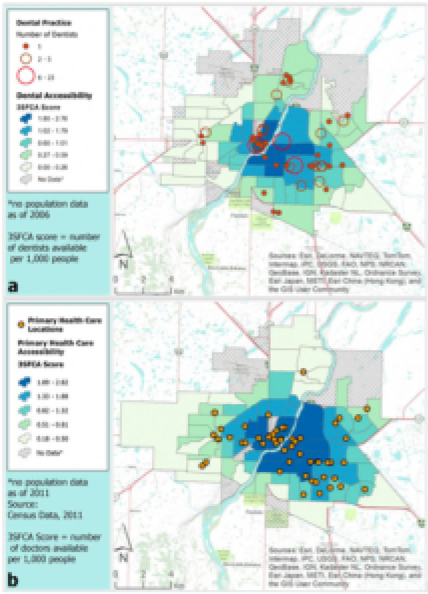
SCOTT BELL

We live on an Earth-sized and Earth-shaped world. Even if we assume away (model) size and shape (flat, spheroid, ellipsoid, or geoid), the phenomena that geographers and others are interested in studying are impossibly complex. In order to explore, examine, model, describe, and understand the world around us, we simplify, generalize, ignore, and represent various aspects of that world. For geographers an important aspect of what we study will inevitably include space. Furthermore, for anyone interested in communicating in a form other than text and numbers, space will be used in a meaningful way to compel our audience to believe our thesis. This presents a conundrum involving what we want to represent (which might be spatial or not) and how we might choose to represent it (spatially or non-spatially).

Geography is a discipline of space and place. Geographers study and use space to explore phenomena. Space influences the outcome of processes and can be used to communicate how a process results in an outcome. With such a tightly bound relationship with space, it is no wonder that representations of space and spatial representations are central to our discipline. While a map might seem the principal or logical representation employed by geographers, we also readily employ diagrams, graphs, charts, cartoons, stories, narratives, models, doodles, and representations of many other types to communicate. Furthermore, the concepts, ideas, theories, etc. that can be communicated might be spatial or non-spatial.

Representation of Space

Reality is composed of spatial and non-spatial elements. We can use many means to define or draw attention to something's spatial character. Take for instance the arrangement of doctors and dentists in Saskatoon; the arrangement is the central element and it is spatial, no matter what units of analysis is used (point locations, aggregations to neighbourhoods, dissemination areas, or tracts) (Aspen, Bell, Shah, & Wilson, 2012; Harrington, Wilson, Bell, Muhajarine, & Ruthart, 2012; Jones, Bell, Hayes, & Uswak, 2015).



To communicate the arrangement, we could write a narrative that describes it, or we can make a map of locations of dentists and doctors. Such a map is a spatial representation of that arrangement, the narrative is not (despite the narratives ability to support the comprehension of the arrangement). We could also take that spatial arrangement and use GIS to calculate the density of doctors or dentists by neighbourhood and rank order the neighbourhoods. While perhaps not as logical (to a geographer), such a ranking would be a more efficient method of identifying the 5 densest neighbourhoods.

On the other hand, there are many non-spatial things (phenomena, process, objects, etc.) that geographers and others are interested in representing or communicating. Some of these might have aspects of space, such as topophilia, but the spatial component might not be of primary importance. Others might have very little or (shudder) no spatial component. Emotions, financial data, feelings, political opinions, attitudes, etc. are examples of non-spatial phenomena. If we take feelings and emotions as an example, there are both spatial and non-spatial representations we can use to summarize the emotion. One could summarize love or hate with words, as text or dialogue, BUT, we might also write a song or poem, or one could produce a painting. The latter would employ space (along with colour, texture, orientation, etc.) to communicate a thing that is decidedly non-spatial.

Spatial Representations

Representations are used to summarize reality so that we can focus more carefully (critically) on a specific topic of interest. Such summaries can come in many forms, some of which have been introduced above, but include many others. Novels, photographs, movies, models, equations, doodles, lists, and graphs are all examples of methods of summarization or representation. Not only are the forms of representation seemingly unlimited, but that which they can be used to represent are also myriad. A hand drawn map of the countries of the world, represents not the world but the "artist's" rendering of what they know (Bell & Archibald, 2011; Golledge, Bell,

&  Dougherty, 1994). A story describing the harvest season on a farm would necessarily have to describe several spatial components of the process of bringing a crop in from the field, but the representation used would be text, not something that is itself spatial and would therefore not be a spatial representation (beyond the order of letters to form words, sentences, and paragraphs). A painting of the farm would freeze in time the spatial arrangement of farmhouse, barn, and yard, requiring the careful use of the spatial extent of the canvas and hence be a spatial representations.

Representations of Space

Spatial Representation	Representations of Space: **Yes** (Represents spatial properties of reality)	Representations of Space: **No** (Does not represent spatial properties of reality)
Yes (Uses space in the representation)	Map Painting of a farm Topological diagram of a subway system	Population Pyramid Chart showing stock market value over time
No (The representation does NOT rely on space)	Speech on the spread of Ebola in 2014 Story about a trip across Canada	Poem about love Song

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=87

II

Maps and Mapmaking

3.

Introduction: What is Geomatics?

Geomatics is a subset of study in the discipline of Geography. Geography is the study of the arrangement, interaction, and change of both physical/natural features and human activity on and near Earth's surface. Human activity affects, and is affected by, the environment; geography is interested in these interactions, but also individual features or distributions of features. Geomatics involves a wide range of methods and technologies for collecting, managing, and analyzing data about Earth and the phenomena arranged on and near its surface. An important component of Geomatics is **Geographic Information Systems (GIS)**; GIS uses spatial data to explore geographic phenomena. GIS includes communicating and visualizing spatial data; these aspects of GIS are commonly (and historically) associated with the field of **cartography** (the art and science of map making).

This text will cover many topics related to Geomatics, GIS, and Cartography. In addition to Geomatics, GIS, and Cartography, some of the topics that will be covered include:

Geodesy: The study of Earth's size and shape, one goal of which is the definition of geodetic datums (models of Earth size and shape).

Remote Sensing & Aerial Photography: Making physical observations without direct contact or touch; aerial photography is limited to those observations made with photographic technology, while remote sensing is often associated with digital images, although the two share many common concepts and methods.

Global Position System (GPS): A system composed of earth orbiting satellites that produce accurate time signals that are used to determine precise and accurate location on Earth's surface using a mathematical technique called trilateration (not to be confused with triangulation). see https://gisgeography.com/trilateration-triangulation-gps/

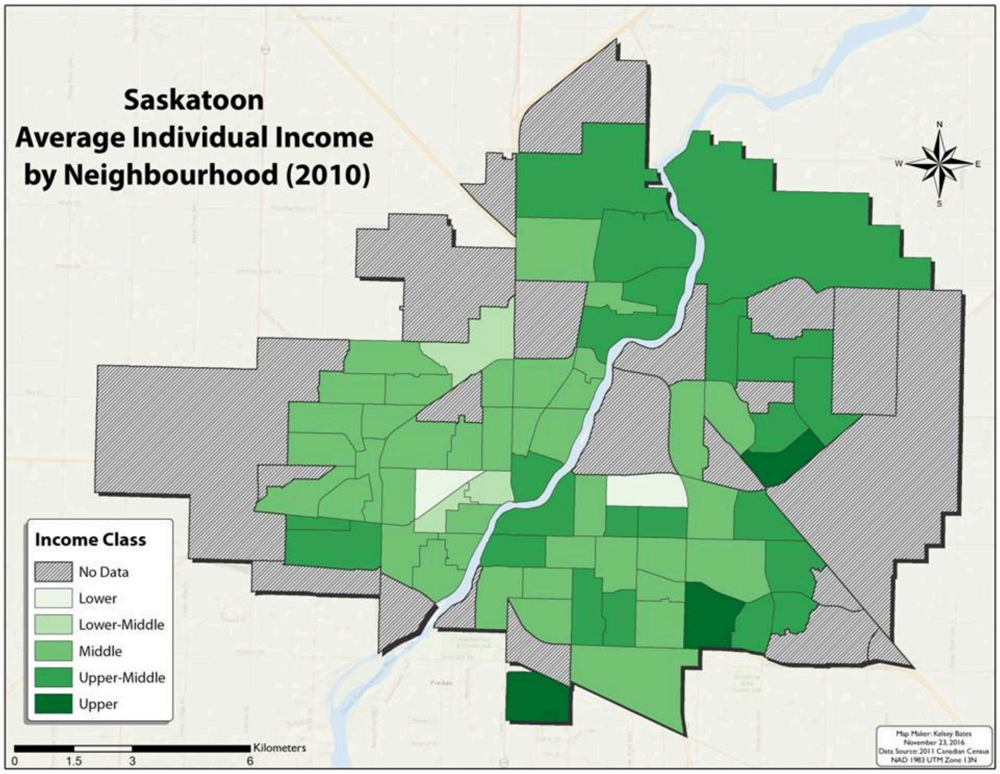
Geographic Information Science: The study and extension of GIS capabilities. While this topic emerged in the 1990s, GIS could not exist without the various advances in what we know about Earth as an object of study, how to systematically simplify it, and how we might examine the patterns that emerge from presenting reality using models (simplified representations).

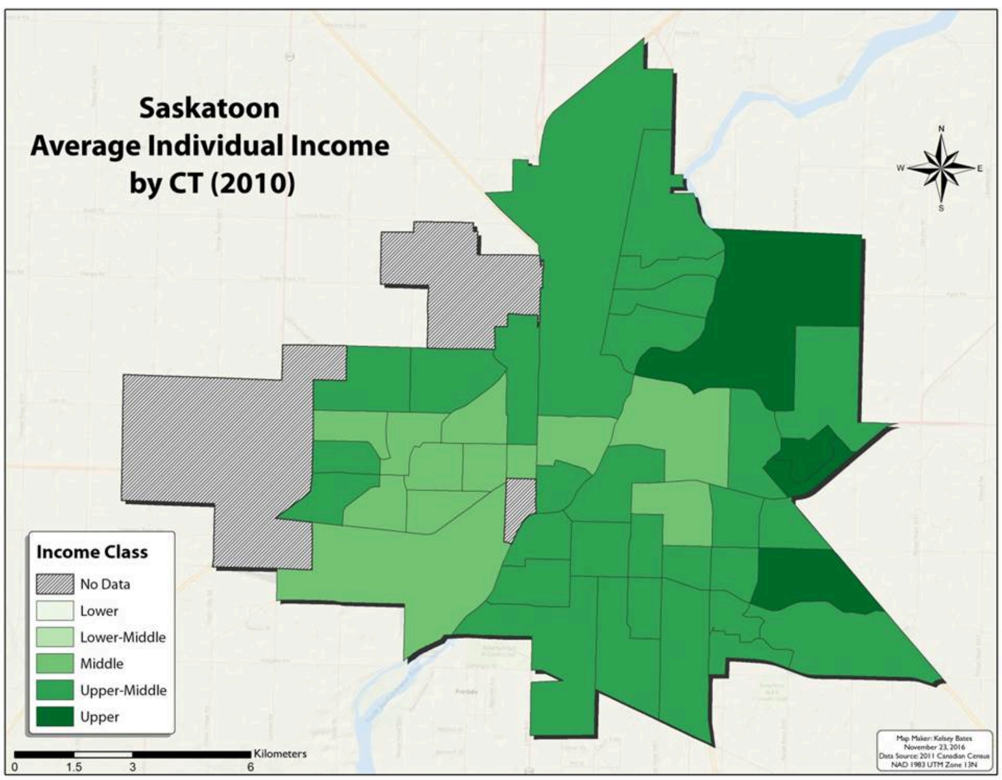
An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=22

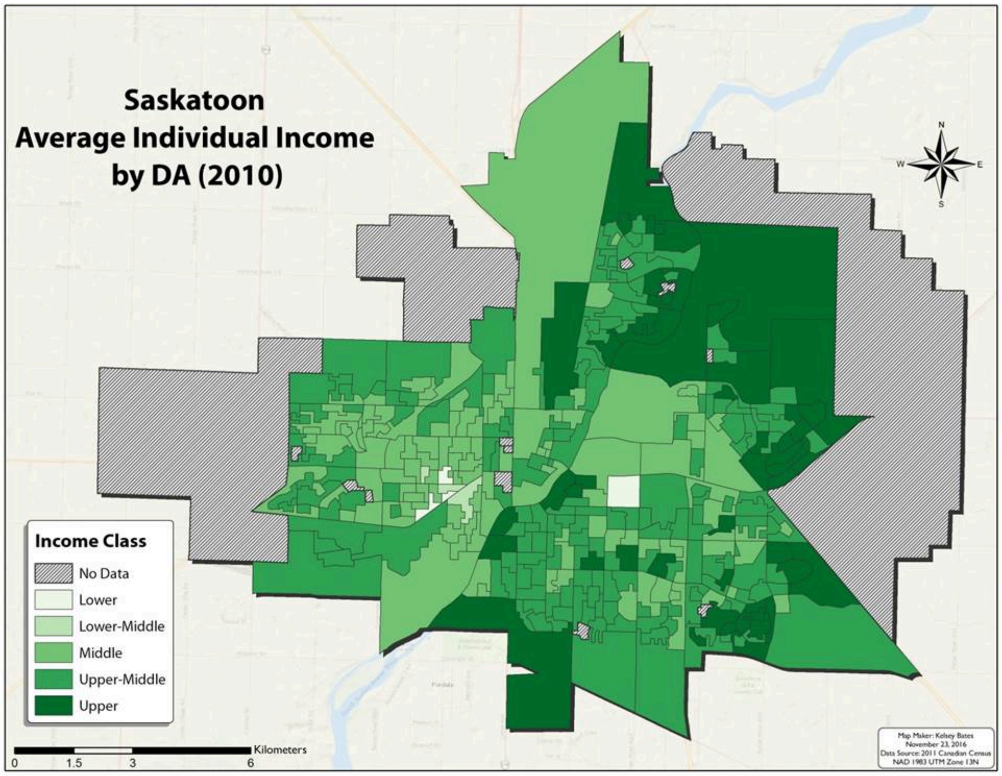
First "Law" of Geography

In 1970, the geographer and cartographer Waldo Tobler coined the 1st law of geography (Tobler, 1970); the "1st Law" states: **Everything is related to everything, but near things are more related**. In many ways this pithy statement summarizes why GIS, geomatics, and mapping are possible. In geography, because the world is so large and the physical and social landscapes are so complicated, it is necessary to model and represent what we know. One of the first steps toward mapping is making models of Earth that simplify its size and shape. When we want to map phenomena that are arranged across Earth's surface, we need to generalize and combine individual observations. Given the fact that the physical landscape varies continuously, and not abruptly, we can make maps that simplify the complexity of the real world. Similarly, when summarizing social phenomena on maps, such as income, voting, residential style, race, ethnicity, etc. we often give areas occupied by many people a single value.

For instance, in the three figures below, income in Saskatoon, SK has been displayed by taking the average income of three different types of areas (neighbourhoods, census tracts, and census dissemination areas). Each "area," or **unit of analysis,** has a single value that value is used to determine what shade of green it is assigned. The value might be the average, total, median, or mode; independent of how we numerically (statistically) summarize the many individual values, the fact that we can simplify reality in such a way. The resulting maps display the geographic trend of income across Saskatoon (the **area of study).** This type of map is called a **choropleth map** (and will be covered in the third part of the Maps and Mapmaking section). By comparing the three maps you should be able to see that the general trend of income is similar, but that each is different in subtle ways. The advantage of generalizing income by taking an average is that the map reader can see the geographic pattern; the disadvantage is that we can't determine the income of any single resident or the specific average income for a neighbourhood. However, if we did place the incomes of each household on the map, the pattern would be lost.


Saskatoon
Average Individual Income
by Neighbourhood (2010)
N
S
W
E
Income Class
No Data
Lower
Lower-Middle
Middle
Upper-Middle
Upper
0
1.5
3
6
Kilometers
Map Maker: Kelsey Bates
November 23, 2016
Data Source: 2011 Canadian Census
NAD 1983 UTM Zone 13N


Saskatoon
Average Individual Income
by CT (2010)
N
S
W
E
Income Class
No Data
Lower
Lower-Middle
Middle
Upper-Middle
Upper
0
1.5
3
6
Kilometers
Map Maker: Kelsey Bates
November 23, 2016
Data Source: 2011 Canadian Census
NAD 1983 UTM Zone 13N



That we accept such maps as reasonable representations of the pattern of income variation in Saskatoon is an example of what can be achieved as a result of the 1st "law" of geography. A final comment of the 1st "law" is my own choice to put the word law in quotation marks. This suggests that the word law should not be take literally. In this case, it is important to accept that the principle of the 1st "law" of geography holds for many phenomena and under many conditions, it is not a physical or immutable law. Even for a pattern like the one above, there are low income residents in high income neighbourhoods and vice versa. Furthermore, we should not assume that everyone who lives in a high income neighbourhood is a high income individual or has an income at or near the value assigned to that neighbourhood (through mathematical or statistical calculation). To draw a conclusion about an individual based on the group to which they belong (in this case an individual residing in a high income neighbourhood) is an example of the **ecological fallacy**, and is a common risk when making thematic maps, such as these choropleth maps.

Geographic Information Systems exist as a tool for geographic exploration and investigation. GIS support a digital representation of geography and as such rely on aspects of data, databases, computation, and abstraction. Because their essence is digital, GIS representations have difficulty portraying aspects of reality that cannot be measured or assigned a value. This includes emotions, feelings, or other qualitatively assessed phenomena. In portraying geography digitally, GIS relies on storing information as data. GIS data generally falls first into two categories: **Spatial** and **Non-Spatial data**. Spatial data provides the spatial structure to which non-spatial data is assigned. The spatial structure must accommodate both geography (political units, physical features, locations, roads, cities, etc.) and cartography (be something that can be "drawn," such as points, lines, and areas). If a line represents a road, and a string of **Latitude** and **Longitude** coordinates define the line's position and extent, then the combination of spatial data (string of coordinates) and non-spatial data (categorized as road) makes it geographic information.

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=22

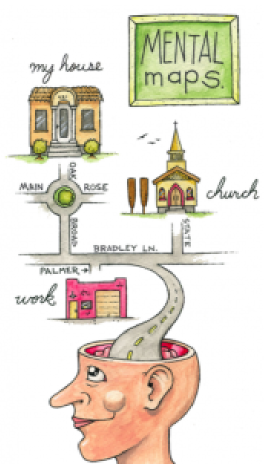
Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(2): 234-240.

4.

Mapping: A Geographic Tradition

(The first two paragraphs of this section are from Campbell and Shin, Essentials of GIS)

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 the imagination of people around the world. As one of the most trusted forms of information, map makers and geographic information system (GIS) practitioners hold a considerable amount of power and influence (Wood 1992; Monmonier 1996).[1, 2] Therefore, understanding and appreciating maps and how maps convey information are important aspects of GISs. The appreciation of maps begins with exploring various map types.



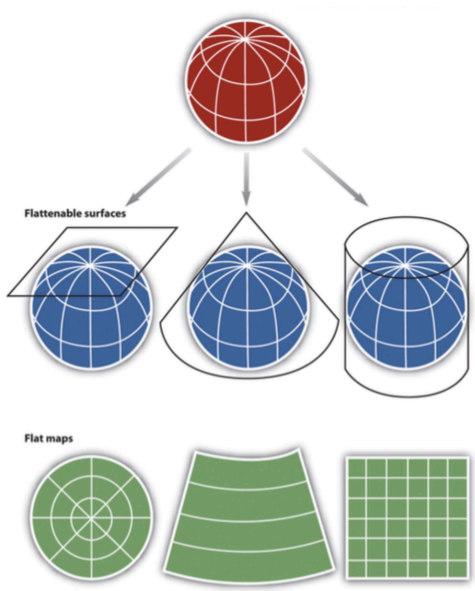
So what exactly is a map? Like GISs, there are probably just as many definitions of maps as there are people who use and make them (see Muehrcke and Muehrcke 1998).[3] For starters, we can define a map simply as a representation of the world. Such maps can be stored in our brain (i.e., mental maps), they can be printed on paper, or they can appear online.

Notwithstanding the actual medium of the map (e.g., our fleeting thoughts, paper, or digital display), maps

1. Wood, D. 1992. *The Power of Maps*. New York: Guilford.
2. Monmonier, M. 1996. *How to Lie with Maps*. Chicago: University of Chicago Press.
3. Muehrcke, P., and J. Muehrcke. 1998. *Map Use*. Madison, WI: JP Publications.

represent and describe various aspects of the world. For purposes of clarity, the three types of maps are the reference map, the thematic map, and the dynamic map.

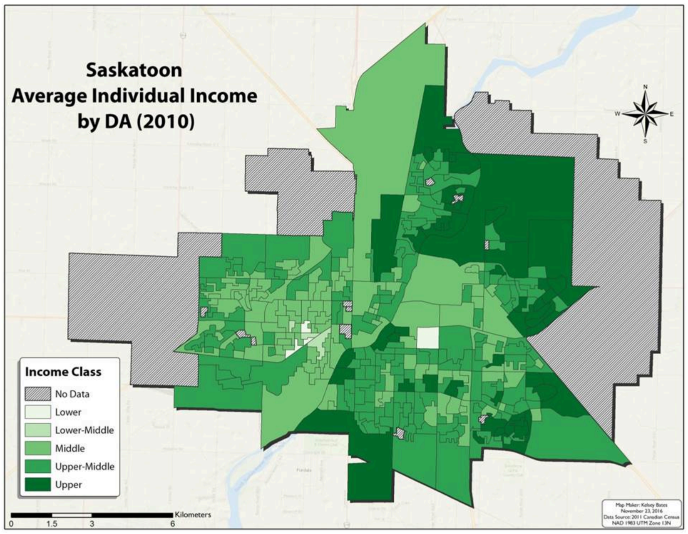
Mapping is the process of making images that are accurate representations of Earth and phenomena on and near its surface, but generalized to two dimensions (flat). Accurate mapping requires precise measurements of Earth's size and shape. From this model of Earth's size and shape (called a **geodetic datum**) we can establish a frame of reference onto which themes, topics, and features can be displayed. Before a map is created from a geodetic datum, two steps are necessary: 1. a cartographic scale is applied; and 2. the surface of the geodetic datum (a sphere or ellipsoid, both three dimensional shapes) must be projected onto a flat surface. The latter step is called **Map Projection**. The term *map projection* can be used as both a noun and a verb. As a verb, it means the process of rendering the surface of a geodetic datum (of Earth) onto a flat surface; as a noun, it means the specific parameters used to establish a single representation of Earth. In GIS software, all data is recorded, stored, and displayed according to a map projection and this information is easily retrievable through the software.



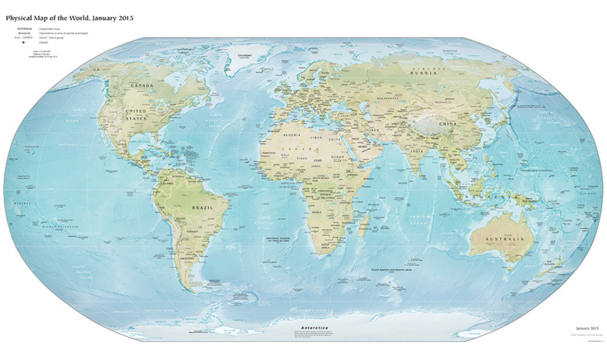
Once projected, a map can be used to display any spatial extent and myriad thematic topics. The spatial extent ranges from world maps that display all of Earth, to maps of large cartographic scale that portray relatively small areas. A cornerstone of maps is that the relative and absolute arrangement of features on the map are accurate to the arrangement seen in reality. If one place is west of another, it will appear this way on a map. This accurate portrayal of spatial relationships can be deployed in other spatial representations, even those not portraying Earth or portions of it. In fact, traditional cartographic scale can be inverted to portray spatial relationships for spaces that are not visible without magnification (such as diagrams of cellular structure or atomic particles).

Once a map projection is selected and applied, and a spatial extent established, a map's content can be portrayed. Maps can be created to achieve many goals. They can be used to communicate a thematic pattern (income variation in Saskatoon, SK), to reveal a pattern that represents an underlying relationship (water contamination that causes a disease outbreak), or simply the accurate portrayal of positions across the map without a specific theme. These three map applications represent three categories of maps, each of which provides guidance for the cartographer (map maker). These map categories are commonly termed map traditions:

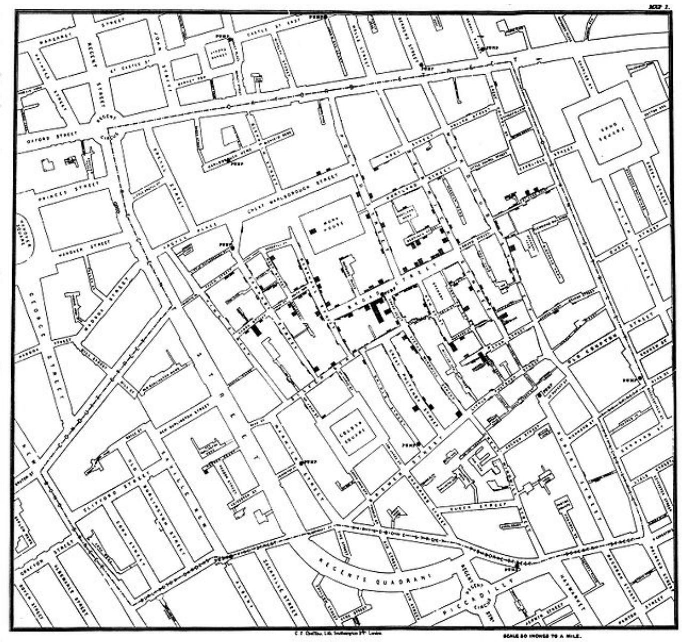
- Communication Tradition: Thematic maps are designed to communicate a single theme, or narrow set of themes, and are intended to help a map reader (user) understand an underlying topic, usually associated with answering a specific question.



- Reference Tradition: Reference maps are designed to portray the accurate spatial location and arrangement of a range of generic features (towns, rivers, boundaries, etc.). Unlike thematic maps there is no single theme on reference maps.



- Analysis Tradition: Cartographic analysis is the process of identifying spatial patterns within and between variables as displayed on a map. The analysis tradition is associated with GIS, spatial statistics, geostatistics, and other analytic methods that exist to quantify spatial arrangements.



In the chapters that follow in this section, map types will be elucidated, along with the principles of map design. Map projections and Coordinate Systems will be examined in a later chapter.

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=31

5.

Maps as Communication

Cartography as Communication

Cartography is a form of communication. The communication is largely visual and relies on various, but stringent, application of rules and guidelines. While language is a less visual form of communication, it is useful to use the concepts of grammar and syntax as they relate to cartography. Unlike language, maps and cartography involve iterative and complex decisions about what to include on the map and how is should be included. Using a variety of tools, such as colour, shape, and pattern in combination with the unalterable spatial relationships of the geography being mapped, the cartographer can compose a meaningful and communicative representation. Additionally, by treating cartography as communication it allows us (students, cartographers, researchers, etc.) to tap into the rich knowledge associated with the semantics of communication.

Grammar: set of rules governing sentence structure

Syntax: how sentences are laid out in order to create grammatical sentences

Semiotics: theory that sumarizes the relationship between reality, reference and those things that represent reality (signs)

Semantics: relationship between a sign (or symbol) and that which it signifies

Pragmatics: practical application of a sign

Unlike many other forms of visual and graphical communication, cartographers are limited by the need to accurately portray the geography of the part of the world being mapped. Furthermore, they need to use established coordinate systems, map projections, and restrict their map to a scale that is reasonable for the spatial extent being mapped and the format of the media being used to display the map. Maps can be displayed in a variety of sizes, from a section of a relatively small-paged book to a wall mounted map several feet in dimension. When combined with the geographic extent of the area being mapped (a city, state, province, country, etc.) and that area's shape, the scale of the final product might be constrained, leaving the cartographer little flexibility.

Maps are symbols, maps are composed of symbols, therefore maps represent a form of symbolic communication.

Furthermore, maps are simplifications of reality. The nature of this simplification is multifaceted and complex. A single map employs a cartographic scale to establish the quantitative relationship between the representation and what is being represented. Once a scale is put in place, the cartographer can set about making decisions about what to include on the map and how to include it, keeping in mind that a theme or topic and spatial extent (area of study) has been set. The unit of analysis might not yet be established, but based on the area of study and theme, it has likely been constrained.

When creating maps we must simplify geography, both its spatial and non-spatial aspects. In chapter 1 (Intro to Geomatics) the concept of combining spatial and non-spatial data to create geographic information was introduced. The non-spatial aspects are generally simplified into variables for which data is recorded. Each variable contains values that are a summary for a spatial entity (like an enumeration area, province, or neighbourhood, but the spatial entity doesn't have to be an area). As a result, a social phenomenon (affluence, poverty, health, etc.) is summarized in a way that can be assessed and then assigned a value. That value is considered data. The value doesn't have to be numeric, but it can be. The amount and type of information that is present in a piece of data is referred to as its **Level of Information**, there are four levels of information: Nominal, Ordinal, Interval, and Ratio, these are summarized below. The spatial/geographic components of what is mapped is also simplified. We've already seen (generally) how cartographic scale, geodetic datums, and map projections simplify reality, but cartography also simplifies the geometry of geography. Features that are "drawn" on a map are drawn as points, lines, or areas (even if they are not recorded as such in a database). By associating the spatial and non-spatial we have enough information to map geography.

Levels of Information

There are four levels of information, each level contains a specific amount of information.

Nominal Information: Data is put into classes with distinct labels or names (nom); classes have no relationship to one another and are given no relative value (eye colour is an example). Nominal data is categorical and is not numeric. A common dilemma with nominal data occurs when the categories are recorded as numbers, it is important to distinguish between numbers that represent nominal categories and cannot (should not) be treated as numbers (added, subtracted, considered larger or smaller based on their numeric value).

Ordinal Information: Data can be placed in ranked categories; categories are differentiated by position in the ranked scale; values can be greater or less than one another, but the magnitude of the difference between neighbouring values on the ranked scale cannot be determined (category 2 is not double or half of category 1). Differences cannot be measured; for height, all we can know is that A is taller than B, B is taller than C, and subsequently A is taller than C).

Interval Information: Data is given a value that is based on a fixed scale with no true zero value (arbitrary zero). Interval data may appear as though it can have complex operations performed on it, such as performing a ratio, division, multiplying, etc., but this is not true (i.e. 10˚C is not twice as warm as 5˚C).

Ratio Information: Data exists on a scale with a constant interval between values and has an absolute 0. Ratio data can be multiplied by a constant, have a ratio applied to observations, and support complex analytical operations.

Levels of information are essential to many aspects of cartography and GIS. This importance is closely associated with how information is used to represent reality, or other information. In cartography, the map is a representation of reality that accurately depicts the geographic and non-geographic aspects of reality (those aspects important to the map's objective).

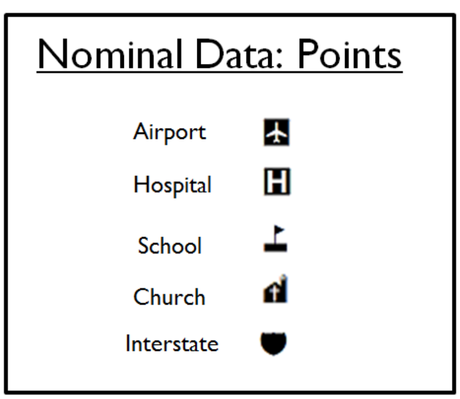
So, from the very first steps in the mapping process the map maker is simplifying reality into information. The amount of information in that initial representation will affect the map making decisions that follow. Once something has been summarized as data, with a certain level of information, that is all the information we have. There is no going back to increase the amount of information available (without recollecting the data). Therefore, we have to be considerate of the amount of information available for mapping.

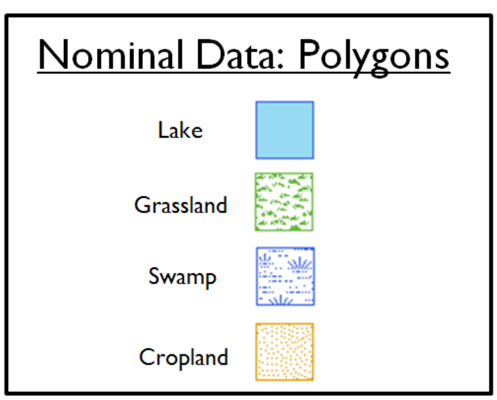
An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=29

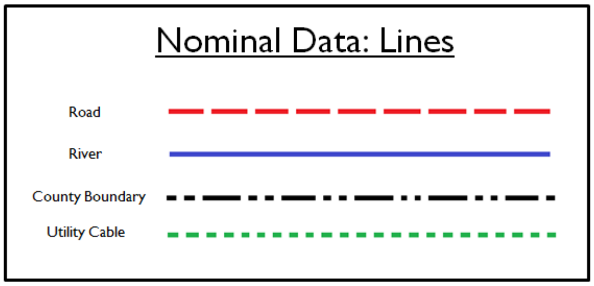
Symbology of Levels of Information

On the map, the symbols we use to display information also have a capacity for information. These are referred to as symbol variables. A symbol can have different colours applied to it (nominal) or be presented in different sizes (ratio, although size can be used to communicate ordinal, interval, or ratio information). A good guideline is to match the level of information in the map symbol to the level of information in the variable being represented. So, if the information is categorical (ethnicity, political party, etc.) then the map symbol variable should also be categorical (colour, shape, etc.). This matches the nominal level of information in the variable being mapped with the level of information in the map symbol.

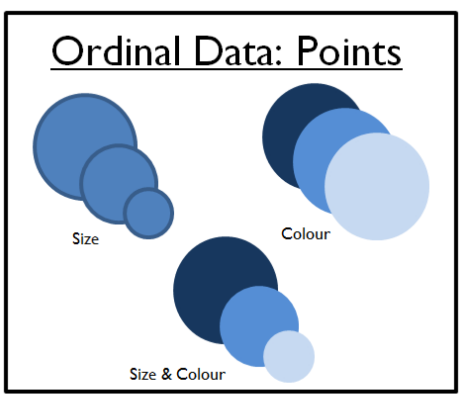
Nominal:

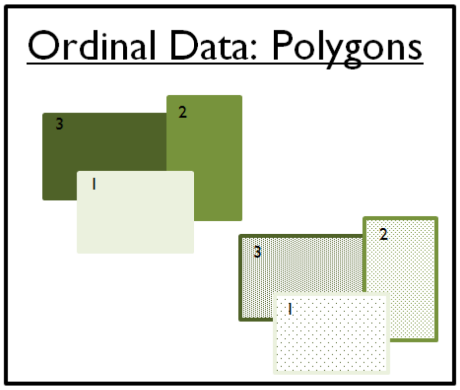


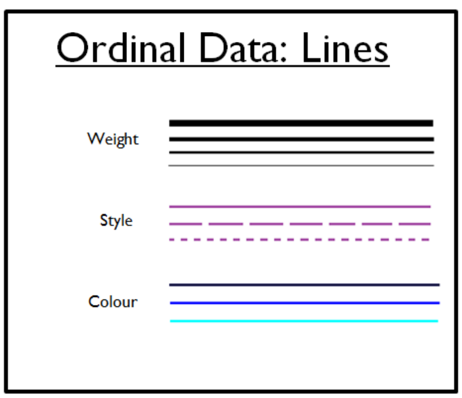



Nominal Data: Lines
Road
River
County Boundary
Utility Cable

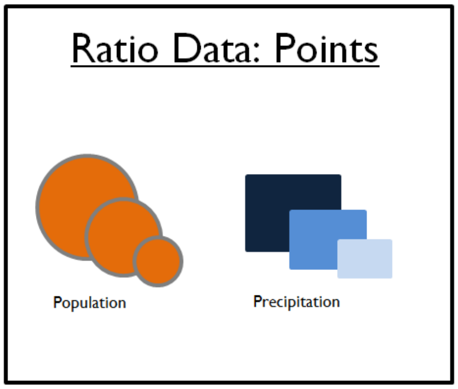
Ordinal:

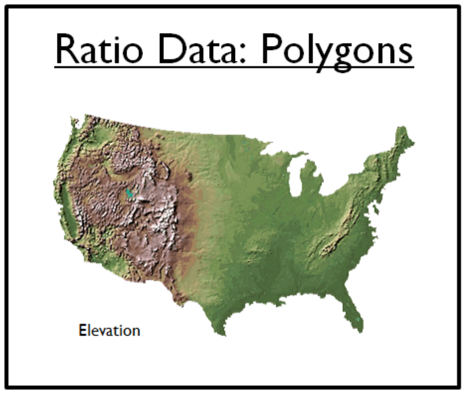

Ordinal Data: Points
Size
Colour
Size & Colour

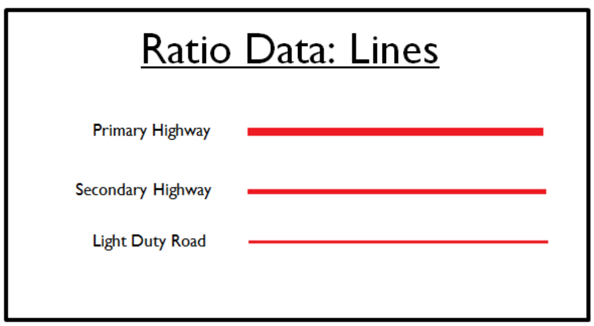

Ordinal Data: Polygons
3
2
1
3
2
1


Ordinal Data: Lines
Weight
Style
Colour

Ratio:



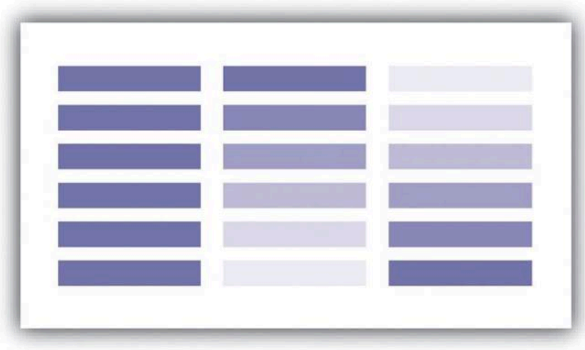




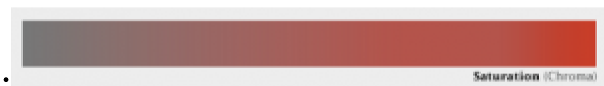
An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=29

In a later chapter we will look at the special properties of colour and how the three elements of colour can be used to convey information. As a teaser, the three elements of colour are: **Hue**, what we would describe as *different* colours (purple, brown, magenta, etc.); **Value**, the amount of grey added to a colour (more grey is LOWER value); and **Saturation**, for a given value, this is the colour’s brightness.

Hue: 



Value:

Saturation: 

<u>Summary of the Levels of Information</u>

The amount of information stored in the data.
Nominal: categorical data (colour, group, nationality, etc.) i.e. bib number in a race
Ordinal: relative value (more, less, greater, smaller) i.e. finishing place in a race
Interval: quantitative difference between data points (no absolute zero, difference can be calculated but ratio cannot) i.e. difference between first finisher and yourself; Celsius and Fahrenheit
Ratio: ratios can be calculated, i.e. ratio of your time to finish race with that of your competitors; Kelvin

6.

Thematic Maps

Thematic maps are intended to communicate a single theme or narrow set of themes. It is important to remember that thematic maps are different from reference maps. The primary purpose of a reference map is to deliver location information to the map user. 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 precedent over any other aspect. Moreover, reference maps generally represent geographic reality accurately. Examples of some common types of reference maps include topographic maps such as those created by the United States Geological Survey and Canada National Topographic Series (NTS) and image maps obtained from satellites or aircraft that are available through online mapping services.

Thematic Map Types

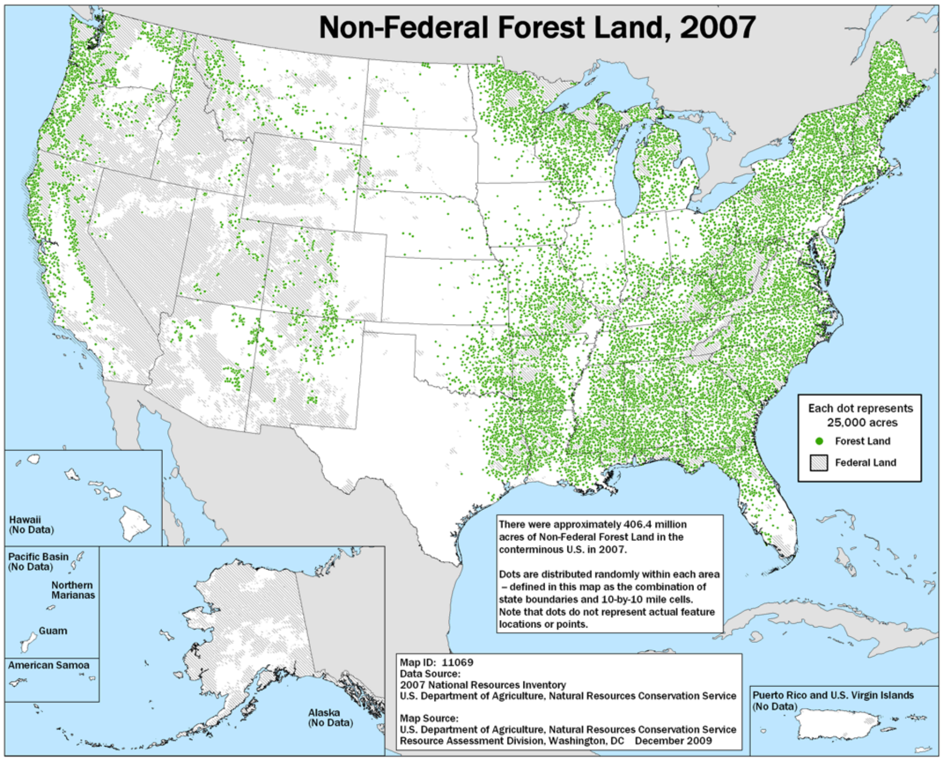
Dot Density Maps

A **dot density map** is a map in which small symbols of uniform size are used to emphasize the spatial pattern of a phenomena. Each dot is equivalent to the same quantity. Dot density maps can be made multivariate (display data associated with more than one variable but using colour; different colours correspond to categories associated with a second variable, hence “multivariate”).

Map symbol is the dot.

Dots are placed randomly within an enumeration zone and are only used to indicate that a specified value of some variable exists in the general area. This type of map is dangerous as it may lead the map interpreter (reader) to believe that an actual occurrence of some variables exists at the point in space occupied by the dot.

Compare the distribution of dots on the accompanying figures.

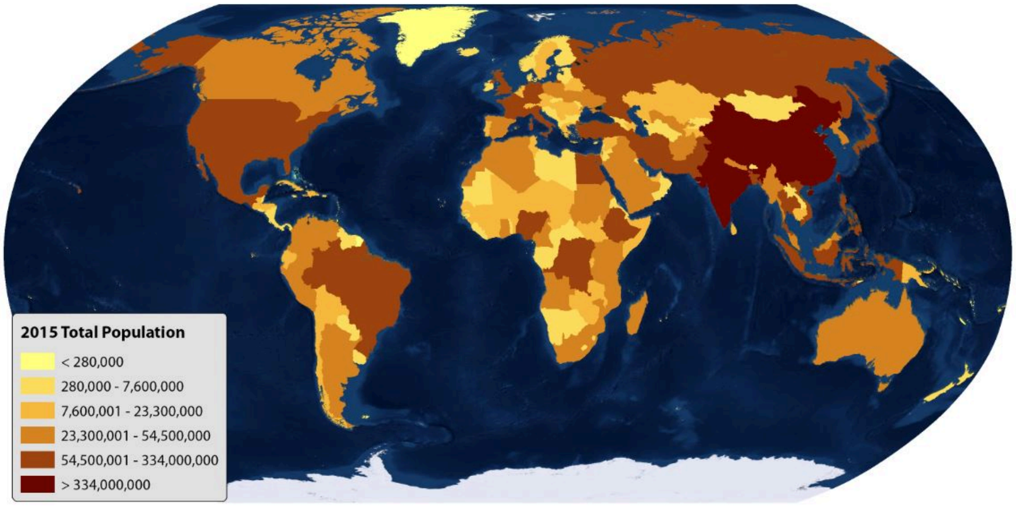


<u>Choropleth Maps</u>

A **choropleth map** is a map in which **enumeration units** (or data collection units) are shaded with an intensity proportional to the data values associated with those units. For instance, darker colours on the map might mean a higher value for the variable be represented. It is important that a choropleth map have clear titles and legends, so the viewer can understand what is being communicated.

Map symbol is the enumeration area.

In this usage, the term enumeration area is used generically, despite the use of the term in the past by Statistics Canada to refer to what is now called a Dissemination Area. Enumeration area is also, and has been, used by other countries to refer to specifically defined areas. For choropleth mapping enumeration areas are the areas with boundaries for which a single value for a mapped variable is available.

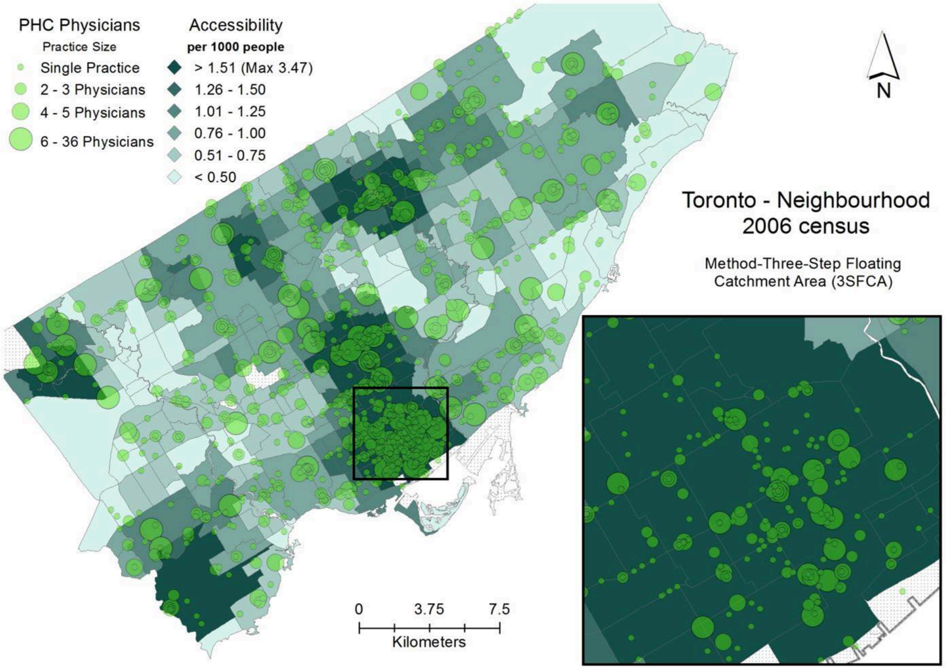


Proportional Symbol Map

Proportional symbol maps have points that are scaled in proportion to the magnitude of data occurring at point locations, such as using circles of varying sizes to represent urban population.

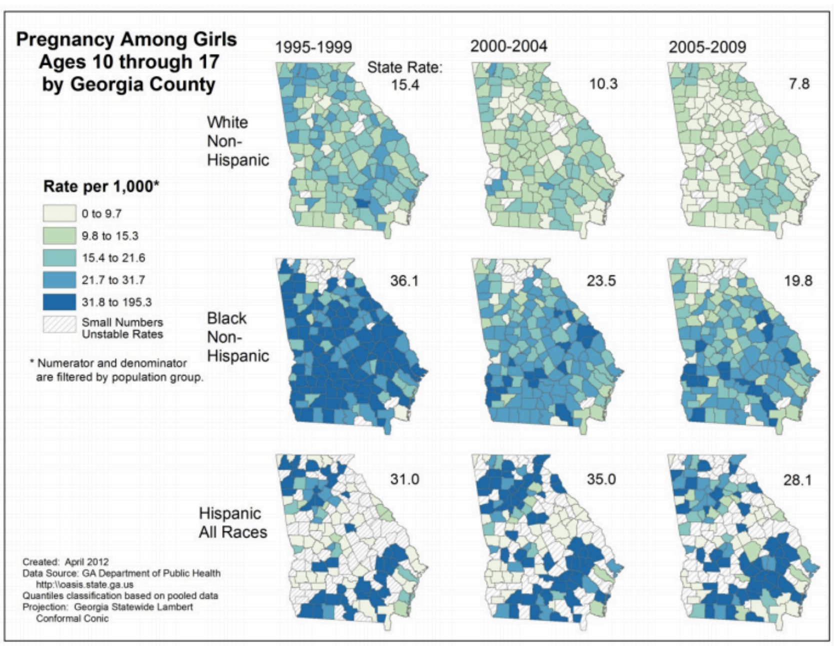
Symbol varies in size and can be of any type the map maker wishes, often a circle.

The size of the symbol (most often it is a circle, but it can be almost anything) is an important consideration for the cartographer. Some mapping packages will use the area as the proportional variable, but some use other values, such as the diameter of the circle (hmmm, how might that be turned into an exam question?). In the most common GIS software there is often a distinction for this kind of map between a *continuous* scaling of symbol size in accordance with the value for the variable being mapped and a *categorized* scaling of the symbol. When scaling is continuous there are as many symbol sizes as there are unique values for the variable being mapped; when scaling is categorical, there are as many symbol sizes as their are categories (as with choropleth maps).



Small Multiples

When a theme or map objective requires the presentation of multiple variables and there are too many to fit on a single map, small multiples can be employed. Small multiples are repeated versions of the same geographic extent the only variation between each map being the variable mapped.



An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=33

7.

Cartographic Design

Maps are designed to communicate themes, reveal patterns, and store information. Achieving these goals relies on design principles that are both general and specific. Specific design decisions include composing a title, choosing a projection, map scale, spatial extent, classification scheme, and sharing references to data and sources. General design decisions include selecting colour schemes and typefaces for text and laying out the geographic space to be mapped and non-map elements (title, legend, north arrow, scale, etc.).

Map Elements

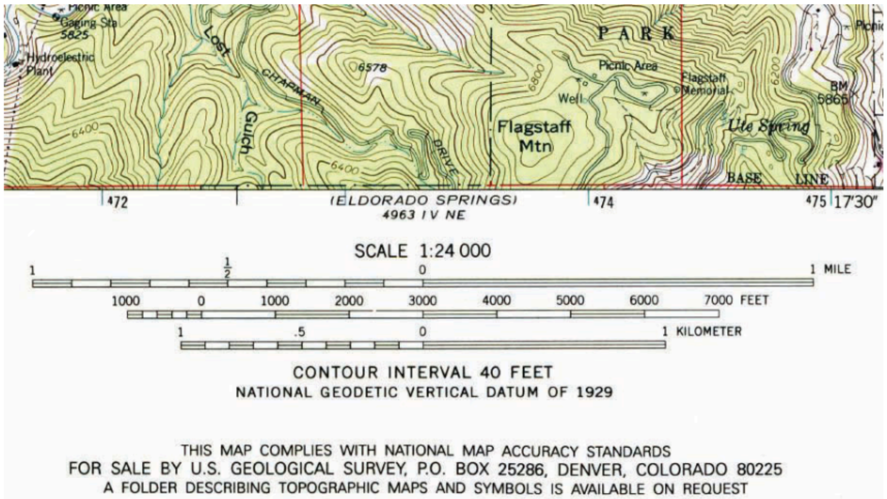
The following map elements are LOOSELY ordered based on the extent to which they are required, thematically essential, or optional. In selecting and applying each element, as with all decisions regarding map layout and design, decisions should made based on their service to the map's objective.

Title: maps need titles, even a reference map. The map should pithily convey the maps theme and not overlap in content with the title of the legend. If the title "needs to be long," this is an opportunity to consider how the title can be split between the Title and the Legend Title.

Legend: communicating the symbology of the map is accomplished via the legend. Legend content should be determined by the map, not the constraints on the legend's position, size, orientation, etc. As a rule of thumb, the title of the legend should focus on the content of the legend as it supports the map's theme. The worst title for a legend is "legend."

Neatline: the line surrounding the geographic extent of the geographic space being mapped. Nothing geographic falls outside the neatline, but other elements can be placed within, outside, or straddling the neatline (legend, scale, title, north arrow, etc.)

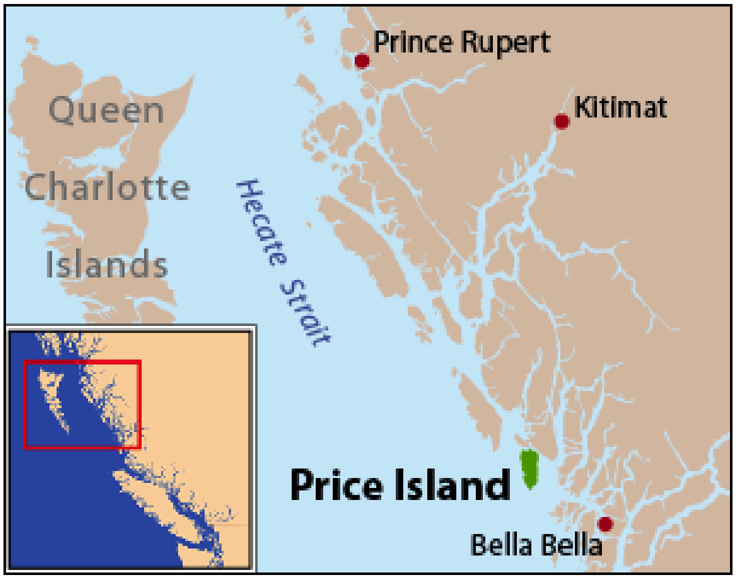
Scale: expressing the spatial relationship between the mapped space and that space's extent in reality can be done in several ways. Like Orientation below, at very small scales, it can be hard to decide how to express scale (since the scale is often different for different parts of small scale maps). Scale can be expressed numerically as a ratio (1:50,000), "verbally" (1cm = 10km), or with a scale bar. Scale bars are elegant in their ability to communicate the scale of the map under conditions of zooming (if the final map is printed and expanded the two maps have different scales).



Orientation (North Arrow): not all maps need to have their orientation specified. Generally, as map scale decreases the need to communicate orientation (where is north) decreases. When the geography being portrayed communicates orientation, a north arrow, or other orientation information, is less necessary. Continental and world maps are good examples of maps that don't need orientation information, in fact for small scale maps north arrows can be confusing. Place of a single north arrow might cause a map reader to wonder, "where is north from any point on the map and is a north arrow specifying the direction to north for its location on the map, for the map's centre, or for some other location?" Orientation can be communicated by north arrow or by the inclusion of coordinate system information (lines of latitude and longitude, for instance).

Labelling: applying text to a map to indicate the names of features is useful for providing thematic context. A section on typography later in this chapter will cover labeling in more detail.

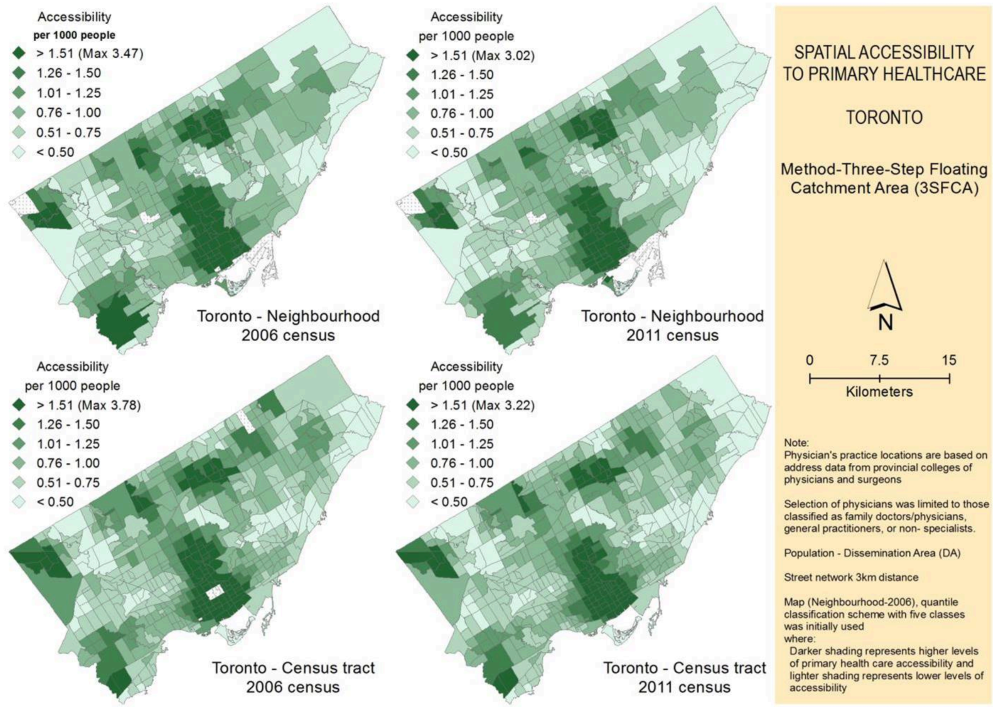
Locator Map: map of smaller dimensions AND smaller scale that indicates the geographic context of the main geographic extent of the map.



Callouts: using lines or other means to connect words or symbols with their associated place on the map. Callouts are especially useful for areas of a map that are dense with information.

Source material: basemap, spatial, and non-spatial information should be attributed to their sources. Generally this information is ancillary and should fall outside the neatline (at the very least) and not attract too much attention. Like references in written work, information about source material should be accessible but not distracting.

Text blocks (explanatory narratives): when a theme is complicated there are several means to provide information. In the previous chapter small multiples were mentioned; blocks of text can be added to provide context, extension, or simply to clarify topics of import that are complex.



Layout

There are many things to consider when laying out a map. The geography being portrayed (and its size, shape, and orientation) will help establish many of the initial parameters. Making a map of Rhode Island, Chile, or Saskatchewan presents a different geography from Montana, Canada, or the Hawaiian Islands (IMAGES). Also, the relationship between the place being mapped and the audience will determine whether a locator or inset map is necessary. The theme being mapped and that theme's complexity will determine how many ancillary elements are required and how much space each might consume (title, legend, source material, etc.).

When a map reader first looks at a map, a cartographer should assume that they will look from top to bottom and left to right (like they might read a block of text). A good wikipedia entry on visual search can be found here: https://en.wikipedia.org/wiki/Visual_search. We can make elements of a map prominent to draw the map reader's eye. This is sometimes called the pop-out effect; for instance, font size and colour (contrast) can be used to draw a map reader's eye. In cartography we are sometimes constrained by the geography of our area of study and what places in that geography are important to the map's theme. If an important cluster is located in the southern area of a north oriented map, the map reader will not immediately be drawn to that part of the map. If there is concern that such a delay might impede the map's utility, the cartography can draw the reader's attention with a design tool, such as drawing a prominent circle around the area or using a clear callout.

In general, when laying out a map the above set of map elements should be considered sequentially. As each element is added or considered, the previously included elements will provide context. Additionally, when a draft layout is accomplished, the map reader should consider the complete layout and return to each element, assessing its utility and whether it compliments the map's theme and the other elements.

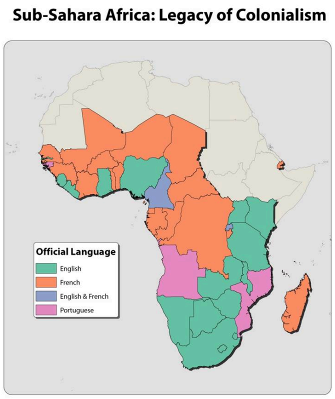


Figure – Ground

In design, the figure-ground relationship establishes, visually, that which appears to be in front of the other. Foreground/Background are synonyms. Various aspects of a visual display can manipulate what appears to be figure (in front) and what is the background (ground). Central position, texture, value (darker), complexity, among other things can make portions of a display appear to be in the foreground.

Typeface and Fonts for Mapping

Typography in cartography is an important ancillary component in the mapping process. The simultaneous goals are:

1. to share information with the map reader that cannot otherwise be communicated visually on or with the map
2. to NOT distract from the presentation of spatial and non-spatial information on the main body of the map (the geographic extent of the map)

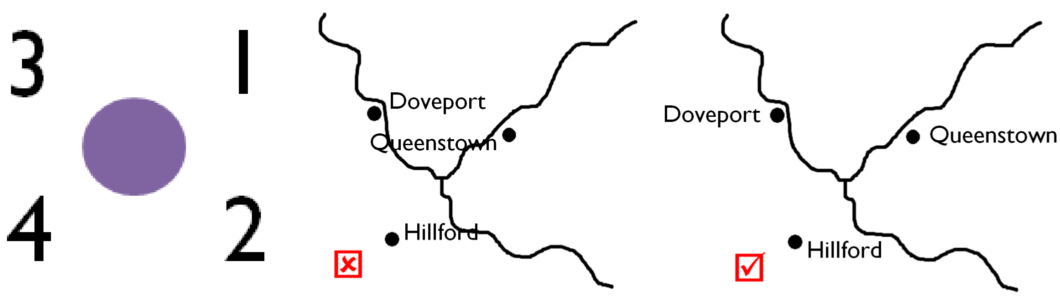
Typeface: the collective shapes of letters that differentiate one collection from another. For a given typeface there can be many fonts (bold, narrow, italics, etc.). When a "font" is named (helvetica, garamond, georgia, etc.) that name refers to the typeface, although font is an acceptable synonym with all but the most pretentious (or typeface/ font nerdy) audience.

Two categories of typefaces that are important to communication are Serif and Sans Serif forms. A serif is a small "tail" added to the of the lines that make up letters. Serif typefaces can be more easily read as the serif helps our eyes make the transition between neighbouring letters and words when reading. Serif type is often used in map titles, but not exclusively. Sans Serif type does not have these small tails. Sans Serif typefaces can be composed to lines of varying width, they are not restricted to a single line width.

Spacing: When using text, the spaces that letters and words occupy can be manipulated. This is not a common practice in settings when large blocks of text are presented (books, manuscripts, essays, etc.) but is useful in graphic design (maps, cartoons, posters, etc.). **Line spacing** occurs when the space between separate lines of text is altered. **Word spacing** occurs when the space between separate words in a set of text is altered. **Character spacing** occurs when the space between separate letters in a word is altered. **Kerning** is the manipulation of letter-by-letter character spacing.

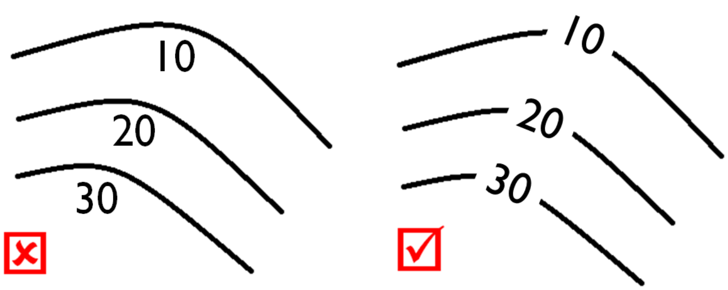
Arrangement: Unlike a lined page or a block of text, typography on a map is less constrained. This provides the map maker both opportunities and challenges. The text is generally shorter groups of words (or single words). Type might be ancillary to the geography (in the case of the legend and title) or it might be labeling the geography (rivers, town, landmarks, regions, etc.). In the latter case the constraint on placement is that the text should be easily associated with the feature it is labeling. There are some guidelines for features of each geometry: points, lines, and areas.

Points should be labeled so the label can be easily associated with the point symbol. The text can be above, below, or beside the symbol; generally a placement that results in the label "leading to" or "leading from" the symbol is preferred.

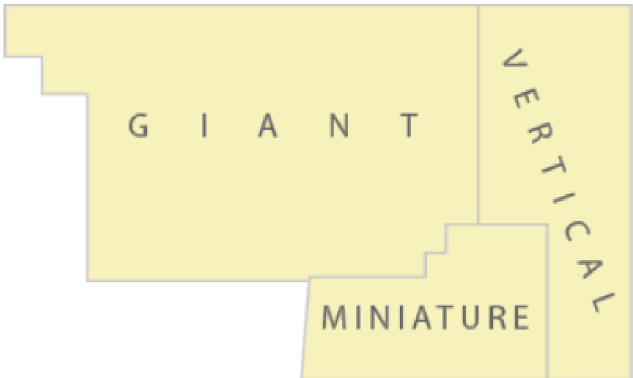


Lines should be labeled so that the text is above the line and follows its general form/shape (if the linear feature is not a straight line). Placement should result in the text flowing in a normal reading orientation (generally, upright from left to right). If the linear symbol is a complex curve (a meandering river or a mountain road)

the curve of the text should be less complex. Some linear features present themselves as special cases; the most common is waterways (rivers, creeks, lakes, etc.) that are linear but have area, (geometric polygons). In these cases the linear feature is labeled inside its area, the label follows its form, and the label might be repeated in case of long features. Furthermore, in the case of topographic maps, there are often even more guidelines that cover typeface and font characteristics. For instance, in Canada water features are labeled with a serif typeface in italics, and blue (darker than the blue of the "water" in the background). When linear features on long, text should be character-spaced, but not so much that the label is potentially unreadable; in the case of waterways, repeating the label is often preferred instead of character spacing.



Areas are labeled similarly to lines (but not identically). Labels should "fill and follow" the form of the area. Features are labeled once. For large features (or where there is a disparity between the space taken up by the label and the feature being labeled) character spacing and complimentary word spacing is recommended.



An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=37

III

Colour Theory and Cartography

Introduction

Colour is a fascinating topic; it is central to cartography, GIS, and Remote sensing. Colour is important to all three cartographic traditions and is an essential element in creating images that can be used for interpretation and presentation. The concept of colour is made possible by our sense of vision. For geomatics and cartography, human vision will be categorized in two ways. First, we will discern between the two different receptors of visual information in the retina of our eyes, Our retina is the rear surface of our eye Rods and Cones (figure 1, anatomy of the human eye). Second, the cones of our retina will be classified as sensitive to different wavelengths of incoming light.

The rods are more sensitive to light, than cones, and, as such, are responsible for visual perception during low light conditions. They are also more useful for picking up movement. Rods are not present at the center of the retina (called the fovea), but are spread out across the rest its surface. Cones are highly concentrated at the centre. Cones are responsible for our ability to perceive colour. They are less sensitive to light, and the light reaching them is best perceived if it is direct.

In almost all humans, our cones can be classified by the wavelength of incoming radiation (light) to which they are sensitive. Humans are called *trichromats*, which means are cones are sensitive to three different wavelengths of electromagnetic radiation (there are three types of cones, one for each range of wavelengths) (see sidebar on electromagnetic radiation). The cones of our retina are sensitive to either red, blue, or green light. The wavelengths for red light are the longest; the cones for red are sensitive to radiation with wavelengths near 565 nm. Green light has wavelengths above and below 535 nm. Blue light has the shortest wavelengths (of the three), ranging around 420 nm. All of the colours we see are a result of different amounts of each of these three wavelengths of radiation reaching our cones. In a later section we will talk about theories of colour mixing and how different colours are created by mixing these primary colours.

(SIDEBAR) The electromagnet spectrum is made up of radiation that ranges from long to short wavelengths (some might describe the range as very, very long to very, very short). Radiation can be of any wavelength, but ranges of wavelengths have been defined to differentiate between meaningful properties of the radiation within a range of wavelengths. The most important range of wavelengths for human vision is called the visible spectrum. It ranges from 360 nanometers (nm) and 780 nm. (SIDEBAR)

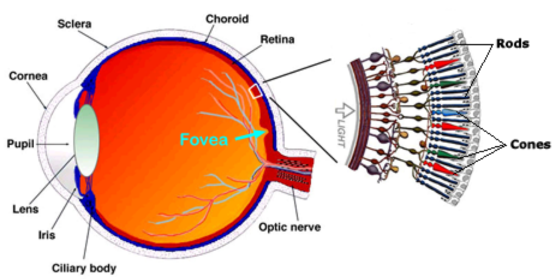
Visual Elements of Colour: Hue, Value, and Chroma

Hue is the aspect of colour that we normally associate with the names of colours, such as purple, brown, red, magenta, and chartreuse. Hues are created by combining different amounts of primary colours.

Value can be summarized as the amount of gray in a colour. At 100% the "colour" is black, but the amount of gray can range from 0% to 100%. Value increases when more of the hue is visually available, so value decreases as gray is added. Different hues respond differently (but continuously for a given hue) to the addition of gray. For instance, yellow can be perceived (as yellow) with the same percentage of gray that will make blue appear black.

Chroma is the intensity of a hue. Like value, chroma can range from negligible (near 0%) to 100%. A colour presented with low chroma will appear light or "faded." A low chroma colour will also be more sensitive to decreasing value (adding gray).

For cartography it is important to appreciate how the visual elements can be used to communicate patterns, differentiate among variables or values, and either confuse or clarify a map's message. Recall from earlier that both what we are communicating and how we are communicating it (the mix of colour elements we are using on a map or graphic) has an amount (level) of information, either nominal, ordinal, interval, or ratio. First, the level of information that each perceptual element conveys needs to be established. While hue (using different colours) can be used to convey value (ordinal and maybe interval, ratio) it is most logically applied to nominal variables.



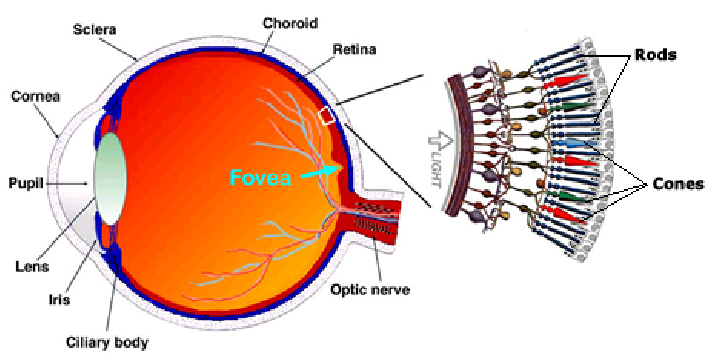
IMAGES: anatomy of the human eye

8.

Introduction and Perceptual Elements of Colour

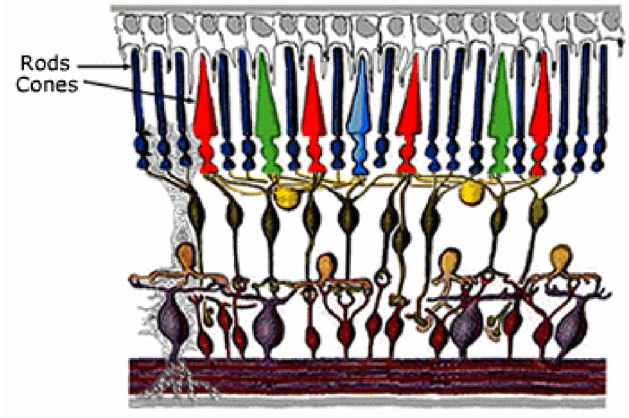
Introduction

Colour is a fascinating topic; it is central to cartography, GIS, and remote sensing. Colour is important to all three cartographic traditions and is an essential element in creating images that can be used for interpretation and presentation. The concept of colour is made possible by our sense of vision. For geomatics and cartography, human vision will be categorized in two ways. First, we will discern between the two different receptors of visual information in the retina of our eyes. Our retina is the rear surface of our eye Rods and Cones (figure 1, anatomy of the human eye). Second, the cones of our retina will be classified as sensitive to different wavelengths of incoming light.



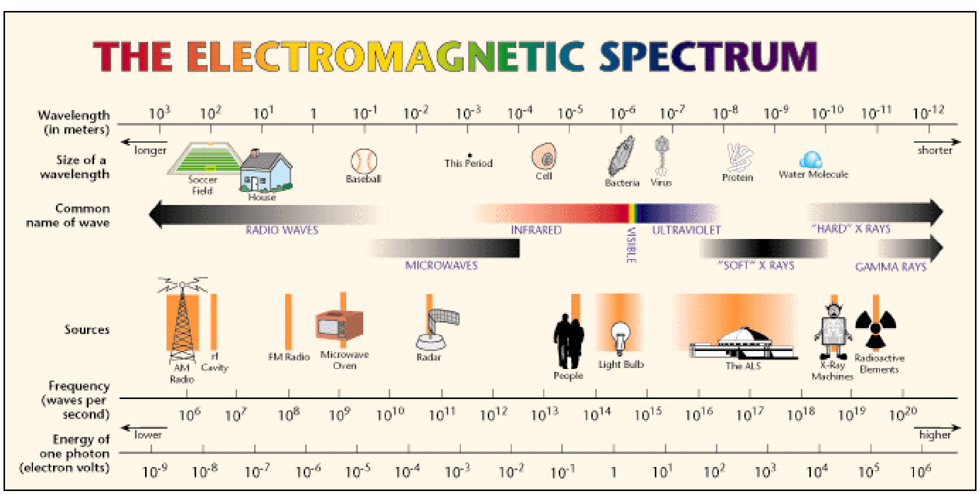
How Our Visual Perception Works

The rods are more sensitive to light than cones, and, as such, are responsible for visual perception during low light conditions. They are also more useful for picking up movement. Rods are not present at the centre of the retina (called the fovea), but are spread out across the rest its surface. Cones are highly concentrated at the centre. Cones are responsible for our ability to perceive colour. They are less sensitive to light, and the light reaching them is best perceived if it is direct.



In almost all humans, our cones can be classified by the wavelength of incoming radiation (light) to which they are sensitive. Humans are called *trichromats*, which means our cones are sensitive to three different wavelengths of electromagnetic radiation (there are three types of cones, one for each range of wavelengths, see sidebar on electromagnetic radiation). The cones of our retina are sensitive to either red, blue, or green light. The wavelengths for red light are the longest; the cones for red are sensitive to radiation with wavelengths near 565 nm. Green light has wavelengths above and below 535 nm. Blue light has the shortest wavelengths (of the three), ranging around 420 nm. All of the colours we see are a result of different amounts of each of these three wavelengths of radiation reaching our cones. In a later section we will talk about theories of colour mixing and how different colours are created by mixing these primary colours.

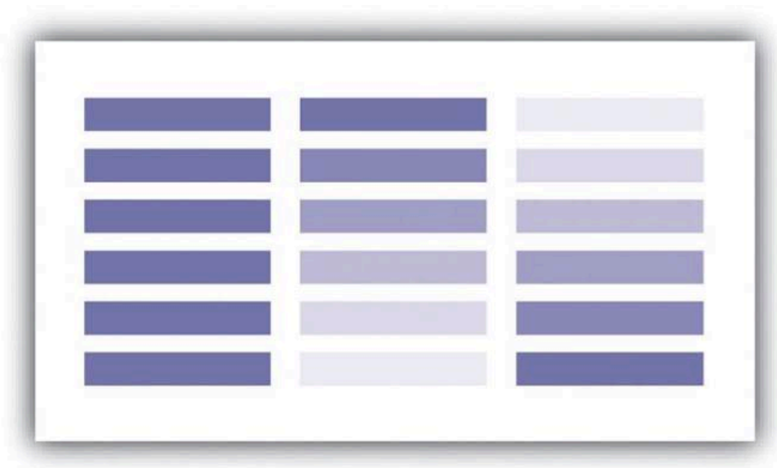
(SIDEBAR) The electromagnetic spectrum is made up of radiation that ranges from long to short wavelengths (some might describe the range as very, very long to very, very short). Radiation can be of any wavelength, but ranges of wavelengths have been defined to differentiate between meaningful properties of the radiation within a range of wavelengths. The most important range of wavelengths for human vision is called the visible spectrum. It ranges from 360 nanometers (nm) and 780 nm. (SIDEBAR)



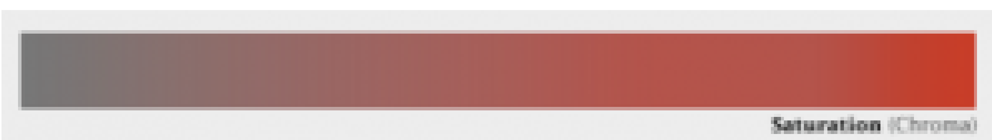
Visual Elements of Colour: Hue, Value, and Chroma

Hue is the aspect of colour that we normally associate with the names of colours, such as purple, brown, red, magenta, and chartreuse. Hues are created by combining different amounts of primary colours.

Value can be summarized as the amount of grey in a colour. At 100% the "colour" is black, but the amount of grey can range from 0% to 100%. Value increases when more of the hue is visually available, so value decreases as grey is added. Different hues respond differently (but continuously for a given hue) to the addition of grey. For instance, yellow can be perceived (as yellow) with the same percentage of grey that will make blue appear black.



Chroma is the intensity of a hue. Like value, chroma can range from negligible (near 0%) to 100%. A colour presented with low chroma will appear light or "faded." A low chroma colour will also be more sensitive to decreasing value (adding grey).



For cartography it is important to appreciate how the visual elements interact with the variables being mapped to communicate patterns, differentiate among variables or values, and either confuse or clarify a map's message. Recall from earlier that both what we are communicating and how we are communicating it (the mix of colour elements we are using on a map or graphic) has an amount (level) of information, either nominal, ordinal, interval, or ratio. First, the level of information that each perceptual element conveys needs to be established. While hue (using different colours) can be used to convey value (ordinal and maybe interval, ratio) it is most logically applied to nominal variables. Value and Chroma can both be manipulated to communicate continuous change in ordinal, interval, or ratio variables. Generally, darker means more, so decreasing value or increasing chroma for higher values of a variable is a logical choice.

What can add complication is the relationship between a map's theme and variable(s) being mapped. For instance, we can measure poverty by calculating the average household income and assigning colour to a choropleth map. However, the variable is income, so higher income indicates less poverty. In this case, it might be useful to reverse to application of the visual element being varied on the map (value or chroma) so that lower income values are assigned darker or brighter colours to indicate higher rates of poverty. It IS complicated; it is possible still, that such a map would be misinterpreted as many people will associated less income with poverty and might expect income to the variable used to represent poverty (or affluence).

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=111

9.

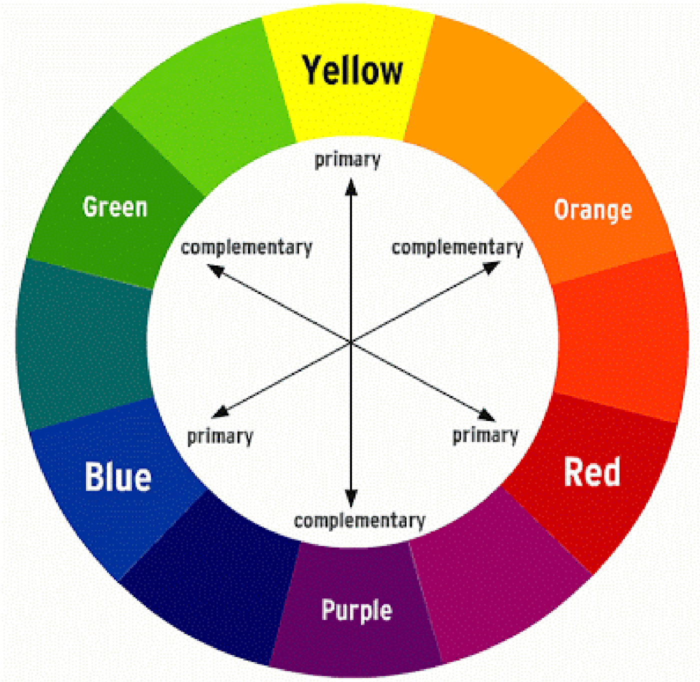
Colour Theory

Introduction

As described in the preceding chapter, colour is the result of how our eye processes incoming radiation in the visual portion of the electromagnetic spectrum. Interestingly, our vision, brain, and interactions between and among colours can have specific implications for what we see when presented with the same colour in different situations. Some of these interactions might be classified as optical illusions, others are simply the result of how colour is seen when placed in the context of other colours.

Colour Wheel

A (the) colour wheel is a useful tool for understanding and explaining the interactions and relationships between and among colour. It is arranged with colours that are near each other having similar properties. These colours (nearby) are called **analogous**. Colours on opposite sides of the colour wheel (such as red and green) have contrasting properties and are called **complementary**. Another property that is visually apparent is the non-visual feeling of colours arranged on the wheel; red, orange, and yellow are consider **warm colours**, which blue, green, and purple as **cool colours**. These properties can be useful on maps, as having a colour scheme range from cool to warm is intuitively interpreted as related to increasing value. Throughout this discussion of colour, it will be helpful to refer back to the colour wheel.

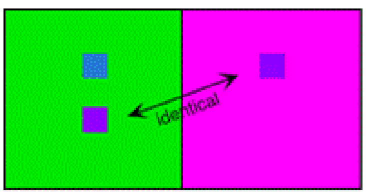


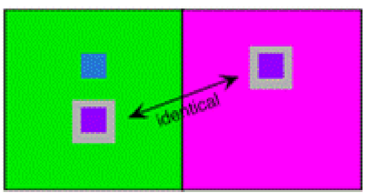
Colour Effects

There are many colour effects that are related to the integration of two or more colours on a single display or with a foreground and background colour. Rather than explain them all, you should read one of several online summaries.

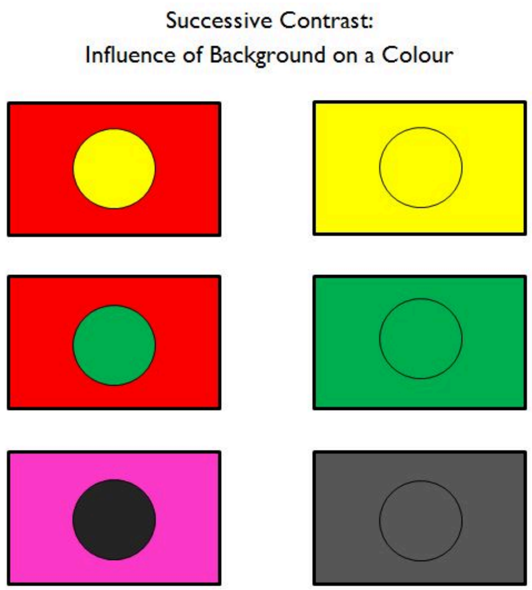
http://www.uxmatters.com/mt/archives/2006/01/color-theory-for-digital-displays-a-quick-reference-part-ii.php

Simultaneous Contrast



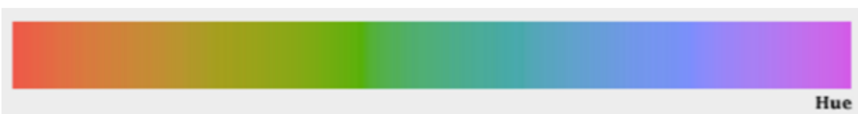


Successive Contrast



Colour Mixing

As described earlier, our eye anatomy is only sensitive to the wavelengths of three visual colours (Red, Blue, and Green). From these three incoming forms of radiation (colours) all other **hues** are can created.



This is accomplished through a process called Colour Mixing. As a result of different amounts of Red, Blue, and Green radiation reaching our retina our eyes generate input for a brain that produces all possible colours. If the only radiation that reaches our retina is in the red part of the electromagnetic spectrum, then we will see red. If equal amounts of red and green reach our retina we will see a different colour… YELLOW!

A good reference for colour mixing can be found at:

http://www.uxmatters.com/mt/archives/2006/01/color-theory-for-digital-displays-a-quick-reference-part-ii.php

Colours are perceived in specific ways and the interaction of incoming red, blue, and green light is always the same for the same "amounts" of each type of radiation. We can explain the perception of different hues in two different way: Additive Colour Mixing or Subtractive Colour Mixing.

Additive Colour Mixing approaches hue production (what hue is perceived by our eye-brain matrix) from the perspective of what proportions of red, blue, and green light reach our retina, are perceived by our eyes' cones, and are processed by our brain. The **primary colours in additive colour mixing are red, blue, and green**.

Subtractive Colour Mixing approaches hue production from the perspective of what proportions of red, blue, and green are MISSING from the light that reaches our retina. **The primary colours of subtractive colour mixing are Cyan, Magenta, and Yellow**. When our mind perceives the colour yellow, the only light reaching our eye is an equal mix of red and green. The primary colours for subtractive colour mixing are each the result of equal amounts of radiation reaching our retina from two of the additive primary colours.

Red + Green = Yellow

Red + Blue = Magenta

Blue + Green = Cyan

As a result, when working with paint, pigment, or other physical liquids or gels that are Yellow, Magenta, and Cyan, we can create by mixing different proportions of each all hues. Looking back at the colour wheel, it is more clear the relationships between the primary colours of each type, for each set of primary colours (additive = R, G, B and subtractive = C, M, Y) the other set of primary colours are the secondary colours.

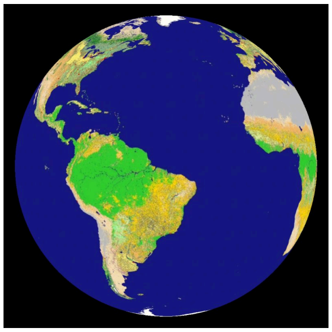
An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=124

IV

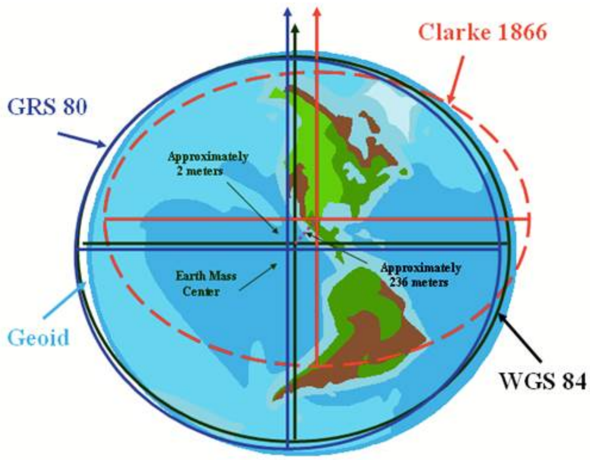
Coordinate Systems and Map Projections

10.

Measuring and Modeling Earth

Earth is Earth-shaped; it is also Earth-sized. For geomatics, accurate estimates of Earth's size and shape are important, particularly when mapping large areas. We can generally agree (and most people will posit)  that Earth is round (like a ball, not a disc). Scientists (mostly geophysicists) have used many methods to measure Earth's size and shape. In the 3rd Century BCE, Eratosthenes (a librarian! like my Dad) used measurements at two locations separated by 805 km, along with arithmetic and geometry to estimate Earth's circumference to be between 40,250 km and 45,900 km. From that time forward estimates have become increasingly accurate. The first estimate of Earth's size (such as Eratosthenes') assumed Earth was spherical.

A sphere is a three dimensional shape with a single radius (distance from the edge to centre is the same everywhere). Modern models for mapping and most geomatics applications assume an oblate ellipsoid. An ellipsoid is a three dimensional shape specified by two radii, one longer than the other and orthogonal to each other (meet at a right angle at the centre of the shape). That means that an ellipsoid is longer in one dimension (it's either "taller" than it is "wide," or "wider"). Earth is bigger around the equator than when measured around the poles (from north to south). The orientation of its shape  is defined by the term oblate. The internet (google) defines oblate as: "flattened at the poles." if an ellipsoid is "taller" that it is “wide,” it is called prolate. **Earth is oblate**. BUT, Earth is covered in water, mountains, rivers, valleys, lakes, etc., it’s topographically more complicated than an ellispoid. Any model for Earth’s size and shape simplifies that complexity, doing so accurately is important. Any estimate of Earth’s size and shape is called a **geodetic datum**.

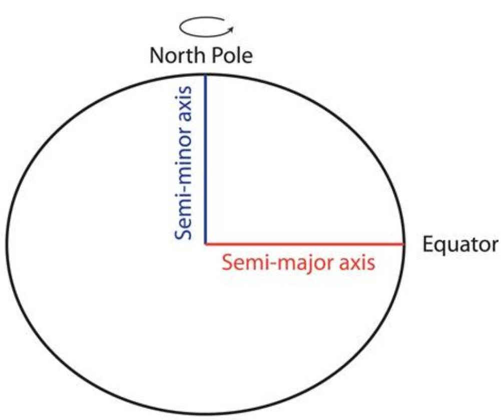


11.

Frames of Reference and Coordinate Systems

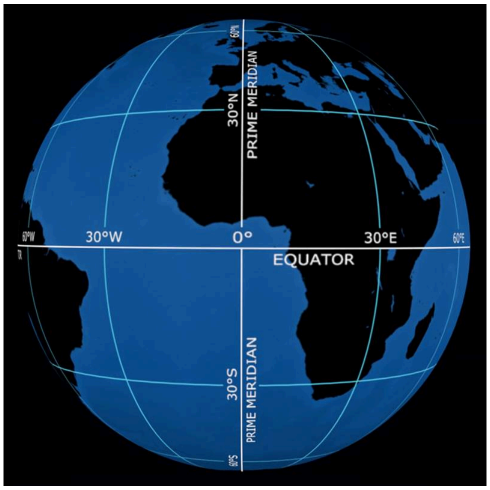
Introduction

In order to place things in space (such as a location on Earth) a frame of reference is required. A frame of reference offers the spatial extent for all that will be placed, or can be placed, on Earth. If we want to place locations on Earth, we need a frame of reference that covers Earth's full extent (all of it). The internet (google) defines a frame of reference as "a system of geometric axes in relation to which measurements of size, position, or motion can be made." For geographic information, we need "axes" that will allow us to place and measure things on Earth. This is a lot easier if we assume that Earth is an easily defined shape (previous chapter), so we almost always use an oblate ellipsoid geodetic datum. Remember that this is a three dimensional roundish shape defined by two radii (one polar and one equatorial).



Once we have the shape and size, we establish units and origin. In the section below, the units and origins for four different types of spaces will be presented and are defined by their shape (planar (flat, square, grid, etc.) or polar (round)) and dimensionality (two or three dimensional). Planar coordinate systems use **distances** away from the origin while polar coordinate systems use **degrees of rotation** away from the baseline around the origin for at least one of the two units.

On Earth, Latitude and Longitude represent a graticule of intersecting lines that serve as a universally accepted frame of reference. The origin is the equator (0 latitude) and the Prime Meridian (0 longitude) with 90 degrees of rotation north or south from the equator to the respective poles and 180 degrees of rotation east or west from the prime meridian. Latitude and Longitude are units in a three dimensional polar coordinate system.



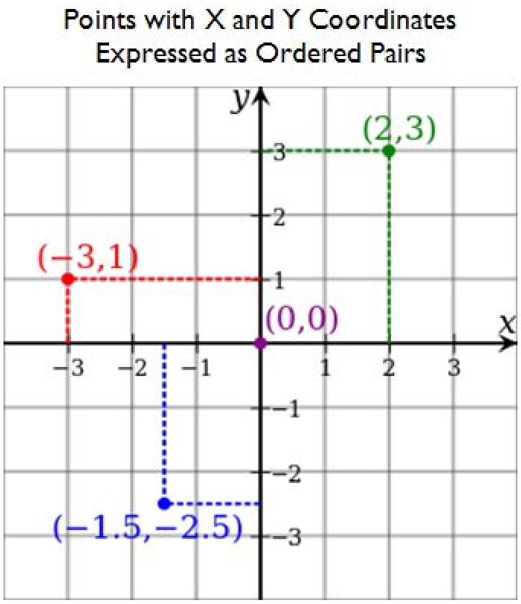
Non-Earth Based Coordinate Systems

There are four classes of coordinate systems relevant to geomatics, they are defined by their dimensionality (two-dimensional and three-dimensional) and the units of measurement (somewhat related to the shape of the space: round or orthogonal): distances or angles. Orthogonal means the space is composed of right angles. The orthogonal frames of reference (coordinate systems) will be referred to as **Cartesian**. Cartesian means the frame of reference is composed of baselines (axes) that intersect at right angles. The intersection point is the frame's origin (0, 0) from which locations are specified by a length (or distance) along two axes. These two axes are often referred to as X and Y. The other set of coordinate systems are called **Polar** and use an angle of rotation around the origin as at least one of the coordinates. (see figures below)

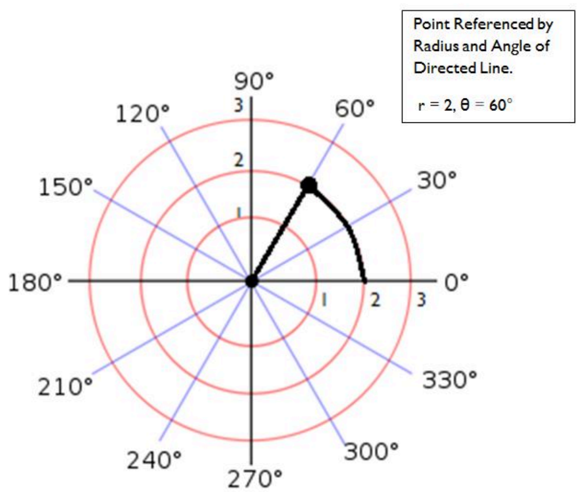
Two Dimensional Coordinate Systems

Two dimensional coordinate systems are also referred to as planar and indicate location on a flat space (or a representation of space). If the coordinate system is Cartesian, the two coordinates indicate distances along one of two axes away from from the origin. If the coordinate system is Polar the pair of coordinates indicate a distance along on an axis and a degree of rotation away from that axis around the origin.

Cartesian:



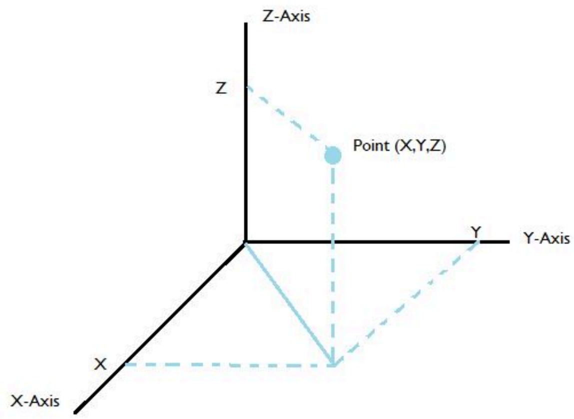
Polar:



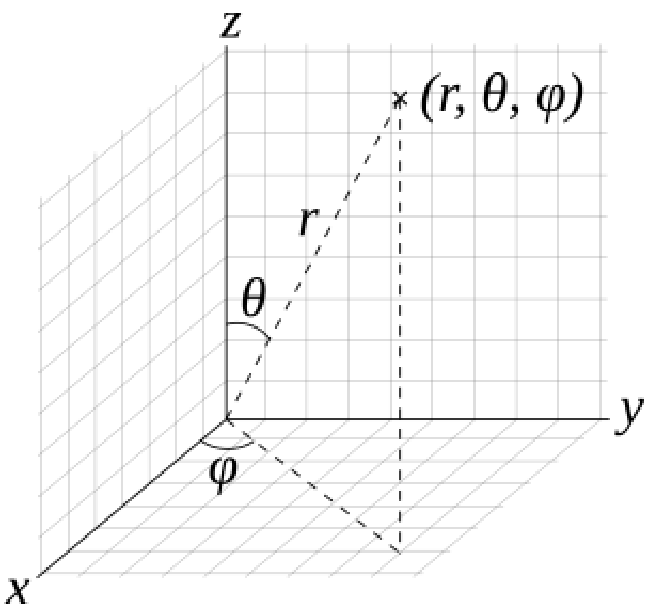
<u>Three Dimensional Coordinate Systems</u>

Three dimensional coordinate systems indicate position in a volumetric space. If the coordinate system is Cartesian the three coordinates indicate distances along one of three axes away from the origin. If the coordinate system is Polar the three coordinates indicate a distance along an axis and two degrees of rotation away from that axis around the origin, each orthogonal to the other.

Cartesian:



Polar:



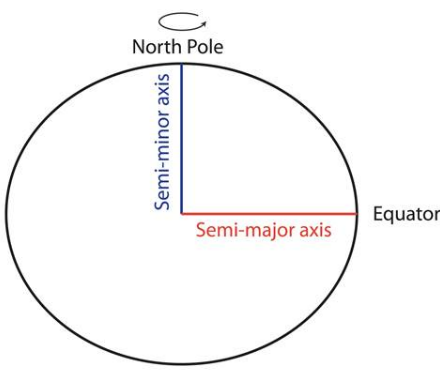
elipsoid, sphere, oblate, and prolate ellipsoids

12.

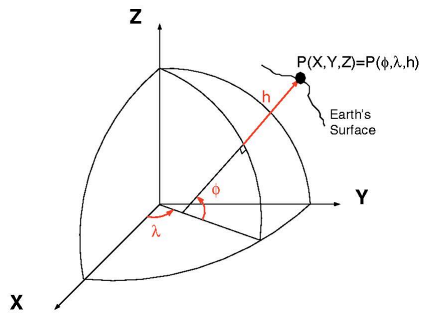
Cartesian/Projected Coordinate Systems, UTM

Introduction

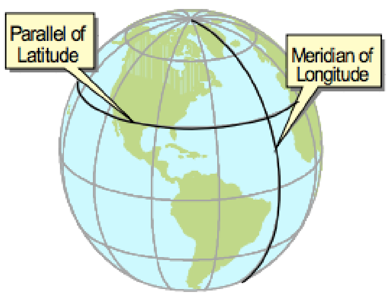
When we translate the previous topic, coordinate systems of different types and dimensions (polar/cartesian and 2D/3D), to Earth, we need to integrate what we know about Earth's size and shape. As summarized in this section's first chapter, Earth is Earth-sized and shaped, but can be simplified to an oblate ellipsoid. An oblate ellipsoid is a 3-dimensional round shape, making its native coordinate system a Polar 3D system.



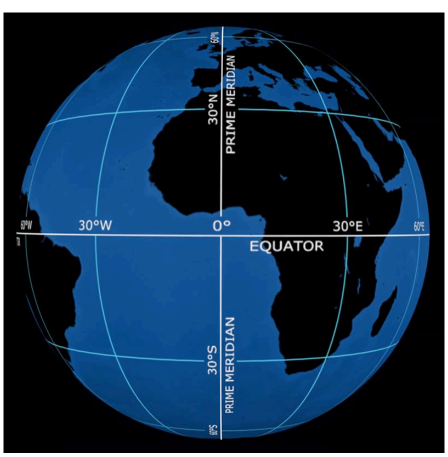
The origin for this system is the centre of the ellipsoid, with the distance unit representing the distance from the origin to the surface of the ellipsoid. This means the distance to the surface changes continuously from the equator to poles. The distance to the surface from the centroid of the geodetic datum (Earth model) is the radius of Earth. For the most part we treat the surface (and corresponding value for the distance to it from the ellipsoid's centre) as a constant.



For the two remaining units, each indicating a number of degrees of rotation around the origin, we use Latitude and Longitude. The origin for these values is the intersection of the Equator and the Prime Meridian. The equator is the midpoint between the two poles. The two poles are defined as the points connected by a line perpendicular to Earth's rotation.The Prime Meridian is an arbitrary, but negotiated, line that passes through Greenwich, England.

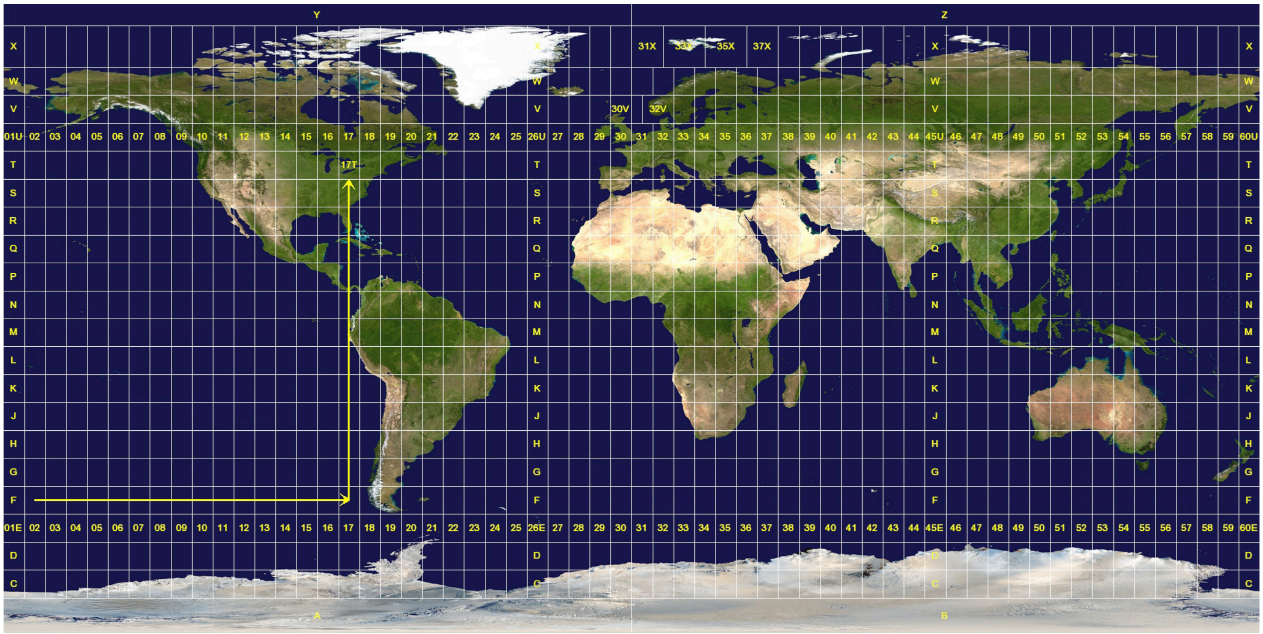


Having a coordinate system that is native to Earth's shape (or close to it) is intuitive and efficient. In particular, once an origin is specified all locations can be expressed with just two values, Latitude and Longitude, in a common unit, degrees. The challenge is using such values in calculations of area and distance. The math is complicated, to say it simply. If we ask a question from the perspective of calculation complexity, such as "what type of coordinate system supports the most simple distance and area calculations?" The answer is 2-dimensional cartesian, such as Universal Transverse Mercator (UTM)

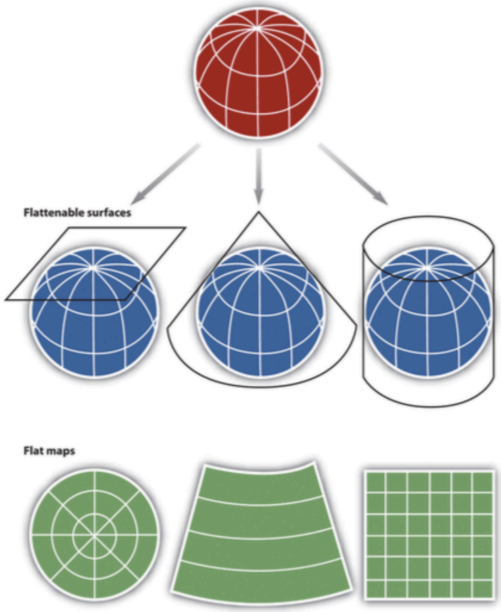


UTM

Universal Transverse Mercator (UTM) is a map projection-based global coordinate system that provides location information using pairs of Cartesian coordinates in metric units (metres). UTM offers location information in a coordinate system (2-D cartesian) that offer a simpler calculation environment than its most popular 3-D polar counterpart (Latitude and Longitude). As you have learned, Earth is a 3-D ellipsoid, making its native coordinate system 3-D polar.



Conversions from one coordinate system to another requires a mathematical process called projection. In this textbook and the accompanying course, this process was initially introduced as a process that uses the term “projected” in a colloquial sense; the surface of Earth was projected onto a paper surface using a light source casting shadows onto a piece of paper. This simultaneously relies on the complexity of systematically converting coordinates in one system to another system. The mathematics aren’t trivial, but by initiating the process with shadows, light, and piece of paper does offer a tantalizing means to “get the idea” of what is happening.



Coordinates in UTM are specified by a Zone (1-60), a hemisphere (N or S), and a Northing (in metres), and Eastings (also in metres). Let's figure those out.

First Conversion

Actually 60 of them. As we have learned, the amount of error, or uncertainty, on a map projection increases as distance from the standard point, line, or lines increases. Close to a standard line we have less to worry about. UTM is composed of 60 separate zones, each a small sliver of Earth centred on a line of longitude. The name of the projection and its orientation is hidden in plain sight (in the name of the coordinate system). The projection is the Mercator and its orientation is transverse. In this situation transverse means rotated 90 degrees from Mercator projection's normal orientation. The Mercator projection's normal orientation is with the standard line at the equator. Lines of longitude run at a 90 degree angle to the equator.

BUT, why 60? As noted, increasing distance from the standard line increases error. Therefore, the UTM system uses a single projection for just a 6 degree swath of Earth. As an ellipsoid, the circumference of Earth is 360 degrees (360/6 = 60 zones!). Each zone has a name, starting with the line opposite the Prime Meridian (180 degrees, east or west), and moving east. The first zone is zone 1, the final zone is zone 60. Saskatchewan is covered by Zone 13.

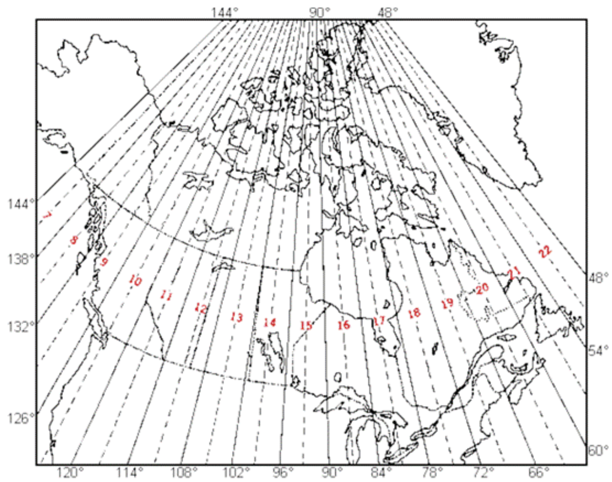


Figure 2 - UTM Zones and Central Meridians for Canada

Coordinates in UTM start with the zone, then the hemisphere: North (N) or South (S). Dividing the system into N and S allows the system to use the equator and its fictional, but parallel equivalent as the baselines for one of the coordinate pairs. This is elegant (it really struck me at learning it, a simple insight, but one that supports an intuitive conversion from polar to Cartesian coordinates). When observing Earth with lines of latitude and longitude depicted and making the centre of the “picture” the intersection of the equator and a line of longitude (like 105 degrees W) looks just like the intersection of two lines in a 2-D cartesian coordinate system. What we AREN’T seeing though, is that each line (0 degrees latitude and 105 degrees west) is receding away from us (the viewer) along the surface of the ellipsoid.

13.

Making Earth Flat: Map Projections

Introduction

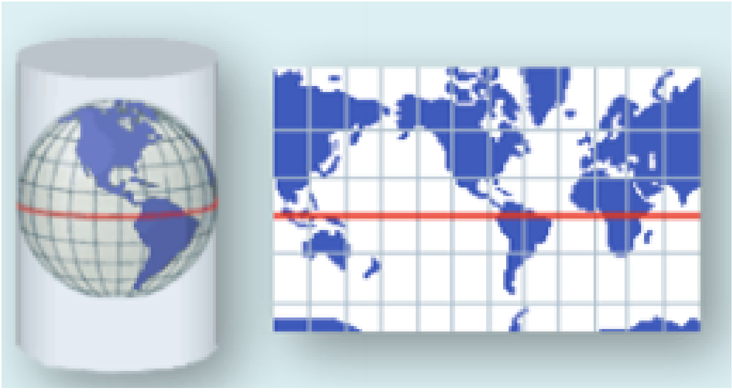
Maps are flat representations of Earth. As we've learned, Earth is large and complicated, but we can simplify it so the information we want to display can be preserved, represented, and analyzed. An important step in the simplification is taking the surface of Earth and presenting it on a flat (planar) surface. When Earth's surface is converted (projected) from 3 dimensions to 2 dimensions, there are trade-offs. In addition to going from round (and 3-D) to flat, the representation must be reduced in size. This is the map's scale (see and Scale chapter for a definition of cartographic scale). All maps are projections, but only on small scale maps (remember **representational fraction**) does the distortion of the projection become readily apparent to anyone but a keen observer (like a student of geomatics). Maps of countries, continents, and the world, show their map projection distortions readily.

Organization of Chapter and Key Aspects of Map Projections

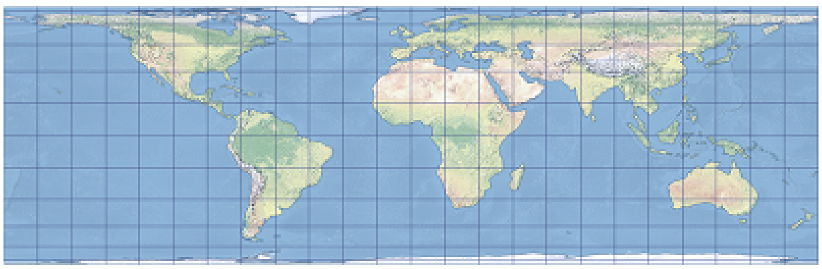
When projecting Earth's surface onto a planar (flat) surface something is lost, or given up. A map project can preserve the accuracy of some aspects of the arrangement and characteristics of features, but at the cost of something else. That which is preserved is one way to classifying map projection.

What Can be Preserved, and What is Lost

Shape: A map that preserve shape are called **Conformal**. The trade-off is that area is **exaggerated**. On a conformal map the shapes (of countries, islands, continents, etc.) are the same as they are on Earth's surface or a globe. That area is distorted means that we can't visually compare the size of different places on the map. On the map below compare (approximately) Alaska and Brazil. They might look similar in size, but Brazil is 5 times the area of Alaska. This is an example of a conformal map that minimizes distortion at the equator and preserves shape, necessarily exaggerating area as the distance from the equator increases.



Area: A map projection that preserves area is called an Equal Area projection. Like the conformal, there is a trade-off. That trade off is that shapes are distorted. Compare Alaska, or any place far from the equator, on this and the previous map.



In the two maps above (and the "unprojected" perspective image of a globe, on the left in the top image) the arrangement of the lines of Latitude and Longitude are an indicator of the nature of the distortion resulting from the projection. Lines (parallels) of Latitude and (meridians of) Longitude are lines that intersect one another at 90 degrees on Earth. Lines of longitude meet at the poles, lines of latitude are a constant distance apart and, other than the equator, are shorter than lines of longitude. A **great circle** is a line that bisects a sphere; all lines of longitude are great circles, the only line of latitude that is a great circle is the equator. Both lines of latitude and longitude follow a direction consistent with a compass bearing, and those respective compass bearings are orthogonal to one another (I believe this is a complicated concept being communicated in some precise language, readers should sit back and ponder what these words mean).

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=47

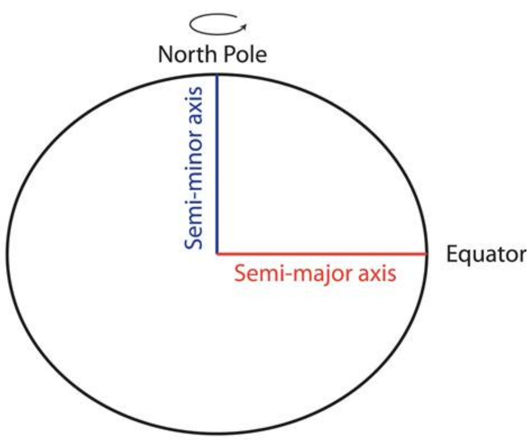
Graticule and Grid

A **graticule** is an arrangement of lines intersecting at 90 degrees ON A SPHERE. A **grid** is an arrangement of lines intersection at 90 degrees ON A PLANE (a flat surface). So, if you are using a projected coordinate system, then locations are specified on a grid. If you are using an unprojected coordinate system ("geographic coordinate system" in Pro/ESRI) then you are specifying locations on a graticule.

Steps to a Projection

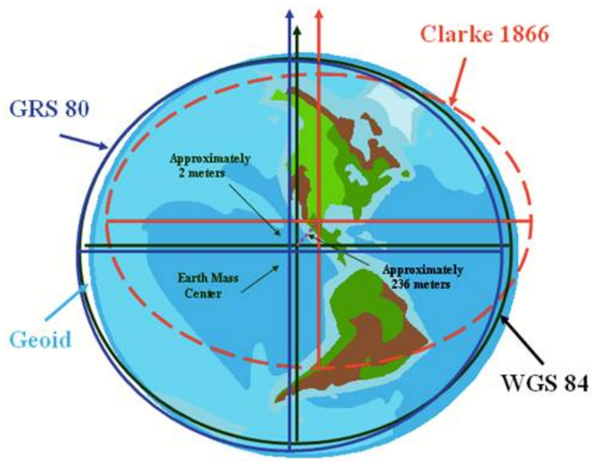
In order to get to a flat map projection there are a few steps.

1. Establish the size and shape for the model of Earth (and an orientation). Earth is big. Earth is Earth-shaped. However, to map it we simplify it to either a Spheroid or an Ellipsoid. A Spheroid is defined by a single diameter, all locations on the surface are the same distance from the centre of the sphere. An Ellipsoid is **defined** by two diameters, one for the poles and one for the equator. This results in a constantly (and smoothly) changing diameter from the equator to the poles. Earth is fatter at the equator than the from pole to pole. So, if we make the north pole the top, Earth is an oblate ("laying down") ellipsoid. With the size and shape defined this is a **Geodetic Datum.**



2. Once we have established a **Geodetic Datum**, the coordination of that shape with Earth can be set. This is similar to how the **standard line** for map projections established where on of the map (at a point or along a line, or lines) the scale is the same as the Geodetic Datum and is an undistorted representation of that point or line(s). Perhaps imagining trying to get two ellipsoids that are slight different into as close alignment with one another is a useful thought experiment. While different coordinate systems or projections might use the same **geodetic datum** (just the size and shape), it can be differently coordinated with Earth. For instance, WGS84 and NAD83 use the same **geodetic datum** (model for Earth's size and shape), **BUT** they differ in where they are placed in relationship with Earth, where they touch Earth. Why? WGS stands for World Geodetic System, NAD stands for North American Datum, so the names offer a clue. WGS places the Datum so that errors are minimized globally,

NAD places the Datum so that errors are minimized for place on or near North America (at the cost of accuracy elsewhere).



“pinning” a geodetic datum to Earth (the relative sizes of the WGS 84 and Clarke 1866 oblate ellipsoids is exaggerated).

3. Projection of the surface of the Geodetic Datum from its 3-D surface to a 2-D plane

Making accurate maps of places requires a projection of Earth’s surface to a plane (flat surface). For this treatment a projection is a mathematical statement (algorithm) that defines the relationship between what is being represented (the surface of the geodetic datum) and the representation (the 2-D version of the surface). In order to portray the Earth’s surface, data must have a recording of the Geodetic Datum and its relationship with the Earth object (itself a type of representation). This relationship includes how they are associated (explaining where they come into contact with each other), the origin for the coordinate system, and the nature of the coordinate system. In the case of a Projected map, the coordinate system is also projected and therefore rectangular, consisting of lines intersecting at right angles.

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=47

V

Geographic Information Systems

14. What is GIS?

GIS stands for Geographic Information Systems. GIS can be thought of as a set of software tools for handling, combining, analyzing, and communicating about our world, particularly in the spatial dimension. Maps are a common product of working with GIS. GIS can be used to created spatial answers (maps) to spatial questions, but GIS can also provide non-spatial answers to questions. Furthermore, the tools necessary for map making and geographic analysis are commonly used with non-spatial data. Mathematical, design, and many other functions, are useful with spatial information, but can be applied to non-spatial data as well.

The software of GIS is often conflated (equated) with GIS. In my teaching, I spend more than half of my practical teaching with sofware from ESRI (environmental systems research institute, but the longer name is almost never used, we just say ehz-ree). The software I am currently teaching with is ArcGIS Pro, this software came after “desktop.” While I try to refer to the current software as “pro” and it’s predecessor as “desktop,” I might be less clear from time to time. There are other software packages out there, some are task focussed, such as GeoDa for spatial statistics and exploratory spatial data analysis. Such software, unless focussed on map layout and design, often have rudimentary mapping capabilities, and maps are created and designed in other software. Other offer similar breadth to Pro, such as QGIS (Q stands for Quantum), and could be worth checking out. In my home department we also teach a series of classes on Remote Sensing, a field focused on satellite observations of Earth. There are several software packages for remote sensing. As computer software and hardware improve the line between GIS and remote sensing is blurred and each is adopting tools and capacities of the other.

One way I like to think about mapping and GIS is to consider how LITTLE information we need for a particular question. A central concern of GIS is the representation of almost anything, spatially. Earth is complex, perhaps unimaginably so. Therefore, it is important to simplify, generalize, and focus how we represent reality and use that representation to answer our questions. We can think of models of reality as being composed of the building blocks, and the geometry of those blocks as the foundation on which GIS is built. There are two ways we represent reality in GIS, one is Raster (like an digital image, made up of rasters, another word for pixels) and one is Vector (points, lines, and polygons). In a coming chapter these two models will be explored in more detail, and their simplicity will be obscured.

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=51

15.

Data and Information

(This chapter is shared from another Open TextBook, Essentials of GIS, Campbell and Shin)

Introduction

To understand how we get from analog to digital maps, let's begin with the building blocks and foundations of the geographic information system (GIS)—namely, data and information. As already noted on several occasions, GIS stores, edits, processes, and presents data and information. But what exactly is data? And what exactly is information? For many, the terms "data" and "information" refer to the same thing. For our purposes, it is useful to make a distinction between the two. Generally, **data** refer to facts, measurements, characteristics, or traits of an object of interest. For you grammar sticklers out there, note that "data" is the plural form of "datum." For example, we can collect all kinds of 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 used to obtain insights, they become **information**. For our purposes, **information** simply refers to the knowledge of value obtained through the collection, interpretation, and/or analysis of data. Though a computer is not necessary to collect, record, manipulate, process, or visualize data, or to process it into information, information technology can be of great help. For instance, computers can automate repetitive tasks, store data efficiently in terms of space and cost, and provide a range of tools for analyzing data from spreadsheets to GISs, of course. What's more is the fact that the incredible amount of data collected each and every day by satellites, grocery store product scanners, traffic sensors, temperature gauges, and your mobile phone carrier, to name just a few, would not be possible without the aid and innovation of information technology.

Since this is a text about GISs, it is useful to also define **geographic** data. Like generic data, **geographic** or spatial data refer to geographic facts, measurements, or characteristics of an object that permit us to define its location on the surface of the earth. 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 data and attribute data, which was discussed in Introduction to Geomatics. Initially, these two categories were presented as **spatial** and **non-spatial** information, here they are **geography** and **attribute**. Where geographic data are concerned with defining the location of an object of interest, attribute data are concerned with its non-geographic traits and characteristics.

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. For instance, we can define the location of your home many ways, such as with a street address, the street names of the nearest intersection, the postal code where your home is located, or we could use a global positioning system–enabled device to obtain latitude and longitude coordinates. What is important is 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 define the location of your home are the attribute data that describe the various qualities of your home. Such data include but are not restricted to the number of bedrooms and bathrooms in your home, whether or not your home has central heat, the year when your home was built, the number of occupants, and whether or not there is a swimming pool. These attribute data tell us a lot about your home but relatively little about where it is.

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=151

Not only is it useful to recognize and understand how geographic and attribute data differ and complement each other, but it is also of central importance when learning about and using GISs. 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, being able to determine which kinds of data you need will ultimately aid in your implementation and use of a GIS. More often than not, 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.

Data of Different Types and Files in which it is Stored

When we collect data about your home, rainforests, or anything, really, 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 useful to record them in digital form so a computer can read them. Though we won’t bother ourselves with the bits and bytes of computing, it is necessary to discuss some basic 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 itself 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., address book). Furthermore, computer files come in all different sizes and types. One of the clues we can use to distinguish one file from another is the file extension. The file extension refers to the letters that follow the period (“.”) after the name of the file. Table 1 contains some of the most common file extensions and the types of files with which they are associated.

Table 1

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 web pages)
filename.xml	Extensible markup language
filename.zip	Zipped/compressed archive

Some computer programs may be able to read or work with only certain file types, while others are more adept at reading multiple file formats. What you will realize as you begin to work more with information technology, and GISs in particular, is that familiarity with different file types is important. Learning how to convert or export one file type to another is also a very useful and valuable skill to obtain. In this regard, being able to recognize and knowing how to identify different and unfamiliar file types will undoubtedly increase your proficiency with computers and GISs.

Of the numerous file types that exist, one of the most common and widely accessed file is the **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 complex 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 data that are organized or structured in some fashion, it is sometimes called a flat file (but the file extension remains the same, i.e., .txt). Generally, 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”
Jake	5’9”
Harry	6’2”

Each row corresponds to one and only one record, observation or case. There are two other important elements to know about this file. First, note that the first row does not contain any data; rather, it provides a description of 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**. "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.

An interactive or media element has been excluded from this version of the text. You can view it online here: https://openpress.usask.ca/introgeomatics/?p=151

The second unseen but critical element to the file is the spaces in between each column or field. In the example, it appears as though a space separates the "name" column from the "height" column. Upon closer inspection, however, note how the initial values of the "height" column are aligned. If a single space was being 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 character that is used to separate columns within a flat file is called the delimiter or separator. 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 Height	Name, Height
Tim 6.1	Tim 6.1	Tim, 6.1
Jake 5.9	Jake 5.9	Jake, 5.9
Harry 6.2	Harry 6.2	Harry, 6.2

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

When recording and working with geographic data, the same general format is applied. 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	Language	Continent
France	65,000,000	French	Europe
Brazil	192,000,000	Portuguese	South America
Australia	22,000,000	English	Australia

Files like those presented here are the building blocks of the various tables, charts, reports, graphs, and other visualizations that we see each and every day online, in print, and on television. They are also key components to the maps and geographic representations created by GISs. Rarely if ever, however, will you work with one and only one file or file type. More often than not, and especially when working with GISs, you will work with multiple files. Such a grouping of multiple files is called adatabase. Since the files within a database may be different sizes, shapes, and even formats, we need to devise some type of system that will allow us to work, update, edit, integrate, share, and display the various data within the database. Such a system is generally referred to as a database management system (DBMS). Databases and DBMSs are so important to GISs that a later chapter is dedicated to them. For now it is enough to remember that file types are like ice cream—they come in all different kinds of flavors.

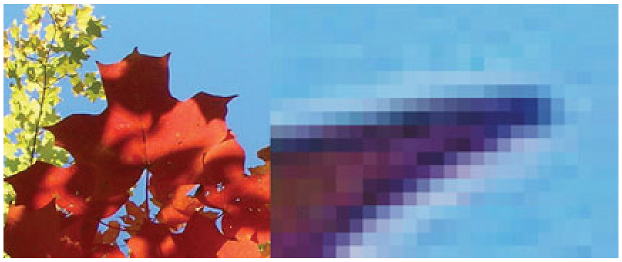
16.

Data Models: Representing Reality as Simply as Possible

Raster Data Model

The raster data model is widely used in applications ranging far beyond geographic information systems (GISs). Most likely, you are already very 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. Take a moment to view your favourite 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). Each of these uniquely colored pixels, when viewed as a whole, combines to form a coherent image (fig. 1).

Figure 1 Digital Picture with Zoomed Inset Showing Pixilation of Raster Image



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 nearly a century. The neoimpressionist artist, Georges Seurat, developed a painting technique referred to as "pointillism" in the 1880s, which similarly relies on the amassing of small, monochromatic "dots" of ink that combine to form a larger image (Fig. 2). If you are as generous as the author, you may indeed think of your raster dataset creations as sublime works of art.

Figure 2 Pointillist Artwork

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, the vast majority of available raster GIS data are built on the square pixel (fig. 3). 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).

Figure 3 Common Raster Graphics Used in GIS Applications: Aerial Photograph (left) and USGS DEM (right)

Source: Data available from U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD.

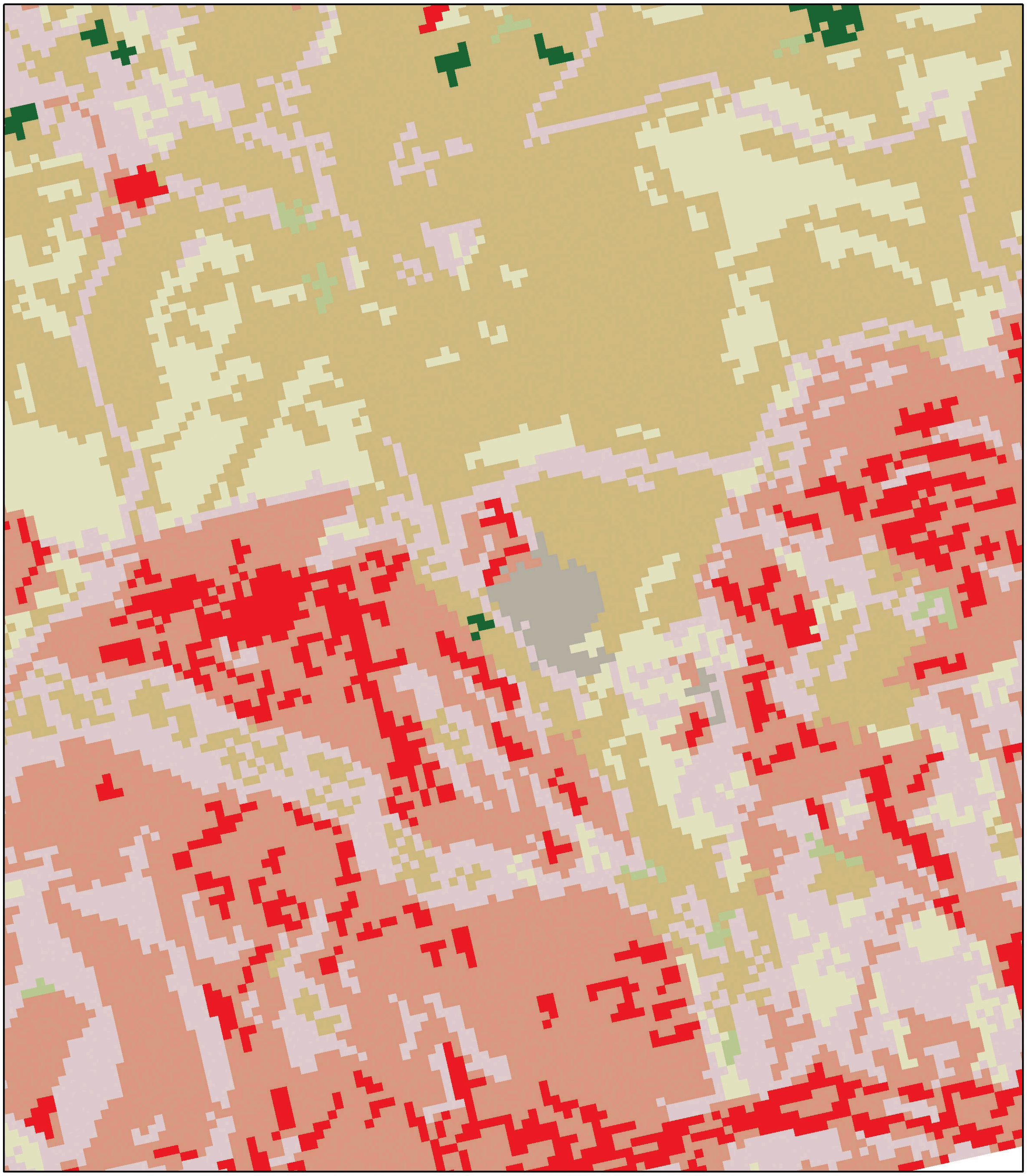
Because of the reliance on a uniform series of square pixels, the raster data model is referred to as a grid-based system. Typically, a single data value will be assigned to each grid locale. Each cell in a raster carries a single value, which represents the characteristic of the spatial phenomenon at a location denoted by its row and column. The data type for that cell value can be either integer (whole number) or floating-point (decimals). 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 of computer technology has made this second methodology increasingly feasible as large 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. The area covered by each pixel determines the spatial resolution of the raster model from which it is derived. Specifically, resolution is determined by measuring one side of the square pixel. 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 overly fine pixel resolution will result in significant increases in file size and computer processing requirements during display and/or analysis. An effective pixel resolution will take both the map scale and the minimum mapping unit of the other GIS data into consideration. 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 entire cell would not be filled by the pole. Rather, 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. Often, an arbitrary, readily identifiable value (e.g., −9999) will be assigned to pixels for which there is no data value. Second, a cell can hold any alphanumeric index that represents an attribute. In the case of quantitative datasets, attribute assignation is fairly 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. 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 (fig. 4). The third property of the raster data model is that points and lines "move" to the centre 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 centre 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.

Figure 4 Land-Use/Land-Cover Raster Image



Source: Data available from U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD.

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

1. Cell-by-cell raster encoding. This minimally intensive method encodes a raster by creating records for each cell value by row and column (fig. 5). This method could be thought of as a large spreadsheet wherein each cell of the spreadsheet represents a pixel in the raster image. This method is also referred to as “exhaustive enumeration.”
2. Run-length raster encoding. This method encodes cell values in runs of similarly valued pixels and can result in a highly compressed image file (fig. 6). The run-length encoding method is useful in situations where large groups of neighbouring pixels have similar values (e.g., discrete datasets such as land use/land cover or habitat suitability) and is less useful where neighbouring pixel values vary widely (e.g., continuous datasets such as elevation or sea-surface temperatures).
3. Quad-tree raster encoding. This method divides a raster into a hierarchy of quadrants that are subdivided based on similarly valued pixels (fig. 7). 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.”

Figure 5 Cell-by-Cell Encoding of Raster Data

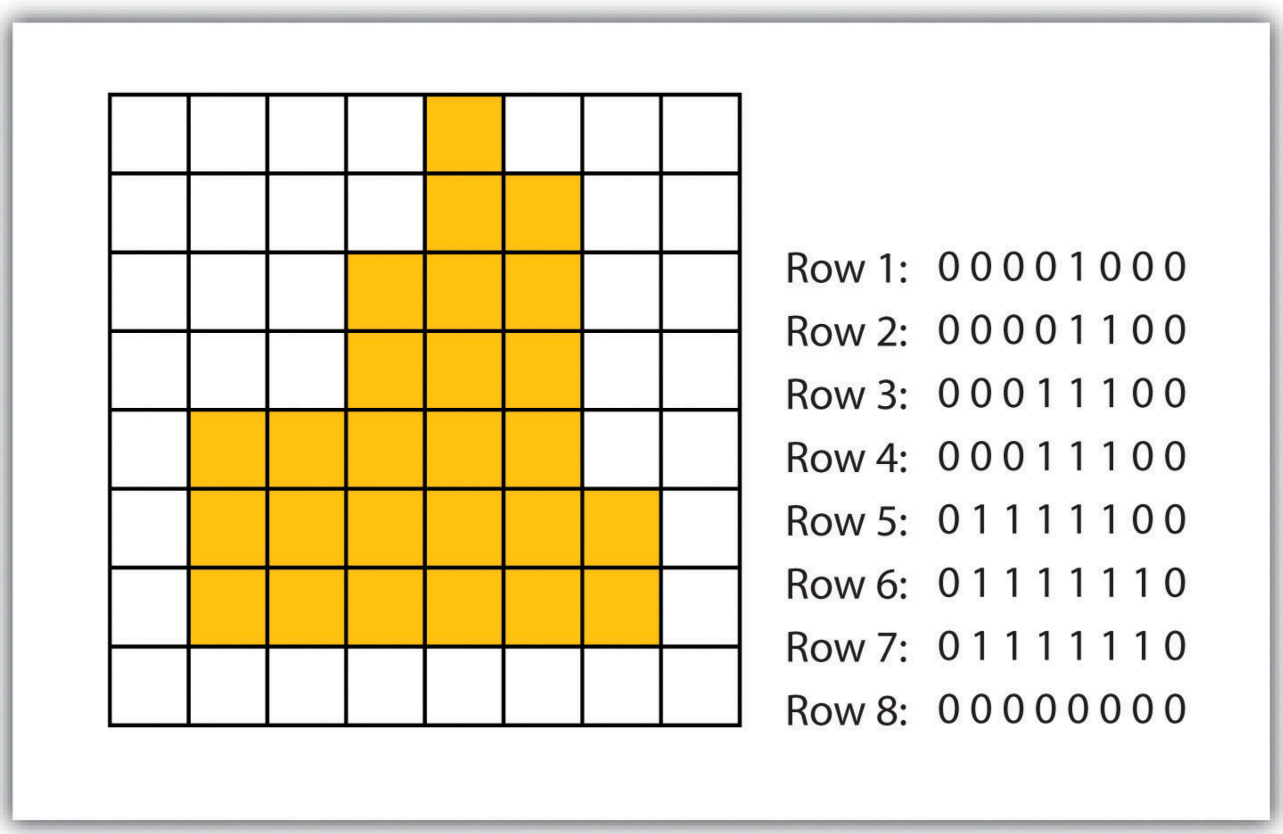


Figure 6 Run-Length Encoding of Raster Data

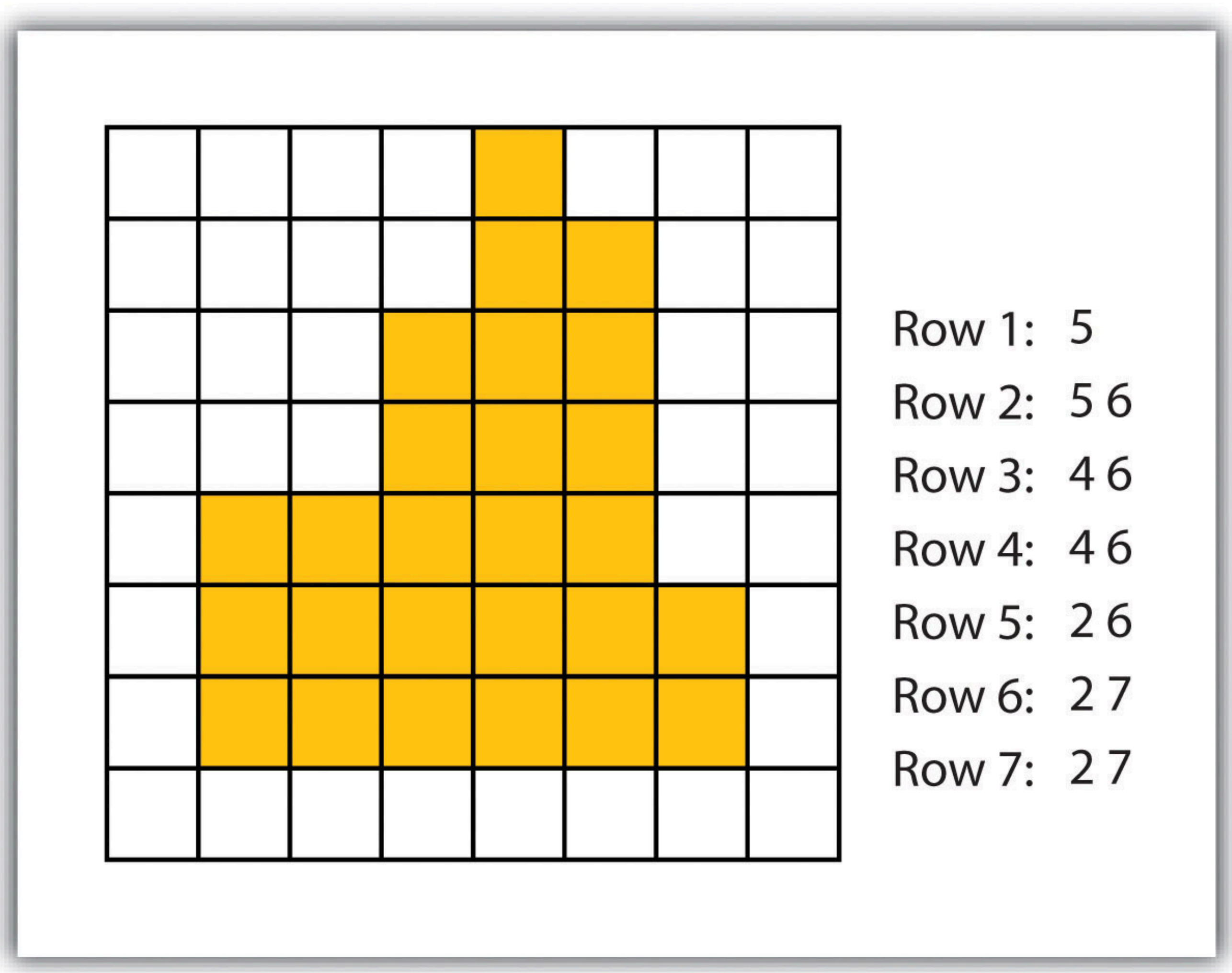
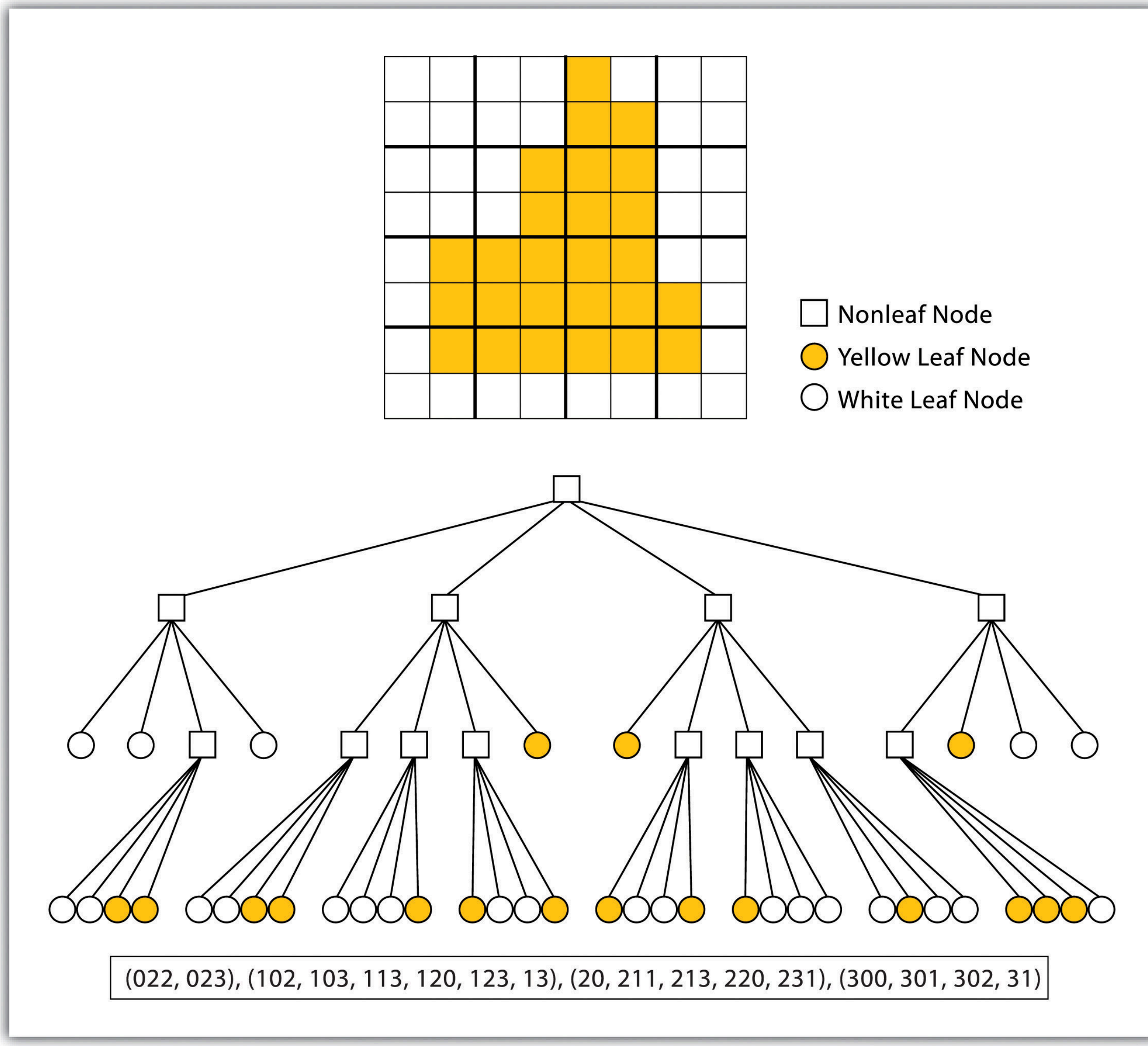


Figure 7 Quad-Tree Encoding of Raster Data



Advantages/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 some sort of raster image generator, namely a digital camera, and few cellular phones are sold today that don't include such functionality. Similarly, a plethora of satellites are constantly beaming up-to-the-minute raster graphics to scientific facilities across the globe. These graphics are often posted online for private and/or 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 represented 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 relatively easy to perform overlay analyses on raster data (for more on overlay analyses, see GIS Analysis chapter). This simplicity also lends itself to easy interpretation and maintenance of the graphics, relative to its vector counterpart.

Despite the advantages, there are also several disadvantages to using the raster data model. The first disadvantage is that raster files are typically very large. Particularly in the case of raster images built from the cell-by-cell encoding methodology, the sheer number of values stored for a given dataset result in potentially enormous files. Any raster file that covers a large area and has somewhat finely resolved pixels will quickly reach hundreds of megabytes in size or more. These large files are only getting larger as the quantity and quality of raster datasets continues to keep pace with 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 fig. 4.1). 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 seemingly randomly coloured grid cells.

The geometric transformations that arise during map reprojection efforts can cause problems for raster graphics and represent a third disadvantage to using the raster data model. These alterations will result in the perfect square pixels of the input layer taking on some alternate rhomboidal dimensions. However, the problem is larger 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).[1]

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

Vector Data Model

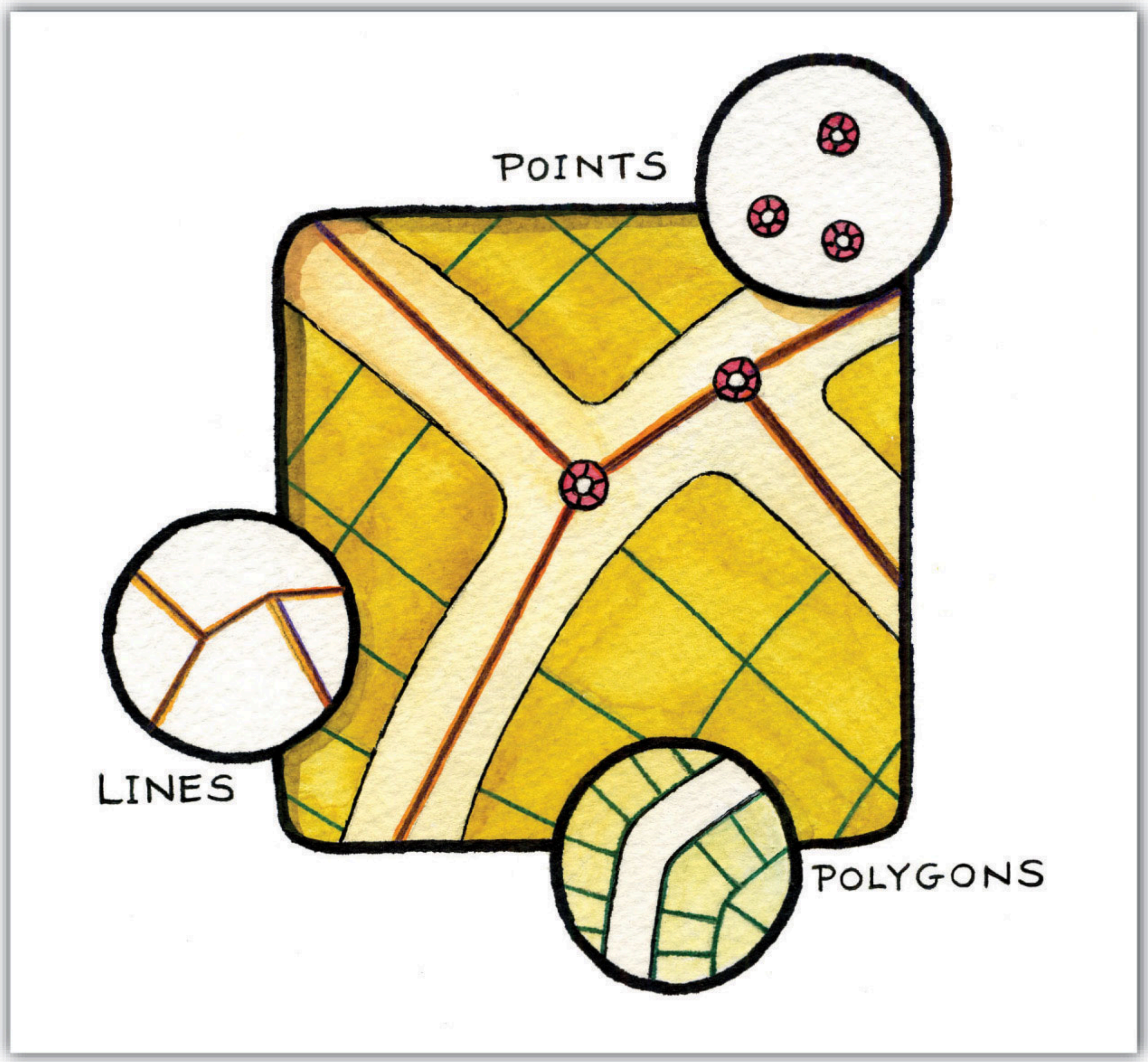
In contrast to the raster data model is the vector data model. In this model, space is not quantized into discrete grid cells like the raster model. Vector data models use points and their associated X, Y coordinate pairs to represent the vertices of spatial features, much as if they were being drawn on a map by hand (Aronoff 1989).[2] 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 that is given to each feature in a map.

Three fundamental vector types exist in geographic information systems (GISs): points, lines, and polygons (fig. 8). 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, and so forth. 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

1. Seong, J. C. 2003. “Modeling the Accuracy of Image Data Reprojection.” *International Journal of Remote Sensing* 24 (11): 2309–21.
2. Aronoff, S. 1989. *Geographic Information Systems: A Management Perspective*. Ottawa, Canada: WDL Publications.

intersecting lines and/or polygons. Vertices are defined as each bend along a line or polygon feature that is not the intersection of lines or polygons.

Figure 8 Points, Lines, and Polygons



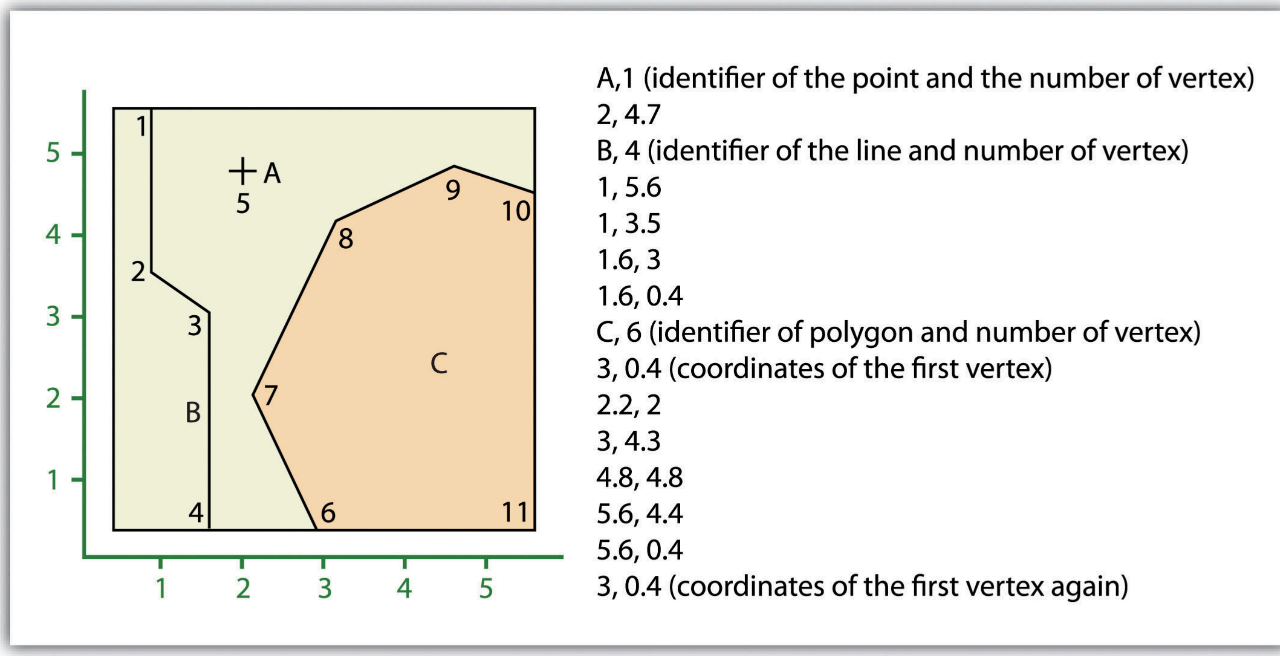
Points can be spatially linked to form more complex features. Lines are one-dimensional features composed of multiple, explicitly connected points. Lines are used to represent linear features such as roads, streams, faults, boundaries, and so forth. Lines have the property of length. Lines that directly 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 are used to represent features such as city boundaries, geologic formations, lakes, soil associations, vegetation communities, and so forth. Polygons have the properties of area and perimeter. Polygons are also called areas.

Vector Data Models Structures

Vector data models can be structured 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).[3] In the spaghetti model, each point, line, and/or 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) with no inherent structure (fig. 9). One could envision each line in this model to be a single strand of spaghetti that is formed into complex shapes by the addition of more and more strands of spaghetti. It is notable that in this model, any polygons that lie adjacent to each other must be made up of their own lines, or stands 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 exact same boundary information. This creates some redundancies within the data model and therefore reduces efficiency.

Figure 9 Spaghetti Data Model



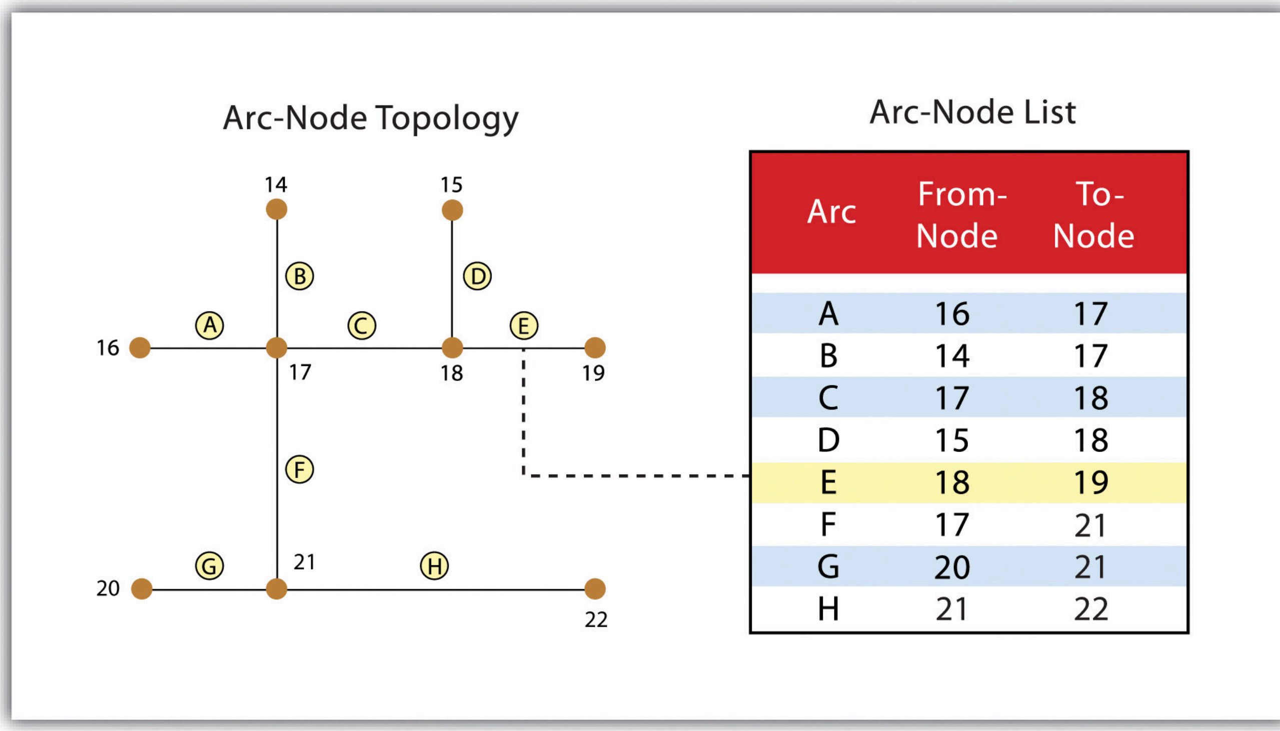
Despite the location designations associated with each line, or strand of spaghetti, spatial relationships are not explicitly encoded within the spaghetti model; rather, they are implied by their location. This results in a lack of topological information, which is problematic if the user attempts to make measurements or analysis. The computational requirements, therefore, are very steep if any advanced analytical techniques are employed on vector files structured thusly. 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.

3. Dangermond, J. 1982. “A Classification of Software Components Commonly Used in Geographic Information Systems.” In *Proceedings of the U.S.-Australia Workshop on the Design and Implementation of Computer-Based Geographic Information Systems*, 70–91. Honolulu, HI.

In contrast to the spaghetti data model, the topological data model is characterized by the inclusion of 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 with 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 that 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 (fig 10). In addition, between each node pair is a line segment, sometimes called a link, which has its own identification number and references both its from-node and to-node. In fig. 10, arcs 1, 2, and 3 all 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 not possible to move from arc 1 to arc 5, as they do not share a common node.

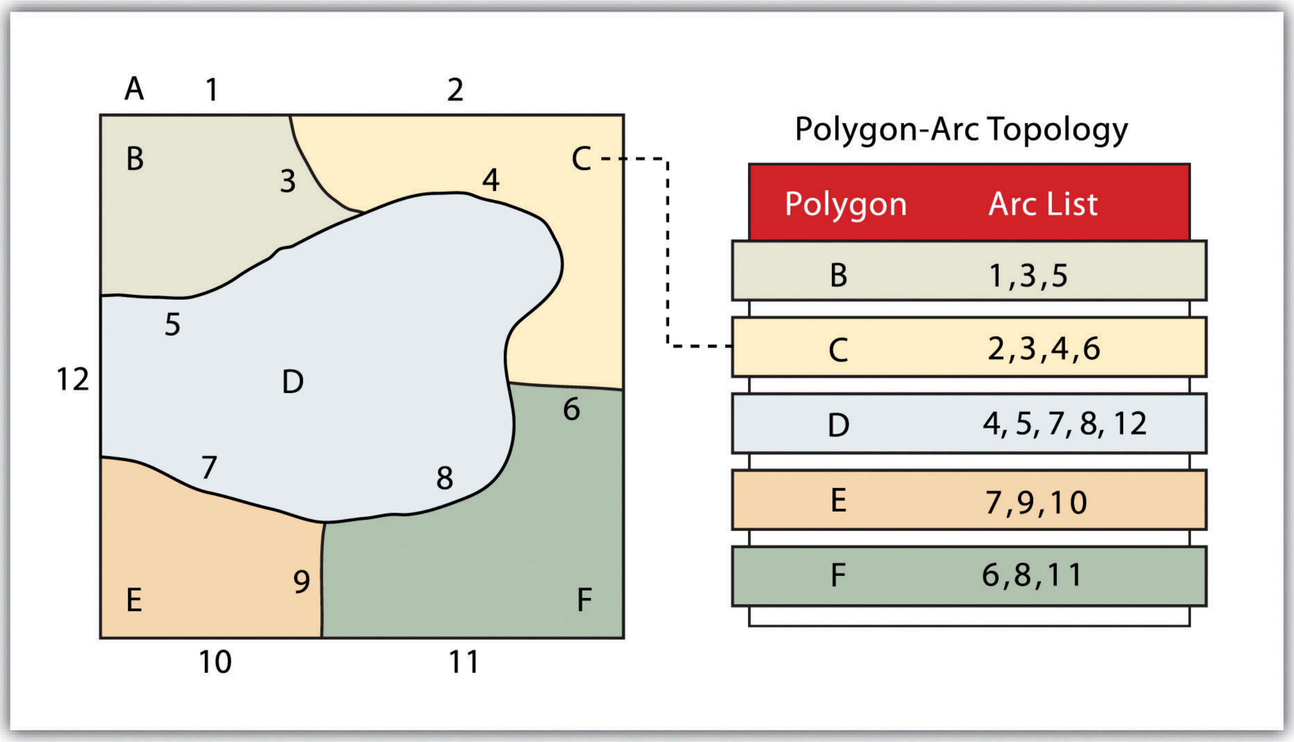
Figure 10 Arc-Node Topology



Arc	From-Node	To-Node
A	16	17
B	14	17
C	17	18
D	15	18
E	18	19
F	17	21
G	20	21
H	21	22

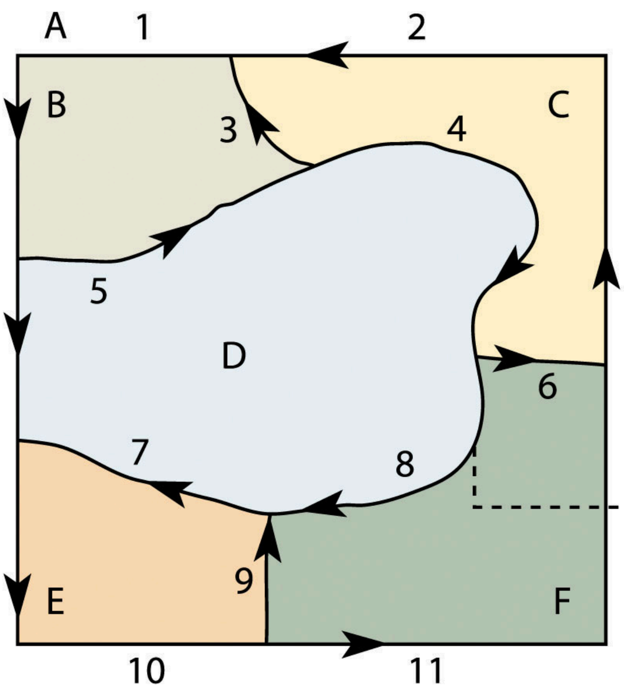
The second basic topological precept is area definition. Area definition states that an arc that connects to surround an area defines a polygon, also called polygon-arc topology. In the case of polygon-arc topology, arcs are used to construct polygons, and each arc is stored only once (fig. 11). This results in a reduction in the amount of data stored and ensures that adjacent polygon boundaries do not overlap. In the fig. 11, the polygon-arc topology makes it clear that polygon F is made up of arcs 8, 9, and 10.

Figure 11 Polygon-Arc Topology



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 (fig. 12). Polygons that share an arc are deemed adjacent, or contiguous, and therefore the "left" and "right" side 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. Fig. 12 shows that arc 6 is bound on the left 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.

Figure 12 Polygon Topology



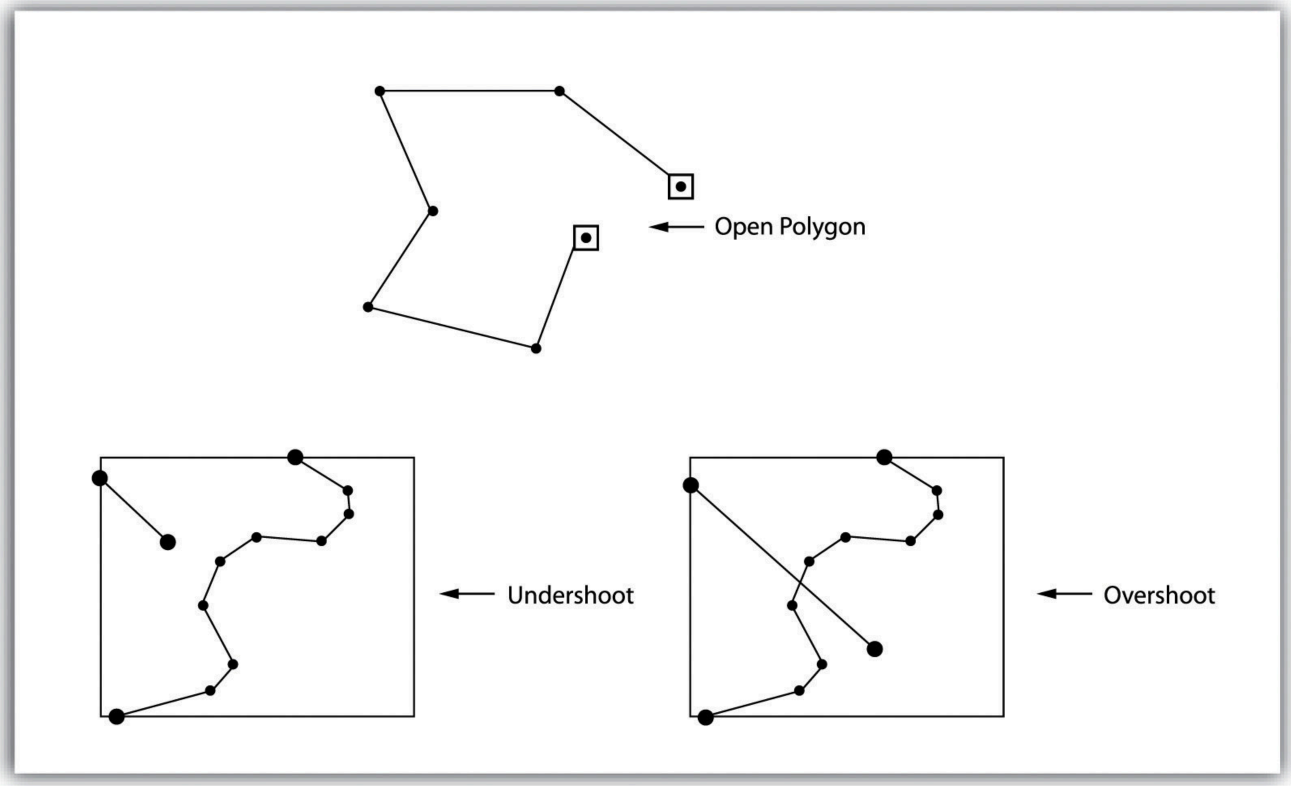
Left-Right Topology

Arc	Left Polygon	Right Polygon
1	A	B
2	C	A
3	B	C
4	C	D
5	B	D
6	C	F
7	E	D
8	F	D
9	E	F
10	E	A

Topology allows the computer to rapidly determine and analyze the spatial relationships of all its included features. In addition, topological information is important 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 do not meet exactly (fig. 13).

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 (fig. 13). The result of overshoots and undershoots is a “dangling node” at the end of the line. Dangling nodes aren’t always an error, however, as they occur in the case of dead-end streets on a road map.

Figure 13 Common Topological Errors



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

Now that the basics of the concepts of topology have been outlined, we can begin to better understand the topological data model. In this model, the node acts as more than just a simple point along a line or polygon. 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/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. 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/or 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 (see fig. 1).

Vector data tend to be more compact in 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 the computer storage space when compared to raster data.

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

Alternatively, there are two primary disadvantages of the vector data model. First, the data structure tends to be much 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 relatively 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, particularly when dealing with large datasets.

17.

GIS and Analysis

18.

Spatial Arrangement

19.

Spatial Statistics: Spatial Autocorrelation

20.

Representing Landscapes and Terrain

21.

Accuracy, Precision, and Uncertainty of Geographic Data

VI

Remote Sensing

22.

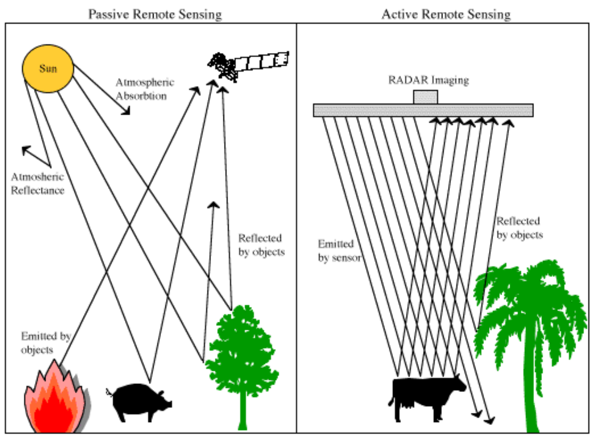
Introduction to Remote Sensing and Elements of Image Interpretation

Introduction

Remote Sensing is the act of perceiving or sensing something without direct contact. In geomatics this act is generally restricted to remotely sensing the surface (or near surface) of Earth from above (by satellites, planes, drones, etc.). While the origins of remote sensing begin with film emulsion systems (cameras) mounted to balloons, and later airplanes, a transition to satellites (as the vehicle to which the camera was mounted) and digital sensing systems (digital "cameras") began in the 1950's.

See: https://www.oneonta.edu/faculty/baumanpr/geosat2/RS%20History%20II/RS-History-Part-2.html

Current remote sensing produces a continuous stream of digital data corresponding to the reflection of solar radiation from Earth's surface. Incoming solar radiation passes through the atmosphere (and some stuff happens to it as it does) and interacts with material at the surface. At the surface the radiation is absorbed, reflected, or scattered. The reflected component of that radiation and its physical consistency is a foundation of the principles of remote sensing.



23.

Resolution and Remote Sensing

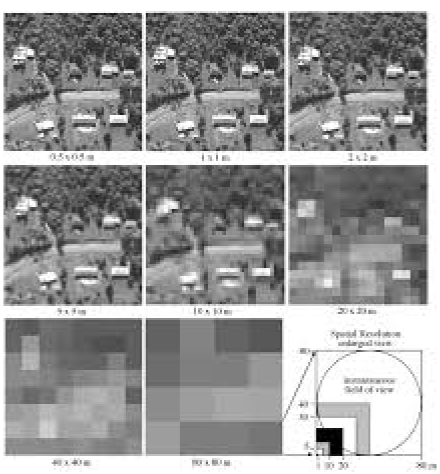
Remote Sensing: Resolution

In remote sensing resolution refers to one's ability to resolve (determine, identify, etc.) what is present in an image. There are four resolution types: spatial, spectral, radiometric, and temporal.

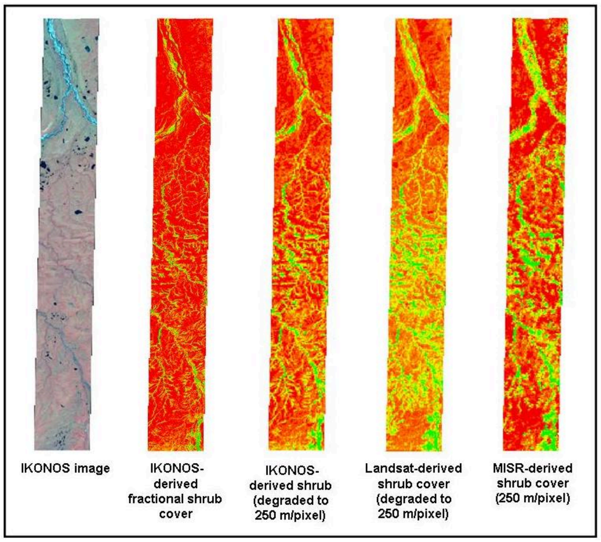
Spatial Resolution

Spatial resolution refers to the smallest item that can be resolved visually or spectrally in an image. The extent to which something (of a certain size) can be resolved is directly related to the pixel size of of the image and sensing system

the multivariate nature of resolution is often conflated to spatial resolution.



Spectral Resolution



Radiometric Resolution

24.

Spectral Signatures

25.

Remote Sensing Platforms

Appendix

This is where you can add appendices or other back matter.

Glossary

Actual Scale (Scale): the scale at any point on a maps surface in relation to the earth (see scale factor and principal scale).

Area of Study (Intro to Geomatics): the geographic extent of the place being studied.

Cartesian coordinates: coordinates indicating location in a space defined by two baselines intersecting one another at right angles. The intersection point is the origin for locations defined by distances along the two intersecting axes.

Choropleth Map (Maps and Mapmaking): A map that assigned to **units of analysis** a colour or fill pattern based on a classification of numeric data. Areas with similar colours have similar values, dissimilar colours on the map represent dissimilar values.

Colour (remote sensing): each distinguishable variation on an image produced by a multitude of combinations of hue, value and chroma.

Conformal: maps that preserve shapes, although continents will appear relatively larger or smaller than one another their shape will be consistent to their appearance on the earth, or reference globe.

Conic projection: The contact line between the plane and the surface of the globe is a small circle (non-great), the same is true for secant conic projections.

Continuous reference system (coordinate systems): lat./long. or UTM identity for any point on earth's surface with infinite accuracy.

Cylindrical projection (map projections): the single contact point (line) between the plane surface and the globe is a great circle, except in the case of a secant cylindrical projection in which case the two lines of intersection between the cylinder and globe are equidistant from a great circle. (see following slides for visual).

Developable Surface (map projections): a surface onto which the graticule of Earth is projected. Specifically, a developable surface is one that must be manipulated before and after projection, such as being wrapped around the reference globe and subsequently unwrapped.

Discrete reference system (coordinate systems: system for referencing discrete units on the earth's surface (Street address, township and range).

Ecological Fallacy (Intro to Geomatics): Drawing conclusions about individuals when only group data is available. If the average income of a group of people is above the national average, it would be an ecological fallacy to conclude that all individuals have above average income.

Equal Area: Areas of shapes (continents, lakes, oceans, countries, etc.) will be consistent with the reference globe and to scale with the earth.

Equidistant (Azimuthal): maps that preserve distances as measured from a central point in any direction. In some cases distances are equal when measured from a circle on the map, if the projection is secant rather than tangent.

Extrapolation (remote sensing): estimating a value that is beyond the range of sampled values (either in a spatial or non-spatial range)

Geocoding: the process of assigning a continuous location reference (global frame of reference) to a discrete object or address.

Geodesy: The science of the earth's shape and size; also related to measuring the variable gravitational force exerted by the earth. Models of the earth's size and shape are used to produce datums upon which frames of reference and mapping systems are based.

Geodetic datum: a definition of the size, shape and orientation of the earth; datums support the development of global reference and mapping systems.

Geographic Information Science: The science related to examining and studying the technologies associated with collecting, storing and processing spatial data. The primary purpose is to study and expand the technology.

Geographic Information System: the integrated hardware and software used to collect, store and process spatial data.

Geoid: a model of the earth's surface (shape) based on the variable gravitational force exerted by the earth as an object.

Georeference: The act of coordinating collected data to a common global reference system, other related topics include geo-coding and address matching.

Gnomic: Map projection with a theoretical light source at the earth's center.

Graticule: the pattern of intersecting lines (latitude and longitude) on Earth, can also be applied to any spherical, ellipsoidal or geoidal object.

Grid: the pattern of intersecting lines on a flat (2-dimensional) surface.

Image (remote sensing): A reproduction or imitation of the form of a person or thing. The optical counterpart of an object produced by a lens or mirror or other optical system.

Interpolation (remote sensing): estimating a value that is within the range of sampled values, but not one of the sampled values (either in a spatial or non-spatial range)

Interval (levels of information): data is given a value that is based on a fixed scale with no true zero value. The interval between observations is constant, but is not related to an absolute zero (measures of temperature other than Kelvin). Interval data may appear as though it can have complex operation performed on it, such as performing a ratio, multiplying, etc., but this is not true (i.e. 10 degrees C is not twice as warm as 5 degrees C).

Line of tangency: Projection surfaces can come into contact with the globe's surface or exist some distance away from the surface. When a developable surface comes into contact with the globe's surface along a single line the projection is considered tangent (the projection surface is tangent to the globe along that line). When a developable surface comes into contact with the globe along two lines the projection is considered secant (the surface "cuts through" the surface of the globe). In the case of a planar projection (non-developed) the projection surface comes into contact with globe at a point for a tangent projection and a small circle for a secant projection. (see following slides for examples)

Nominal: data is put into classes with distinct labels; classes have no relationship to one another and are given no relative value (eye color is an example).

Ordinal: data can be placed in ranked categories, categories are differentiated by their position in the ranked scale, values can be greater or less than one another but the differences cannot be measured (A is taller than B, 1st, 2nd, 3rd, 4th, etc.).

Orthographic: a view of the world from a vertical perspective, as though one were looking at the entire earth from directly above the entire surface at once. Orthographic projections are computed with a theoretical light source placed infinitely far away from the earth.

Pattern (remote sensing): Pattern is the spatial arrangement of objects and is a macro image characteristic.

Photo Interpretation (remote sensing): The act of examining aerial photographs/images for the purpose of identifying objects and judging their significance.

Photogrammetry (remote sensing): The science or art of obtaining reliable measurements by means of photography.

Photography (remote sensing): The art or process of producing images on a sensitized surface by the action of light or other radiant energy.

Planar Projection: the plane onto which the globe's surface is "projected" is not altered following or preceding projection. The plane is tangent to the globe at a point or along a small circle in the case of a secant projection.

Polar Coordinates: coordinates indicating location in a space defined by a single baseline. One end of the baseline is defined as the origin and locations are indicated by a distance (in some measurement system) away from the origin and an angular deviation around the origin away from the baseline.

Principal Scale: (scale of globe radius to actual earth radius) scale of the model of the earth being used for a map to the earth. Principal scale and actual scale will be the same all over the reference globe.

Public Particpations GIS (PPGIS): studies related to the societal context of Geographic Information Systems.

Query: question asked of a database.

Ratio (levels of information): data exists on a scale with a constant interval between values and has an absolute 0. Ratio data can be multiplied by a constant, and have a ratio applied to observations, supporting complex analytical operations.

Resolution (remote sensing): Resolution is defined as the ability of the entire photographic system, including lens, exposure, processing, and other factors, to render a sharply defined image. An object or feature must be resolved to be detected and/or identified.

Scale Factor: Actual scale divided by the Principal scale; SF will therefore be 1.0 everywhere on the reference globe. On the projected map surface the SF may be higher or lower than 1.0 due to distortions associated with transforming the reference globe into a flat map.

Site (remote sensing)- How objects are arranged with respect to one another; or with respect to various terrain features.

Spatial Autocorrelation: A measure of the spatial relationship between the occurrence of an attribute in different locations.

Stereographic: Map projection with a theoretical light source at the antipode (opposite side of the earth's surface being mapped).

Texture (remote sensing): The frequency of change and arrangement of tones. Texture is a micro image characteristic and is related to the visual impression of smoothness or roughness of an area.

Tone (remote sensing): Tone can be defined as each distinguishable variation from white to black.

Unit of Analysis (Intro to Geomatics): the individual units that are used to create an aggregate pattern for an **area of study**. The unit of analysis might be a single observation (i.e. household income) or a summary of observations (average income for a neighbourhood). Units of analysis are studied or presented together so that a pattern across a larger spatial extent can be communicated.