An Open Geomatics Textbook

UBC

2022-01-06

Contents

Preface

This is the very first part of the book, which will eventually include the textbook's introduction. For now, here's some useful info for you:

0.1 Contacts

Paul Pickell, paul.pickell@ubc.ca
Evan Thornberry, evan.thornberry@ubc.ca
Francois du Toit, fdutoit@mail.ubc.ca

0.2 Project Wiki

github.com/ubc-geomatics-textbook/docs/wiki

0.3 Style Guide

0.3.1 Audience

1. Audience is undergraduate of graduate student studying GIS, geomatics, and remote sensing with no prior knowledge in these subject areas (i.e., introductory).
2. Assume only first year-level knowledge (or equivalent concurrent learning) of mathematics, science (biology, chemistry, physics), and geography.
3. Assume a multicultural reader who is not necessarily familiar with Canadian geography and history.

0.3.2 General Style

1. Word spellings should follow *The Oxford Canadian Dictionary (2 ed.)*.
2. Every chapter begins with 1-3 paragraphs of introductory text. The introductory text should be for general interest and not introduce any important terms that will be defined later in the chapter. The last sentence of this introductory text should summarize what students will learn.

3. Posing questions to readers is encouraged in all sections. For example, "Have you ever wondered…?" "How do you think X relates to Y?"
4. At every opportunity, authors should highlight Canadian examples of technology and science in geomatics. Examples of geomatics applications are highly encouraged in the Canadian context. For example, the following list of environmental management problems that are important to Canada should be discussed whenever possible:
   - Northern communities
   - First Nations
   - Climate change
   - Boreal forest
   - Endangered wildlife
   - Freshwater management and ecosystems
   - Fisheries
   - Glaciers/ice monitoring
   - Environmental justice
   - Resource extraction

0.3.3 Learning Objectives

1. Every chapter will have a numbered list of learning objectives that follow the introductory text.
2. There should be no period at the end of each listed learning objective.

0.3.4 Summary

1. All learning objectives should be addressed in the summary section.
2. The summary section should never introduce any new concepts, terms, or definitions and should never reference figures, tables, or equations.

0.3.5 Key Terms

1. Every chapter will have an alphabetical, but unnumbered list of key terms.
2. At first mention in the chapter text, key terms should be boldened and defined.

0.3.6 Headings and Labels

1. Chapter titles should use title-case and are numbered.
2. Chapter sub-titles are also numbered and in title-case. Sub-titles should go no lower than level 3 heading (i.e., 1.2.3).
3. Level 4 headings are not numbered, all letters are capitalized, and should only be used in special call-out boxes:
   - LEARNING OBJECTIVES
   - REMEMBER THIS?
   - YOUR TURN!

- CASE STUDY

0.3.7 Formulae

1. Do not format formulae using Microsoft Word or LaTeX. Instead, formulae should be formatted with RMarkdown.
2. Coordinates and Greek letters should always be formatted as formulae with RMarkdown.

0.3.8 Units

1. Standard International (SI) units should be used for the following:
   - Length = meter (m)
   - Time = second (s)
   - Amount of substance = mole (mole)
   - Electric current = ampere (A)
   - Temperature = Kelvin (K)
   - Luminous intensity = candela (cd)
   - Mass = Kilogram (kg)
2. Angle degrees are preferred over radians (rad) when referencing geographic position.
3. Rates should be expressed with a dot operator and negative exponent rather than a divisor (e.g., m · s-1 or W · m-2).

0.3.9 Numbers

1. Scientific notation is the preferred way to represent large and small numbers and should use the × operator (not dot or asterisk) and be formatted as a formula (see Formulae): 1 × 102.
2. Scientific notation should be limited to four significant figures (e.g., 1.234 × 100) except for specific numbers where the precision is important or meaningful like the speed of light (2.99792458 × 108 m · s-1) or Planck's constant (6.62607004 × 10-34 J · s-1).
3. Constants (like above) and other physical variables should use common notations (e.g., c for speed of light and h for Planck's constant) and be formatted as formulae (see Formulae).

0.3.10 Dates and times

1. The Gregorian calendar should be adopted for recent dates. In these cases, use Common Era (C.E.) to indicate dates after 0 A.D. and Before Common Era (B.C.E.) for dates before 0 A.D.
   - For specific recent dates, use the format "20 February 2021" and omit C.E.
   - If many dates need to be summarized in a table, use the format "DD-MM-YYYY"

2. Times should be specified in either Local Standard Time (LST) or Coordinated Universal Time (UTC) using a 24-hour clock:
   - 00:00 = 12 A.M. midnight LST
   - 12:00 = 12 P.M. noon LST
   - 23:00 = 11 P.M. LST
3. For non-recent dates or when referring to geologic time scales, use the following:
   - Thousands of years before present = kilo annum (ka)
   - Millions of years before present = mega annum (Ma)
   - Billions of years = giga annum (Ga)

0.3.11 Tables

1. Tables are numbered in the order that they appear in text and begin with the number of the chapter:
   - Table 1 in Chapter 1 = 1.1
2. A short, descriptive caption should be written for a table.
3. Tables should only include information that is discussed or referenced in the chapter text.
4. Every table must be referenced in the chapter text.

0.3.12 Code blocks

1. Avoid code blocks in chapter text. Instead, try to place code blocks in TRY THIS! or CASE STUDY sections.
2. Only R code blocks should be embedded using RMarkdown.

0.3.13 Abbreviations

1. Abbreviations are shortened form of a word or phrase and should be punctuated with periods:
   - e.g.
   - Dr.
   - Ph.D.

0.3.14 Initialisms

1. Initialisms are the first letters of several words and should always be defined at first use in the chapter text regardless if the initialism is introduced and defined in an earlier chapter.
2. Do not introduce initialisms in figure or table captions or table text.
3. Except for the specific cases in this style guide, do not punctuate initialisms with periods:
   - AVHRR
   - NDVI

0.3.15 Acronyms

1. Acronyms are combinations of the first letters of several words and are pronounced as words. Acronyms should never be punctuated with periods.
2. Many satellites and remote sensing systems have acronyms that vary capitalization.
3. Following are some preferred acronyms:
   - Light Detection and Ranging = LiDAR
   - Radio Detection and Ranging = RADAR
   - Moderate Resolution Imaging Spectroradiometer = MODIS

0.3.16 Punctuation

1. Use serial comma (Oxford comma) in lists: Yukon, Northwest Territories, and Nunavut.
2. Use italics for internal dialogue or when you infer what the reader might be thinking:
   - "At this point, you might be wondering, *why am I reading this sentence?*"
3. Avoid the use of semi-colons.
4. Use and punctuate common Latin abbreviations with periods:
   - "For example" = exempli gratia (e.g.)
   - "That is" = is est (i.e.)
   - "And other similar things" = et cetera (etc.)
5. Avoid phrases in parentheses () or brackets []. Instead, place the phrase in a proper sentence.
6. Use single spaces between sentences.
7. Use double quotation marks for direct quotes, but avoid reproducing verbatim large texts. Paraphrasing with proper citation is preferred to direct quotation.
8. Bullet points are preferred over long lists in sentences.

0.3.17 Citations

1. Style should follow American Psychological Association (APA) format.
2. In-text references are encouraged where necessary, especially in case studies.
3. References and Recommended Readings section is placed at end of each chapter. Where possible, Recommended Readings should be populated with Open Educational Resources.

Chapter 1

What is Geomatics?

Chapter 2

Mapping Data

You probably already accept that the Earth is "round" and not "flat". You have probably held and touched a globe at some point in your life. But have you ever wondered how we describe location and measure something as large as the Earth? In this chapter, we will explore fundamental concepts for how we measure the Earth and orient ourselves with coordinate systems.

Learning Objectives

1. Understand the models of Earth's figure and shape
2. Describe different vertical datums and how they are used to reference height
3. Understand the difference between cartesian, celestial, geographic, and projected coordinate systems
4. Recognize the differences among major types of map projections
5. Explore how projected coordinate systems distort and represent the world around us

Key Terms

Antipode, Great Circle, Small Circle, Geodesy, Vertical Datum, Horizontal Datum, Deflection of the Vertical, Ellipsoid, Spheroid, Geoid, Elevation, Orthometric Height, Geoid Height, Geodetic Height, Coordinate System, Celestial Coordinate System, Cartesian Coordinate System, Geographic Coordinate System, Projected Coordinate System, Map Projection, Tissot's Indicatrix

2.1 Introduction to Geodesy

Geodesy is the fascinating science of measuring the shape, orientation, and gravity of Earth. Naturally, some of the questions that come to mind when thinking about such a grand topic are *I thought the shape of Earth is a sphere?* and *How do we orient ourselves on Earth?* and *What does gravity have to do with mapping location?*

All of these questions stem from need to represent **location**. For our purposes, location is the position of something relative to something else. In order to actually describe a location on Earth, we first need to know the size and shape of Earth. Some of the first estimations of Earth's size and shape were made by Eratosthenes, a Greek mathematician from the second and third centuries B.C. Eratosthenes was responsible for many concepts we use in our everyday lives:

- Conceiving the first spherical model of Earth
- The first accurate measure of Earth's circumference
- Calculating the tilt of Earth's axis
- Calculating the distance of Earth to the Sun
- Invention of the leap day

Eratosthenes accurately calculated the circumference of Earth by noticing how the Sun shone directly down the bottom of a well in Syene (modern Egypt) at noon on the summer solstice. He later made a second observation at Alexandria at noon on the summer solstice with a pole and noticed a shadow. He measured the angle of the shadow and inferred the circumference of Earth, which was already known to be spherical (Figure **??**).

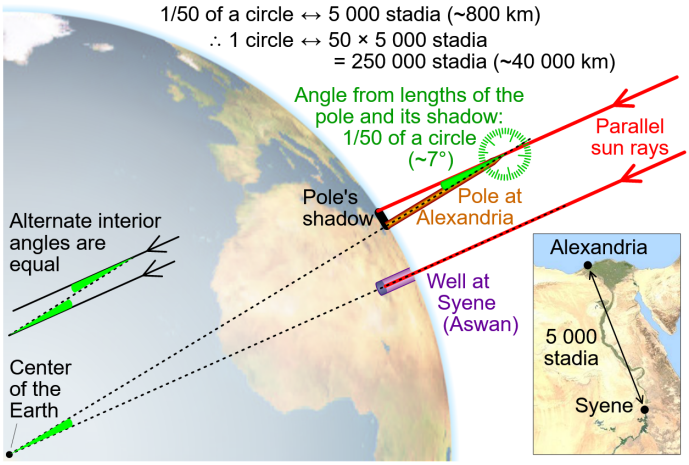


Figure 2.1: Diagram showing how Eratosthenes estimated the circumference of Earth by observing the angle of a shadow that was cast about 800 km north of Syene in present-day Egypt. David Monniaux, CC-BY-SA-4.0.

Pretty simple, right? Turns out, Eratosthenes was off by only 75 km or less than 0.2% in his calculation! The actual North-South circumference of Earth is about 40,075 km. His calculation worked because the Sun's rays are nearly parallel when they strike Earth. So if you observe the Sun at the same time in two locations on Earth on the North-South axis, you will notice the Sun has a different elevation above the horizon, which means different lengths of shadows will be cast on the ground. This is also a way to prove that the Earth is in fact round because a flat Earth would have equally-sized shadows everywhere at any given time of day.

2.2 Models of Earth

Here is a simple thought experiment to consider. Suppose you are trying to measure your own height. You probably have not given much thought about how to technically do this because it seems intuitive: place a measuring tape at the bottom of your feet and mark the measurement at the top of your head. If we break this down, there are some important rules to follow (Figure **??**):

1. The measuring tape must originate somewhere. In other words, we need to define a reference point or surface of zero height (i.e., the ground).
2. The measuring tape must be a straight line and originate at a 90-degree angle, perpendicular to the ground.
3. The measurement must terminate at a point along an imaginary line that is tangential to your head, and yes, that line must be perpendicular to the measuring tape and also parallel with the ground.

Whenever you measure your height, the ground is easy to define. It is whatever point you are standing on. This starting point it also known as a **datum**. A datum is simply a reference point, set of points, or a surface from which distances can be measured. It does not matter if you are below sea level, atop Mount Everest, or on the 30th floor of a skyscraper. You will always get an accurate and repeatable measure of your height using a datum that is defined directly below your feet. But what about measuring the height of terrain on Earth? Whenever we measure the height of Earth's terrain above some reference surface, we are measuring **elevation**.

The same rules above apply when we measure elevation. In order for elevation measurements to be comparable across the world, we need to define a reference surface, a datum, for the entire planet. There are actually several ways that we can model the shape of Earth in order to produce a datum. Models of Earth's shape are often referred to as either vertical datums (the plural of datum) if you are referencing elevation or horizontal datums if you are referencing location. A **vertical datum** is a 3D surface model that is used to reference heights or elevations for the Earth. A simple question like *How high is Mount Logan in Yukon, Canada?* is complicated by the need for a reference surface and the fact that Earth's shape is irregular. In this section, we will review three types of

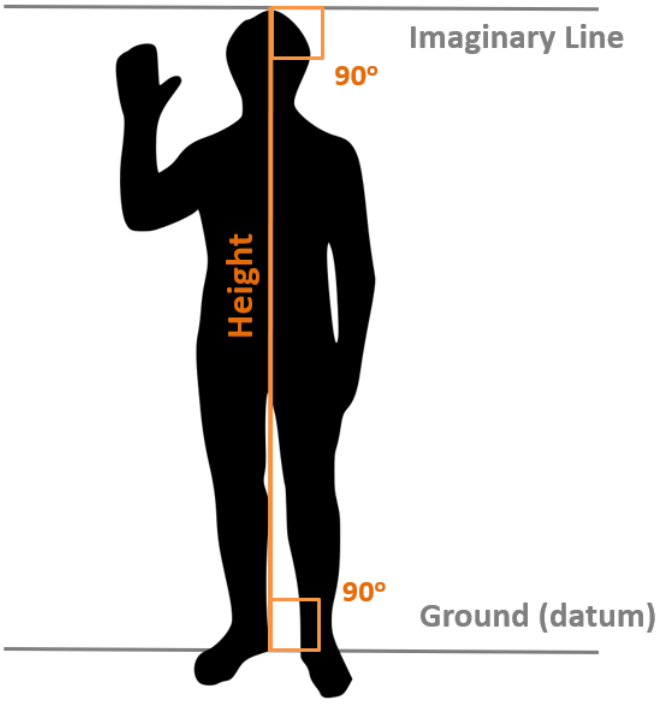


Figure 2.2: Diagram for measuring height above a datum. Pickell, CC-BY-SA-4.0.

vertical datums:

- Geodetic - based on geometry
- Tidal - based on sea level
- Gravimetric - based on gravity

2.3 Geodetic Vertical Datums

A **geodetic vertical datum** is one that describes the Earth's shape in the simplest possible terms using standard geometry. Despite what a globe might lead you to believe, the Earth is not perfectly spherical, but it is close to being spherical. In fact, the radius of Earth varies by no more than 22 km or 0.35%, hardly anything you would ever notice if you were holding it in your hand. That small difference is, however, significant enough to lead to mapping inaccuracies at the local level if a spherical model of Earth was adopted (Figure **??**). Instead, we frequently describe Earth's shape as an oblate ellipsoid, which is essentially a sphere that has been flattened, and we define this ellipsoid with a semimajor and semiminor axis. Sometimes you will see the term *spheroid* used, which just means "sphere-like" and is interchangeable with the term *ellipsoid*.

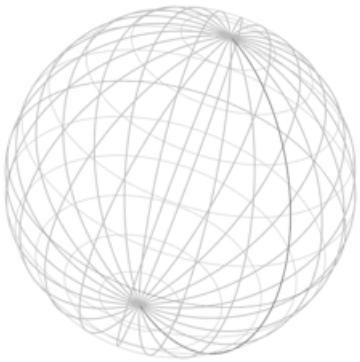


Figure 2.3: Spherical geodetic datum. Pickell, CC-BY-SA-4.0.

There are many different ellipsoids that have been defined and are currently in use as datums. The most commonly used ellipsoid is called the World Geodetic System of 1984 or usually abbreviated as WGS 1984 or WGS 84. In fact, there are hundreds of ellipsoids that have been defined over recent centuries to model the shape of the Earth. The reason for so many other ellipsoids is due in part to technological advances that have improved the accuracy and precision of surveying as well as estimation of the ellipsoidal parameters. Many of these ellipsoids are not **geocentric**, that is, not originating from the center of mass of Earth. These datums are known as **regional datums**, which still describe the dimensions that approximate the shape of Earth, but are instead oriented so that the surface of the ellipsoid is congruent with a particular regional surface

of Earth. For example, the European Datum 1950, the South American Datum 1969, the North American Datum 1983, and the Australian Geodetic Datum 1966 conform well to their respective continents, even better than WGS 1984 in most cases, but poorly anywhere else in the world.

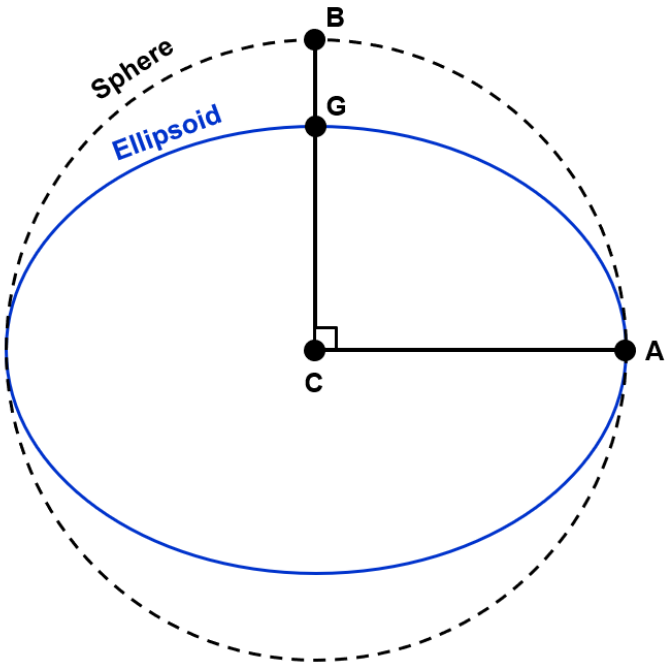


Figure 2.4: Sphere versus ellispoid. Pickell, CC-BY-SA-4.0.

Figure **??** greatly exaggerates the flattening of the ellipsoid to illustrate the above points. In reality, the sphere is flattened using a flattening factor calculated as $f = (CA - CG)/CA$ and defined exactly as $f = 298.257223560$ for WGS 1984. Thus, the semiminor axis (i.e., rotational axis) for the WGS 1984 ellipsoid (meters) is

$$CG = CA - (CA \times \frac{1}{f}) = 6378137 - (6378137 \times \frac{1}{298.257223560}) = 6356752.3$$

where G is the North Pole and A is a point on the Equator. The sphere, of course, is much simpler where $radius = CB = CA = 6378137$.

2.4 Tidal Vertical Datums

A **tidal vertical datum** is likely one that you are familiar with. The premise of a tidal vertical datum is to use mean sea level as a reference surface, above

which are positive elevations and below are negative elevations. This has a lot of advantages, like it is intuitive and oceans cover more than 70% of the planet's surface so much of Earth's land mass is near an ocean. However, the disadvantages are that sea level changes over time with tides and also with climate change. The not-so-obvious problem with a tidal vertical datum is that the sea level is actually not constant around the planet not only due to tides, but also temperature, air pressure, and gravity. In other words, mean sea level measured at a gauge station in Halifax on the Atlantic Ocean will not be the same distance from the center of Earth as mean sea level measured at Victoria on the Pacific Ocean (Figure **??**). The primary challenge with a tidal vertical datum is extending it away from the coastline through a network of survey points using a process known as levelling, and even still, it is only meaningful during the epoch in which the mean sea level was measured at a number of tidal gauge stations.

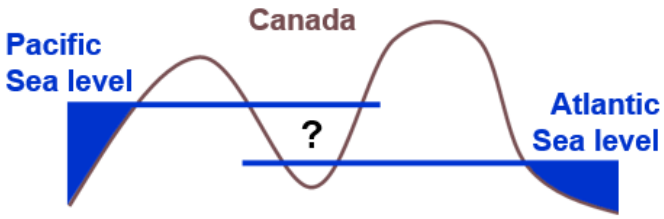


Figure 2.5: Conceptual tidal datum for Canada. Pickell, CC-BY-SA-4.0.

2.5 Gravimetric Vertical Datums

The **geoid** is a physical approximation of the figure of Earth. The shape represents Earth's surface with calmed oceans in the absence of other influences such as winds and tides. It is computed using gravity measurements of Earth's surface and is best thought of as the surface or shape that the oceans would take under the influence of Earth's gravity and rotation alone. In other words, the geoid represents the shape Earth would take if the oceans covered the entire planet. More specifically, the geoid is a **gravimetric** model of Earth's shape that is defined as an equipotential surface from a constant gravity potential value. Due to the distribution of mass on Earth, gravity is not constant across the planet's surface. As a result, the surface of Earth's oceans is not smooth like a sphere, but instead undulates depending on where gravity forces water to remain at rest. You can think of Earth's gravitational field as a series of parallel lines extending outwards from the center of mass of Earth into space. Any of these lines that you choose is an equipotential surface where the force of gravity is constant. Keep in mind that the force of gravity is stronger nearer the center of mass of Earth and weaker as you move away from it. Thus, the geoid is an arbitrary equipotential gravity surface that is chosen to roughly coincide with

present-day mean sea level.

When you measure the height of something relative to a gravimetric vertical datum like the geoid, you must level your instrument. Levelling forms a vertical line that is orthogonal or perpendicular to the geoid, known as a **plumb line**. It is incredibly easy to visualize a plumb line. Simply tie a rock to the end of a string and hold the string with your outstretched arm. The length of the straightened string traces a plumb line to the center of mass of Earth, wherever you are. Because gravity changes with location on Earth and all plumb lines are converging on a singular point, plumb lines are never parallel. This phenomenon has important implications for comparing observations on the the ground with a geodetic model of Earth like an ellipsoid. In other words, the plumb line that you traced with your string is pointing to the center of mass of the geoid, but the center of the ellipsoid is often in a slightly different direction. This difference is known as the **deflection of the vertical** and is measured as the angular difference between the centre of the geoid and the centre of a reference ellipsoid. Like other measurements of geodetic location (i.e., latitude and longitude), the deflection of the vertical is comprised of two angles: ξ (xi) representing the north-south angular difference and η (eta) representing the east-west angular difference.

It should be evident by now that the reference surface that you choose as a vertical datum will determine the measured elevation of Earth's terrain. Additionally, We frequently need to convert elevations between geodetic and gravimetric vertical datums. For example, when you use a Global Navigation Satellite System receiver, you are provided with an elevation that is relative to the WGS 1984 ellipsoid (more on that in Chapter 4). The difference in height between an ellipsoid and the geoid is referred to as **geoid height (N)** while the difference in height between an ellipsoid and Earth's surface is referred to as **geodetic or ellipsoidal height (h)**. The difference in height between the geoid and the Earth's surface is called **orthometric height (H)** (Figure **??**), and is given as:

$$H = h - N$$

To illustrate the concept of a gravimetric datum, suppose we constructed a large, straight tunnel through the physical Earth that was tangential to the ellipsoid. If we allowed the oceans to flow freely through this tunnel, your experiences might convince you that water would flow from one end to the other. But in fact, this tunnel is so large, that the gravity field is changing. So the water would actually come to rest at the surface of the geoid or gravimetric model, as shown in Figure **??** below.

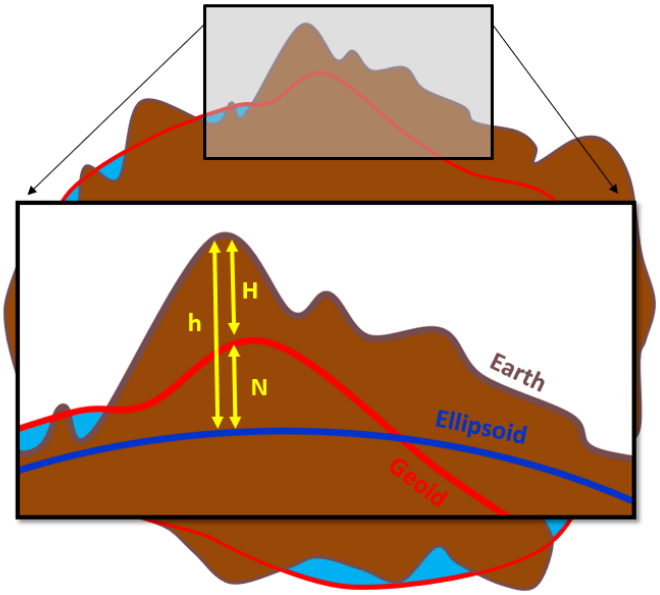


Figure 2.6: Orthometric Height (H) is the ellipsoidal height (h) less the geoid height (N). Pickell, CC-BY-SA-4.0.

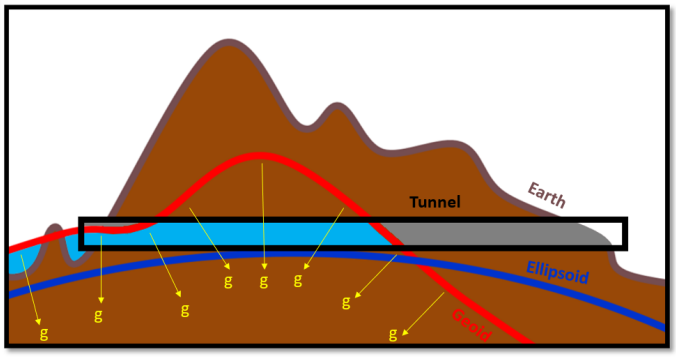


Figure 2.7: Thought experiment showing where water would be at rest within a tunnel through the geoid due to the equipotential force of gravity (g).Pickell, CC-BY-SA-4.0.

2.6 Case Study: The Canadian Geodetic Vertical Datum of 2013

The Canadian Geodetic Vertical Datum of 2013 (CGVD2013) is the current gravimetric vertical datum used in Canada to reference heights. It is defined with a potential gravity value of 62,636,856.0 m^2 $s^{-}2$. The pervious vertical datum in Canada - the Canadian Geodetic Vertical Datum of 1928 (CGVD28) - was actually a tidal vertical datum that corresponded to mean sea level measured at Yarmouth, Halifax, Pointe-au-Père, Vancouver and Prince-Rupert, and a height in Rouses Point in New York. It turns out that Halifax referenced to CGVD2013 is 64 centimeters *below* Halifax referenced to CGVD28!

For reference, CGVD2013 is 17 centimeters below mean sea level measured in Vancouver at the Pacific Ocean, 39 centimeters above mean sea level in Halifax at the Atlantic Ocean, and 36 centimeters above mean sea level in Tuktoyaktuk at the Arctic Ocean. The older CGDV28 did not have any survey benchmarks in the far north of Canada and, with the advent of more reliable satellite-based measurements, was modernized in 2015 to CGVD2013. The United States currently uses the North American Vertical Datum of 1988 (NAVD88), which was never adopted by Canada, but the United States will be modernizing their vertical datum by adopting a gravimetric model with the same gravity potential value as Canada as early as 2025.

2.7 Referencing Location

2.8 Cartesian Coordinate Systems

Now that we have explored how to reference heights to vertical datums, we will now turn to considering how to reference location on Earth. Before we we jump to the three dimensional case of Earth, consider how you would map your room and identify your location within that room. Assuming you are in a rectangular room, you could easily pick a corner and first measure the distances between the four corners of the room, giving you the dimensions. You could then proceed to measure your distance to any two walls and quite easily define your position within the room relative to the first corner that you picked. This is an example of a **coordinate system** that provides reference for the relative locations of anything contained within the extent of the coordinate system (i.e., the four corners of your room).

Fundamentally, a coordinate system is defined by a common unit of measurement (e.g., meters, feets, degrees), an orientation defining the direction that measurements positive or negative, and an **origin**, which is an arbitrary point where the measurements begin at zero. On a one dimensional line, you can define any other location on the line as a measured distance from the origin at [0]. On two dimensional maps, like your room, we have two axes that are per-

pendicular from which we can define any location with measured distances from the origin $[0, 0]$. We can also extend this to the three dimensional Earth, which simply requires another axis to define the origin at $[0, 0, 0]$. All of these cases are referred to as **Cartesian coordinate systems**, so-named after French philosopher Descartes (also known for the phrase, "I think, therefore I am") who first described the two dimensional case in 1637 in *La Géometrie*.

2.9 Celestial Coordinate Systems

You might be wondering, *How do we reference locations on Earth?* Before the technological era of geoids, astronomical observations provided the basis for the earliest coordinate systems. The ancient civilizations of Greece, Egypt, and Babylon all recognized the use of celestial coordinate systems for defining the location of stars, planets, and other celestial bodies in the sky. Recall that Eratosthenes, a Greek mathematician living about 2300 years ago, had already worked out the spherical shape of Earth. A celestial coordinate system simply extends the spherical shape of the Earth outward into space to locate objects using angular measurements. It is a *geocentric* coordinate system, that is based on an origin at the center of Earth with an orientation following the rotational axis of Earth (i.e., spinning around the semiminor, North-South axis). This is also known as an equatorial coordinate system, because it is oriented relative to the Equator of Earth. The equatorial coordinate system is the reason why you can navigate by Polaris, the North Star, which would appear nearly directly overhead if you were standing at the North Pole. Even though Earth is not perfectly spherical, astronomical observations have been reliably used for millennia to transit the irregular oceans and terrain of Earth.

2.10 Geographic Coordinate Systems

Geocentric coordinate systems are essential for global navigation and mapping. In the modern era, we use **geographic coordinate systems** (sometimes abbreviated GCS) that are oriented to the rotational axis of Earth, much like the equatorial coordinate system. The origin of a geographic coordinate system is the center of Earth and the units of measurement are degrees of **longitude** and degrees of **latitude**. Degrees of latitude (denoted by lambda, λ) measure the angle from the equitorial plane North (+) or South (-), while degrees of longitude (denoted by phi, φ) measure the angle from the polar plane East (+) or West (-). Thus, geographic coordinate systems use angular units of measurement. Any combination of latitude and longitude gives coordinates $[\lambda, \varphi]$ on a sphere or ellipsoid. Positive values of latitude put you in the Northern Hemisphere, while negative values of longitude put you in the Western Hemisphere. Constant lines of latitude, known as **parallels** because they are always parallel to one another, and lines of longitude, known as **meridians**, form a grid that fits over a sphere or ellipsoid called a **graticule**.

For most of your life, you have probably believed that there is a singular combination of capital-L Latitude and Longitude values that absolutely define some location on Earth in perpetuity. This is perhaps the most profound geographic lie that we were taught as young school children. In fact, there are as many "types" of latitude as there are geographic coordinate systems. By now, you should be able to recognize the difference between *geocentric latitude* that is referenced to a sphere and *geodetic latitude* that is referenced to an ellipsoid. The main difference is that geocentric latitude is the angle relative to the centre of the sphere at the equatorial plane and geodetic latitude is the angle relative to the equatorial plane (i.e., not necessarily the centre). Figure **??** illustrates how the same angle can put you in two very different places depending on the geographic coordinate system you are using.

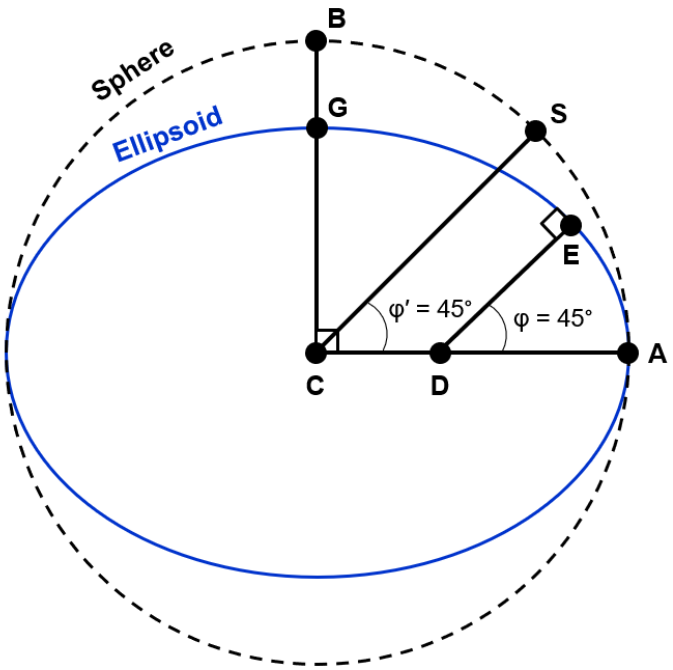


Figure 2.8: Geocentric versus geodetic latitude. Pickell, CC-BY-SA-4.0.

CS represents the line connecting the centre of a spherical geographic coordinate system to the surface of the sphere at a geocentric latitude $\varphi' = 45°$ and DE represents the line connecting the equatorial plane to the surface of the ellipsoid at a geodetic latitude $\varphi = 45°$. Notably, CS is parallel to DE and CS intersects with the surface of the sphere at a 90° angle and DE intersects the surface of the ellipsoid at a 90° angle.

A coordinate system that is referenced to an ellipsoid is known as a **horizontal datum**. For example, the World Geodetic System of 1984 (WGS 1984) is a *geodetic* datum that has a geographic coordinate system referenced to it. So the

origin of WGS 1984 is also the origin of the GCS from which latitude and longitude measurements are derived. So if you measured your latitude and longitude using the WGS 1984 horizontal datum, you would be placed somewhere on the ellipsoid also defined by WGS 1984. It is important to note, however, that the same exact *geocentric latitude* and longitude measures would place you somewhere else on a sphere. Other factors such as plate tectonics, glaciation, and ocean tides cause the Earth's surface to be in constant motion underneath any fixed horizontal datum. For example, Europe has drifted about 60 meters from North America since Eratosthenes first calculated the circumference of Earth. Therefore, any "type" of latitude or longitude is only useful during a particular epoch of time.

If you mapped the Earth as an ellipsoid in a three dimensional Cartesian coordinate system, you could describe location using three coordinates $[x, y, z]$, with $[0, 0, 0]$ being the center of Earth. However, we do not often express mapped coordinates of Earth using Cartesian coordinates. Instead of referring to the North Pole at $[0, 0, 6356752.3]$ meters (the polar radius of Earth) we usually refer to it with a single coordinate, 90° N. Why? The North and South Poles are the only points on Earth that can be defined with a single coordinate because they coincide with the orientation of a geographic coordinate system. Imagine placing a protractor at a 90° angle relative to a table. If you rotate it along that perpendicular axis, one arm of the protector would spin in a 360° circle, but the other arm of the protractor would always point up or down at 90° in the same direction. For all other locations on Earth, a pair of $[\lambda, \varphi]$ coordinates are needed to define location. With space-based global navigation systems it is more common to combine coordinates in both horizontal and vertical datums together. For example, $[\lambda, \varphi]$ expresses your location relative to the horizontal datum while $[\lambda, \varphi, h]$ expresses your ellipsoidal height (h) at a location and $[\lambda, \varphi, H]$ expresses your orthometric height (H) at a location.

2.11 Projected Coordinate Systems

Despite everything covered so far in this chapter, we very rarely see or display geographic data in 3-dimensions. In fact, most geographic data you will likely encounter will be either 1- or 2-dimensional (more on that in Chapter 3). Geographic coordinate systems are incredibly important for understanding how geographic data are fundamentally "attached" to the Earth. However, geographic coordinate systems are not suitable for creating maps, those 2-dimensional spatial models that can be easily displayed on your computer screen or printed on a sheet of paper. Instead, cartographers rely on **projected coordinate systems** that flatten a 3-dimensional geographic coordinate system to a 2-dimensional map. Really, these projected coordinate systems involve transformations called **map projections** that convert 3-dimensional coordinate space into 2-dimensional coordinate space, which means the map units are linear such as meters. Whenever we move from a geographic to a projected coordinate

system, we lose information, and distortion results.

Cartographers have wrestled with how to project Earth onto a printed page for millennia. The fundamental mathematics for map projections were first comprehensively described by Claudius Ptolemy around 150 C.E. Ptolemy's work *Geography* was one of the earliest treatises on cartography and map making that included an atlas of regional maps of Europe, Africa, and Asia. Ptolemy's work built on and came several centuries following Eratosthenes and earlier Greek geocentric observations by Plato and Aristotle. Ptolemy observed that a globe was the best way to represent the intervals and proportions of Earth's surface without distortion. However, globes are not very useful for looking at regions in detail and you can only see part of a globe at any given time. Thus, a mathematical language is needed to translate a geographic coordinate system to a planar or projected coordinate system. *Geography* was lost to antiquity before it was rediscovered, copied, and translated centuries later, first by Muslim cartographers in the 9th century C.E. and later by Italian cartographers in the 15th century C.E. during the Renaissance, which gave rise to the many types of map projections that we see today.

Because all map projections result in distortion from the loss of the third spatial dimension, it is useful to think about map projections in terms of what they preserve. There are four main characteristics that can be distorted or preserved, which give rise to the primary types of map projections that are in use for environmental management applications:

- **Conformal** projections preserve shape and angles
- **Equal-area or equivalent** projections preserve area
- **Azimuthal** projections preserve direction
- **Equidistant** projections preserve scale and distances

Some map projections can preserve several of these characteristics at once, but only a globe can simultaneously preserve area, direction, distance, and shape. Any map projection will have inherent trade-offs representing these characteristics accurately. It is beyond the scope of this textbook to discuss all map projections. Instead, we will focus on several key examples of map projections that are commonly used for environmental management applications. For a more comprehensive discussion of map projections generally, the reader is referred to *Map Projections: A Working Manual* by (**?**). In the next section, we look at how map projection distortation can be measured.

2.12 Measuring Map Projection Distortion

Tissot's Indicatrix is often used to visualize distortion from map projections, named after Nicholas Auguste Tissot. The metric is relatively simple: Tissot's Indicatrix is a perfect circle on the surface of a 3-dimensional globe, but will form an ellipse whenever projected to a 2-dimensional coordinate system. For this reason, Tissot's Indicatrix is sometimes referred to as Tissot's Ellipse. Since

ellipses can vary along two axes, Tissot's Indicatrix can represent areal, angular, and linear map distortions both longitudinally and latitudinally at any location in the map. This is very handy, because we can place Tissot's Indiciatrices (the plural of indicatrix) at different locations and examine how distortion changes throughout the map projection.

So how do we use this tool? The quotient between a line projected onto a map a and the same line on a globe a' is $\frac{a}{a'} = 1$ when there is no distortion on that axis of the ellipse. This quotient is also called a scale factor because it is showing how the map scale is modified locally by a map projection. Figure **??** below shows an example of a reference indicatrix that is a perfect circle on a globe with axes a and b.

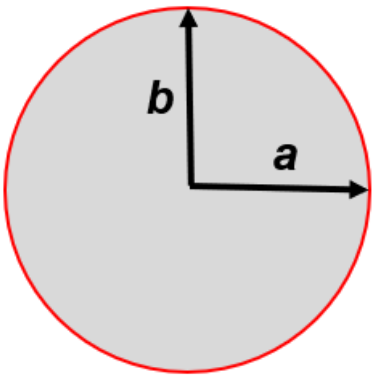


Figure 2.9: Reference Tissot Indicatrix. Pickell, CC-BY-SA-4.0.

Let us assume that $a = b = 1$. Then, the reference indicatrix has the following properties: $a = b$, $a \times b = 1$, and $Area = \pi ab = \pi^2$. If $\frac{a'}{a'} > 1$, then we can conclude that the the projection is *expanding* the distance along the a line. If $\frac{a'}{a'} < 1$, then we can conclude that the projection is *compressing* the distance along the line a. For example, suppose $a = 2$ and $b = 0.5$, then we have modified the scale of the indicatrix along both axes, but the areal solution is the same as the reference indicatrix shown as the red dashed circle in Figure 2.10 below.

The indicatrix above in Figure **??** is an example of an *equivalent* indicatrix, which has the following properties: $a > b$, $a \times b = 1$, and $Area = \pi ab = \pi^2$.

If $\frac{a}{a'} \times \frac{b}{b'} > 1$, then we can conclude that the projection is *inflating* the area. If $\frac{a}{a'} \times \frac{b}{b'} < 1$, then we can conclude that the projection is *deflating* the area. Consequently, whenever the quotients of both axes are equivalent (i.e., $\frac{a}{a'} = \frac{b}{b'}$), Tissot's Indicatrix forms a perfect circle and the ellipse is conformal with angles true to the globe. For example, suppose $a = b = 2$, then we have modified the scale of the indicatrix along both axes, but with the same factor. This results in a *conformal* indicatrix that is not equivalent, shown in Figure **??** below.

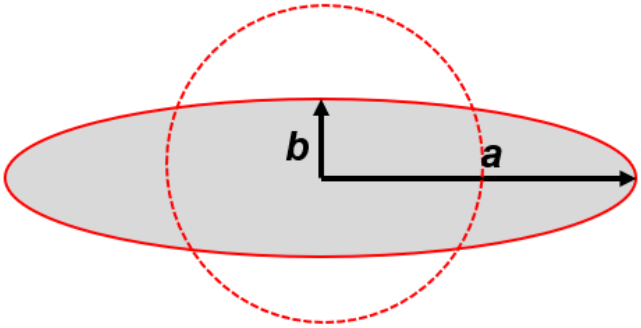


Figure 2.10: Equivalent Tissot Indicatrix. Pickell, CC-BY-SA-4.0.

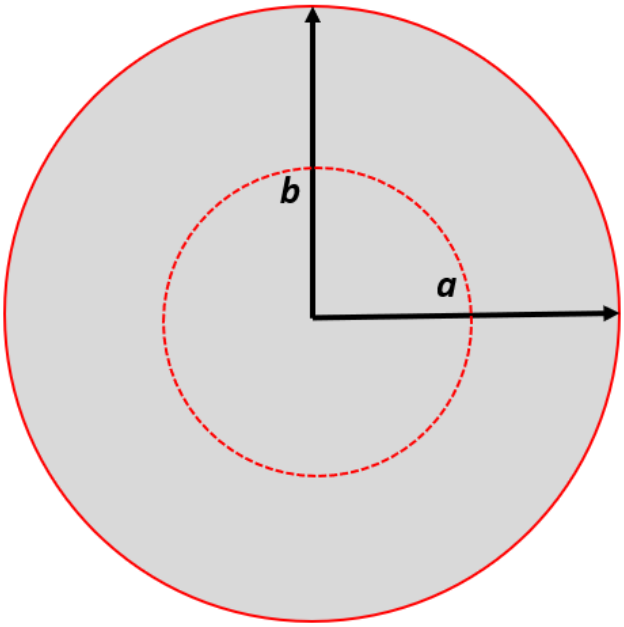


Figure 2.11: Conformal Tissot Indicatrix. Pickell, CC-BY-SA-4.0.

This conformal indicatrix has the following properties: $a = b$, but $a \times b \neq 1$, and therefore $Area = \pi ab = 4\pi^2$. As you can see, whenever $a \times b = 1$ the indicatrix is equivalent (equal-area) and whenever $a = b$ the indicatrix is conformal.

2.13 Map Projections for Environmental Management

2.14 Mercator

Mercator is a cylindrical map projection that represents meridians and parallels as straight lines. The cylindrical surface is oriented such that the rotational axis of Earth runs through the openings of the cylinder and the Equator represents the tangent where the cylinder meets the Earth's surface. Scale along the tangent is true because the translation from the spherical Earth to the cylindrical surface is one-to-one at the tangent, so this is also the location on the projection where there is no distortion. This has the effect of accurately representing the shape and angles on the map (i.e., conformal), but greatly distorts area as you move away from the Equator. In fact, the North and South Poles are represented as 2-dimensional lines at the top and bottom edges of the map instead of 1-dimensional points. Although scale and area change as you move North or South along a meridian, scale and area are equivalent along any parallel, but not necessarily true to a globe. In other words, you can compare area or scale anywhere along a parallel, but only at the Equator is the area and scale true to the globe.

The Mercator map projection is perhaps the most pervasive and reproduced projection around us (Figure **??**). Because angles are preserved, you can easily and accurately navigate long distances across Earth, and this was exactly the purpose that Gerardus Mercator envisioned when he first identified the projection for sea-faring Europeans in 1569. You may also recognize the Mercator projection from web mapping applications like Google Maps, which use it because it ensures that North-South roads intersect at right angles with East-West roads.

2.15 Universal Transverse Mercator (UTM)

Universal Transverse Mercator (UTM) is very similar to a Mercator projection except that the cylinder is rotated or transverse by 90° so that the opening of the cylinder is perpendicular to the rotational axis of Earth. This has the effect of moving the tangent from the Equator to any Meridian. In fact, you can rotate the cylinder at any angle you want where 0° is a true Mercator, 90° is a transverse Mercator, and any other angle is considered an oblique Mercator. UTM is actually a system of 60 different transverse Mercator projections that are defined to represent 6° Longitudinal intervals of Earth's surface (60 zones x 6° = 360°). Each projection is defined as a zone, which is also divided into

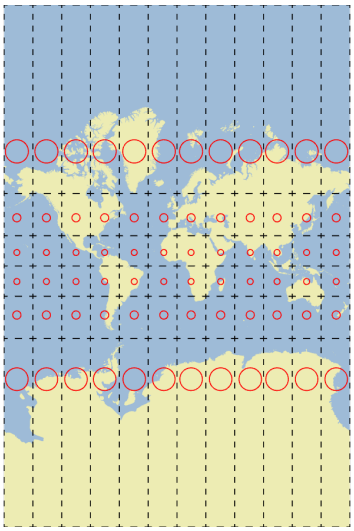


Figure 2.12: Mercator map projection with Tissot's Indicatrices in red. Pickell, CC-BY-SA-4.0.

North and South zones depending whether you are in the Northern or Southern Hemisphere. Canada spans 15 UTM zones from Zone 7 North in the Yukon to Zone 22 North covering Newfoundland. Figure **??** below shows a map of UTM Zone 13 over Saskatchewan.

Besides defining the orientation of the cylinder, we can also specify its size or diameter. When the diameter of the cylinder is equivalent to the diameter of Earth at the Equator, a single tangent line is formed. If the diameter of the cylinder is smaller than Earth's diameter at the Equator, then the cylinder has two lines that contact Earth's surface, known as **secants**. The purpose for having two secants is that the projection distortion can be more evenly distributed across the map. In the case of UTM, a scale factor of 0.9996 is applied to shrink the transverse cylinder slightly, forming two secants that are 360 km apart East-West. In between the two secants lies the **central meridian**, which is used to define the origin for the projected coordinate system. It is important to realize that the secants are parallel to each other and the central meridian, which means the secants are not meridians on Earth and form what are called **small circles**, a line that does not divide Earth into two equal portions.

UTM uses a unique coordinate system that deserves some explanation. Just like with latitude and longitude coordinates $[\lambda, \varphi]$, an arbitrary origin needs to be defined so that we know where we are in relation to the origin on the map. For projected coordinate systems like UTM that use linear units of measure such as meters, the origin is defined as the intersection of the central meridian and the Equator. Simple enough, right? There is one catch: UTM does not use any negative coordinates by convention. Thus, the origin of the coordinate system

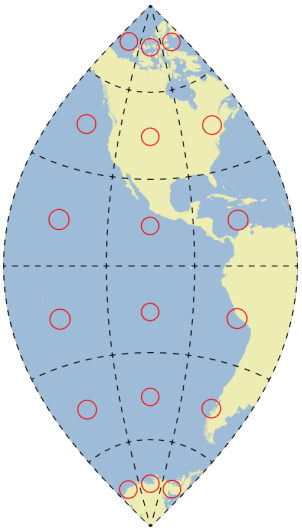


Figure 2.13: Universal Transverse Mercator Zone 13 map projection with Tissot's Indicatrices in red.Pickell, CC-BY-SA-4.0.

for each zone must be moved so that coordinates West of the central meridian and South of the Equator are positive. To do this, a constant value is added to all East-West coordinates (known as **Eastings**) and all North-South coordinates (known as **Northings**) to create what are known as **False Eastings** and **False Northings**, respectively. A value of 500,000 m is added to all Eastings so that the western limit of the zone is located at 0 m, the central meridian is located at 500,000 m, and the eastern limit of the zone is located at 1,000,000 m. A value of 10,000,000 m is added to all Northings in the Southern Hemisphere so that the Equator is at 10,000,000 m and 0 m is near the South Pole for all southern UTM zones. You might recognize the importance of the 10,000,000 m value because this represents approximately one-quarter of the Earth's North-South circumference.

2.16 Sinusoidal

Sinusoidal is a *pseudocylindrical* map projection, so-named because these projections approximate a true cylindrical projection except that Meridians are curved instead of straight like with Mercator or UTM. Because Meridians are curved, Sinusoidal maps represent the North and South Poles as single points instead of lines as is the case with Mercator. Thus, Sinusoidal maps are not conformal and distort shape, but they are in fact equal-area (Figure **??**). Equal-area map projections like Sinusoidal are important for accurately accounting for land cover and other global mapping efforts. Therefore, it is common to find global datasets distributed in a Sinusoidal map projection.

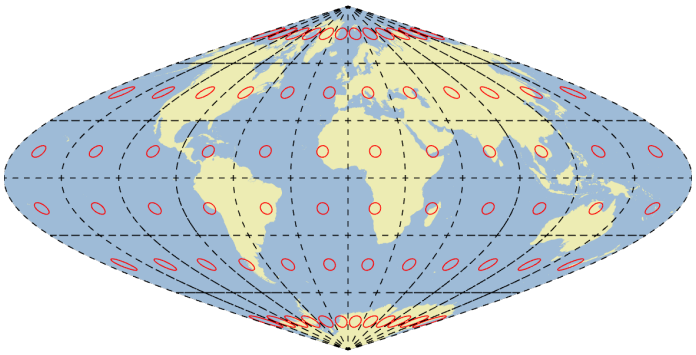


Figure 2.14: Sinusoidal map projection with Tissot's Indicatrices in red. Pickell, CC-BY-SA-4.0.

2.17 Albers

Albers is another example of an equivalent map projection, but unlike Sinusoidal, an Albers projection uses a cone as a projection surface instead of a pseudocylinder. Like the cylindrical case, the cone can be sized and oriented in any way, but the cone of an Albers projection is typically oriented so that the vertex of the cone aligns with the rotational axis of Earth. The base of the cone has a diameter that usually results in two secants, known as **standard parallels** on an Albers map. Thus, Albers projections tend to distort latitudinally as you move North-South away from the standard parallels, but even more so as you move toward the base of the cone (Figure **??**). Besides being an equal-area projection, Albers is a good choice for mapping regions because shape and scale are mostly preserved near the standard parallels. For example, the province of British Columbia has adopted a modified Albers projection that situates the standard parallels at 50° N and 58.5° N, which are near the northern and southern latitudinal limits of the province. This narrow band of latitude between the standard parallels ensures that there is relatively little distortion in shape and scale within the province, which is comparable to UTM. However, British Columbia is a longitudinally wide province, spanning 6 of Canada's 15 total UTM zones, so Albers has a distinct advantage of being able to show the entire province with little distortion. For the same reasons, you will often find Canada-wide data distributed in an Albers projection.

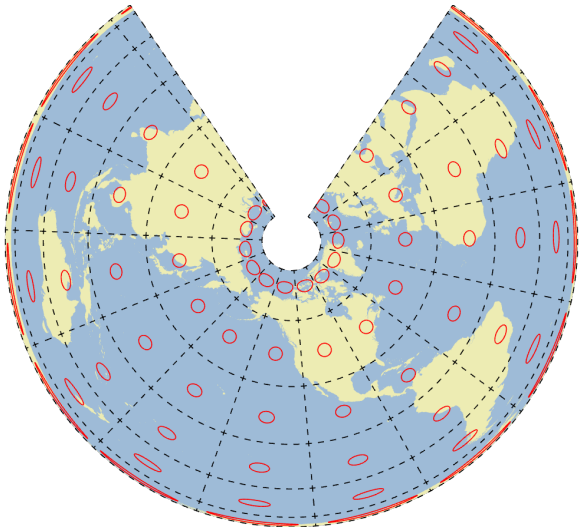


Figure 2.15: British Columbia Environment Albers map projection with Tissot's Indicatrices in red. Pickell, CC-BY-SA-4.0.

2.18 Azimuthal

Azimuthal projections use a flat circular plane to project the Earth onto a map. This plane is usually oriented so that it is tangent at a single point, usually the North or South Pole. In the case of a polar azimuthal projection, meridians radiate outward as straight lines from the pole to the edge of the circular plane and the parallels are represented as concentric circles. As a result, distortion increases as you move away from the centre point of the map with the outer edge of the plane representing an **antipode** or an opposite point on Earth. The primary benefit of azimuthal projections is that they preserve direction and distance between the centre point and any other point in the map (Figure **??**). The shortest geographic distance between the centre point and any other point creates a line known as a **great circle**, which divides the Earth into two equal portions. So another benefit of an Azimuthal project is that great circles can be mapped as straight lines. Azimuthal projections are commonly used when distance and direction are important, such as weather RADAR stations or air traffic control towers. It is important to realize that the centre point for an azimuthal projection can be any point on Earth and the equidistant property can be exploited for a number of applications.

2.19 Summary

In this chapter, you learned about the science and technology of geodesy that goes into mapping data. We described the different models of Earth's shape and

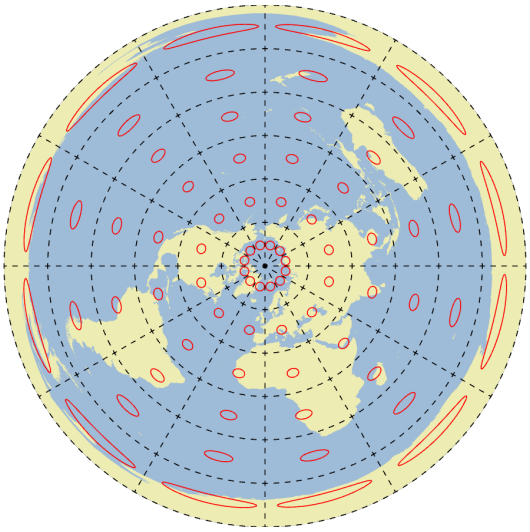


Figure 2.16: North Pole Azimuthal map projection with Tissot's Indicatrices in red. Pickell, CC-BY-SA-4.0.

the advantages and disadvantages that each model has. More generally, models of Earth represent vertical datums to which heights are referenced. When we think of "where" something is on Earth, we must use horizontal datums to reference location. There are two types of horizontal datums that are very important to geomatics: (1) a geographic coordinate system uses lines of latitude and longitude to define locations on Earth's 3-dimensional surface while (2) a projected coordinate system flattens Earth into a manageable 2-dimensional coordinate space. In the case of projected coordinate systems, we have many choices when deciding which map projection to use when we map data, each with its own uses and distortions. The next time that you look at a map, ask yourself with your new found appreciation of geodesy, how is this information being misrepresented to me?

Reflection Questions

1. Describe the process of measuring the height of something on Earth.
2. Explain the difference between a geographic coordinate system and a projected coordinate system.
3. Name as many projected coordinate systems as you can.
4.

Practice Questions

1. Choose a continent to roughly draw and then draw a grid of latitude and longitude lines in 3 different projected coordinate systems on a piece of paper. Describe the distortions that you expect and point to where they appear on your continent.
2. xx

Recommended Readings

Ensure all inline citations are properly referenced here.

Chapter 3

Types of Data

‿ Can we change the chapter name? I think something like **Spatial Data Models & Data Types** would be more appropriate.

In the previous chapter, we discussed some of the unique challenges associated with representing spatial data in a GIS, and how to account for these these with geographic coordinates systems and map projections. In this chapter we will discuss more broadly how to represent both spatial and non-spatial data in a Geographic Information System. We will introduce the different types of data that can represent non-spatial attributes and discuss the different scales this data can be measured on. Then we will introduce the different *spatial data models* we use to link the spatial and non-spatial data. Finally, we will cover some of the different file types that can be used to store data.

Learning Objectives

1. Data types and measurement scales
2. Introduce spatial data models
3. File types we use in GIS

Key Terms

Attribute, Qualitative, Quantitative, Discrete, Continuous, Vector, Raster, Measurement Scale

3.1 Types of Phenomena

There are two kinds of phenomena we observe in the real world and represent in a GIS **Discrete Objects** and **Continuous Fields**. Discrete objects have

distinct, well-defined boundaries meaning their geography can be exactly measured (e.g. a building, a street, a tree). They are also countable, meaning there are a finite number of them. Continuous fields lack clearly-defined boundaries and can be measured at an infinite number of points (e.g. elevation, precipitation, land cover). The values of a continuous field can vary dramatically over short distances. Both kinds of phenomena can be represented in a GIS, but they requrie different considerations and are usually use different data models.

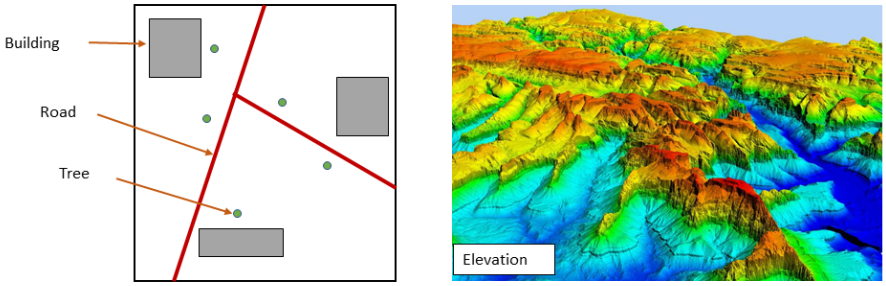


Figure 3.1: One the right, we see discrete objects, on the left we see a continuous field. These images are from my slides. I'm not sure the original source of the DEM image, it would be easy enough to just make my own down the line though.

3.2 Types of Data

Within the context of a Geographic Information System, each piece of information pertaining to a phenomena can be referred to as an **Attribute**. An phenomena can have many different attributes associated with it, but each attribute can broadly be said to address one of three questions: **What**, **When**, or **Where**? Attributes that describe *where* are known as **Spatial Data** while all other attributes are **Non-Spatial Data**. All data, spatial and non-spatial, can broadly be classified as either **qualitative** or **quantitative**. These data types fundamentally different and are therefore measured on fundamentally different scales. The types of analysis we can conduct with qualitative data are more limited than quantitative data, but that does not necessarily mean quantitative data are "better" than qualitative.

3.3 Qualitative Data

Qualitative data are categorical; they are strictly descriptive and lack any meaningful numeric value. They describe the qualities of an phenomena, without giving us any numeric information. Most qualitative data you will work with in a GIS are textual or coded numerals, but there are circumstances where you may encounter non-textual data (e.g. images, sound clips, videos) in a dataset.

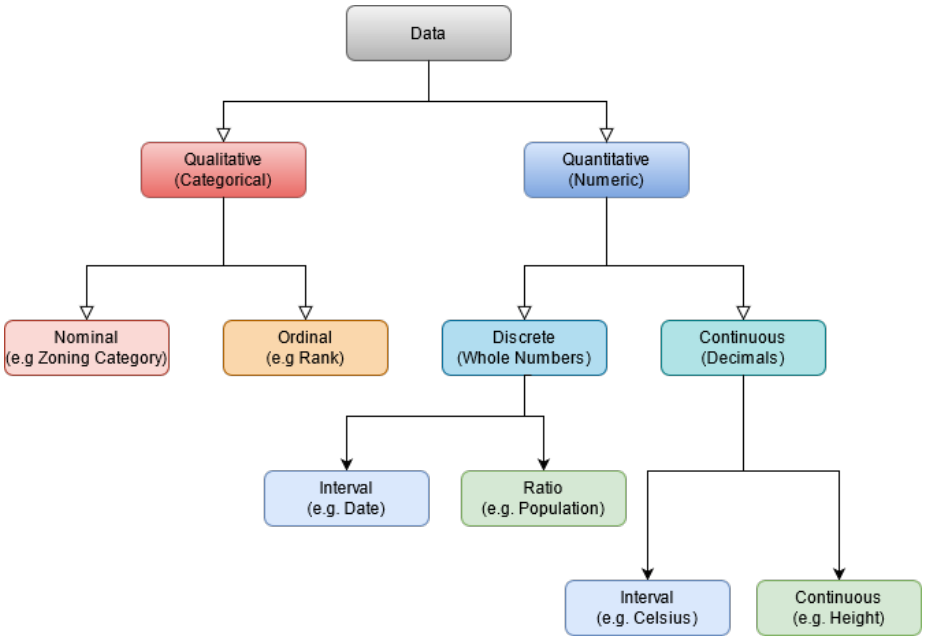


Figure 3.2: Summary of the types of data.

Qualitative data can be "spatial" in nature (e.g. relative directional descriptors: left/right, near/far, north/south), but because they lack numeric values, they cannot be used for spatial analysis. Qualitative data can be measured on either a **Nominal** or **Ordinal** scale.

3.4 Nominal Scale

These are data that just consists of names or categories with no ranking or direction are nominal. One category is not more or less, better or worse than another, they are just different. A good example would be flower types (see fig). Other examples would be zoning categories, colors, flavors of ice cream, place names, etc.

3.5 Ordinal Scale

These data are categories that also have a some ranking or directionality. A good example would be relative sizes (see fig). Some other good examples of ordinal data include spice levels (mild, medium, hot), residential zoning density (low, medium, high), and survey responses.

The only arithmetic operations we can do with nominal data are checking for equality (True/False), counting occurrences (frequencies), and calculating the

Figure 3.3: Each flower is different, but no flower is "more" or "less" a flower than any of the others.

mode (most frequent occurrence). With ordinal data, we can these operations as well, plus a few more. We can check the order/rank (greater than, less than) and in some circumstances we can calculate the median (see figure).

3.6 Graded Membership

When trying to group real world phenomena into categories, there are often "exceptions" that blur the lines a bit. Take this example: you are trying to develop a land cover classification scheme for Garibaldi Provincial park in British Columbia. Some of the land surface is unquestionably alpine tundra and some is certainly forest area. However, the transition between forest and alpine meadow is not an abrupt line. How/where do you draw the line? Examples like this are known as fuzzy variables, and we often use a **Graded Membership** scale to assign them to categories. With the landscape classification, a simple approach would be a "winner take all" approach. If a plot is 5% bare rock, 40% forest, and 45% alpine meadow, the area will be classified as alpine meadow. From that point forward, in the GIS, that area will be treated as alpine meadow, any information about the variability within the area will be lost. In practice, many of the qualitative data we work with in GIS, especially those describing natural phenomena, are actually graded membership variables.

3.7 Quantitative Data

Quantitative data are numeric; they describe the quantities associated with an phenomena. The numerical values that are separated by a unit that has some

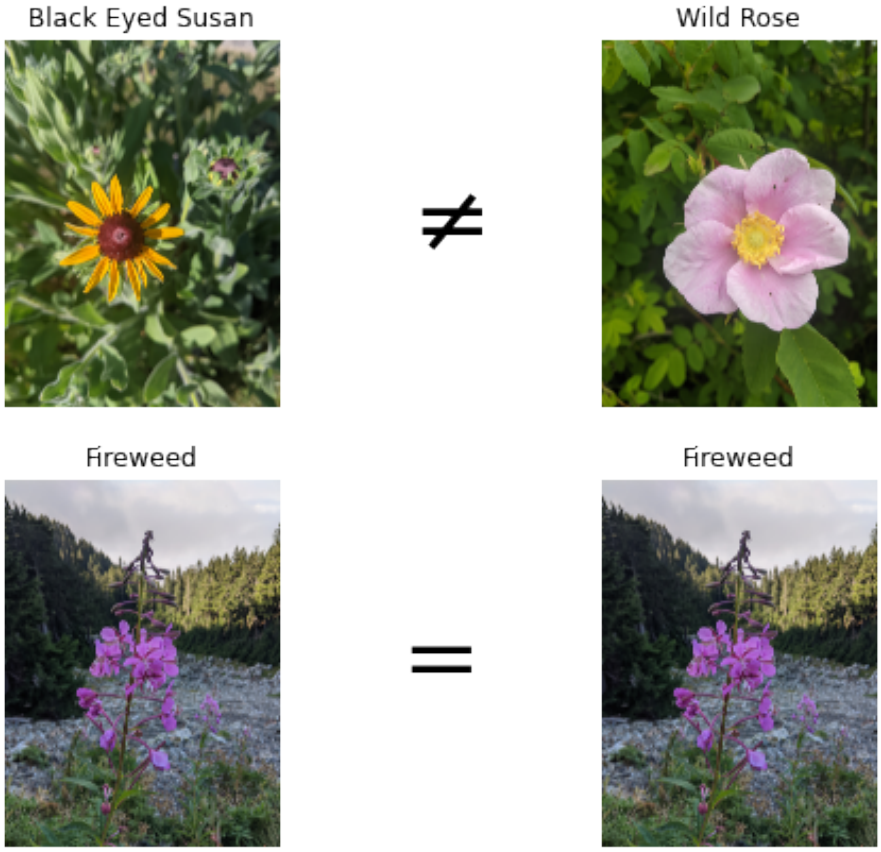


Figure 3.4: With nominal data, you can check for equality, and count occurenecees, but that is it.

Figure 3.5: We can see Yarrow is taller than her sister Shamsa, so we can rank these dogs by height. However, we haven't measured their heights, so we don't know how much taller Yarrow is than Shamsa.

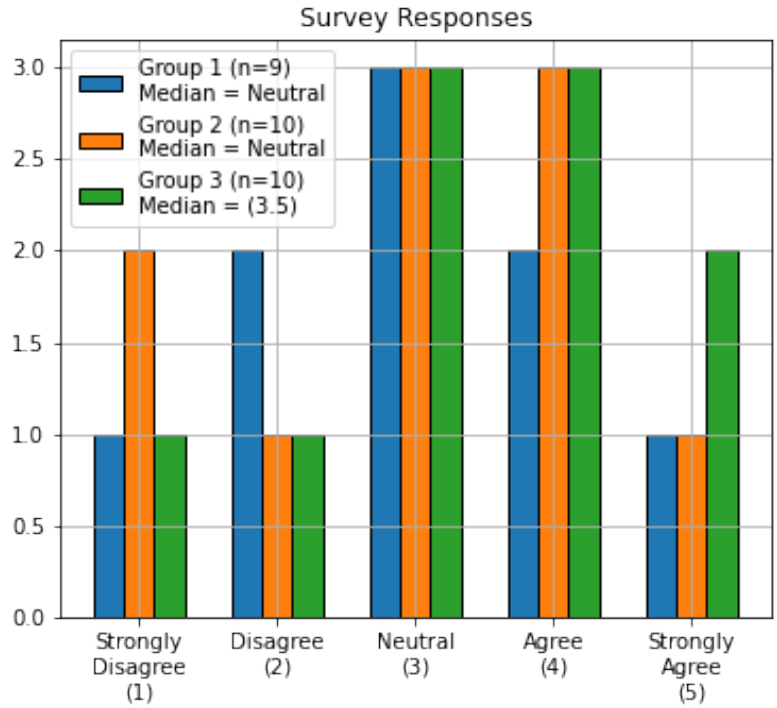


Figure 3.6: In some circumstances, we can directly calculate the median (middle value) of an ordinal set. With odd numbered sets (e.g. Group 1), the median, is simply the middle value of the set, when sorted lowest to highest. We can always take the median when we have an odd number. With even numbered sets, its a bit more complicated. The median, is the average of the middle two values. For Group 2, the middle values (5th and 6th) are both "Neutral", so we don't have an issue. But for Group 3, the 5th value is "Neutral" and the 6th value is "Agree". We can't directly average these two ordinal values. One solution is to arbitrarily assign a numeric score to the ordinal categories (e.g. 1-5). This would then allow you to show the median is between "Neutral" and "Agree".

Figure 3.7: Will annotate to show better example.

inherent meaning (as opposed to the arbitrary numeric codes like in the ordinal data example). This allows us to conduct a wider range of arithmetic operations on quantitative data. In addition to the operations we perform on Qualitative data; with numeric data we can always calculate measures of central tendency (mean/median) and we can add/subtract values to calculate differences.

Numeric data can be either **discrete** or **continuous**. Discrete variables (e.g. population) are obtained by counting and values within a range cannot be infinitely subdivided. You can have a population of 1, 37, or 179 but you cannot have a population of 2.3. Continuous variables (e.g. temperature) can take an infinite number of values a given range, but they cannot be counted. You can have temperatures of 10, 10.5, or 10.1167 °C, but a temperature of 10°C does not mean you have 10 individual degrees of temperature. Quantitative data (both discrete and continuous) can be measured on either an **Interval** or **Ratio** scale. These types of quantitative data are closely related, but have one important distinction.

3.8 Ratio Data

These data have fixed, meaningful, absolute zero points. The absolute zero point means ratio data cannot take negative values. It also means that we can multiply/divide two values to calculate a meaningful ratio between them (hence the name). A good example of ratio data are population total (see figure). Population counts start at zero and go up from there. A population of zero means there are no residents, and its impossible to have a negative population. Other examples of ratio data include: temperature (*in degrees Kelvin*), precipitation, tree height, income, rental cost, and units of time (years, seconds, etc.)

3.9 Interval Data

These data on the other hand, have an arbitrarily set zero point. This means they can can take negative values. Because the zero point is arbitrary, we cannot multiply/divide two values or calculate meaningful ratios between two values. A good example of interval data is temperature measured in Celsius, and comparing it to Kelvin highlights the difference between the two data types (see fig). The conversion between Kelvin (ratio) and Celsius (interval) is very simple: °C = °K-273.15. Zero Kelvin is "Absolute Zero" - ie. the lack of temperature, while zero Celsius is the freezing point of water (273.15 degrees above absolute zero). Other examples of interval data include: the pH scale, IQ test scores, elevation (relative to a datum) dates (April 12th, 2011), and times (11:00 A.M.).

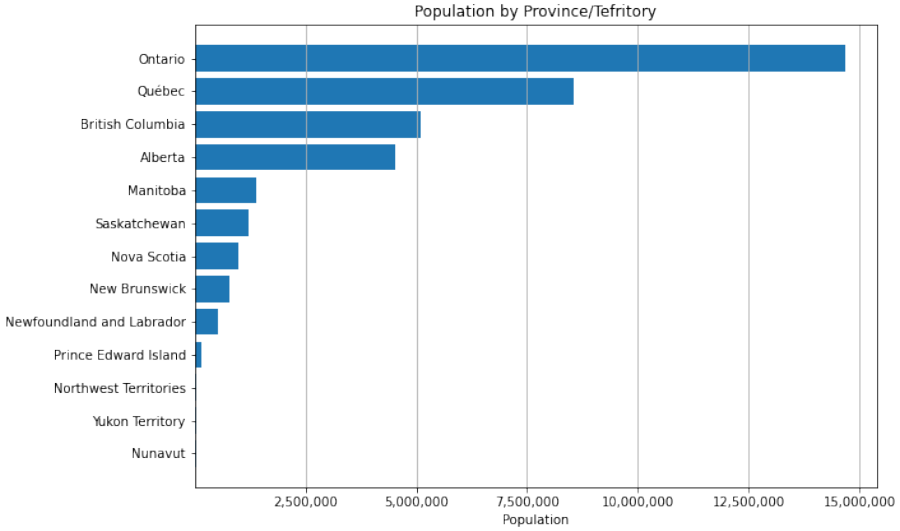


Figure 3.8: Because of the fixed, meaningful zero point, we can calculate ratios between populations: e.g. Manitoba's population is 1/10th that of Ontario, British Columbia has 129 times as many people as Nunavut.

3.10 Derived Ratio: Normalizing Data

Sometimes we want to account for the influence of one variable when analyzing another. To do this, we can divide one value by by another to get the ratio of the two, also known as a **derived ratio**. This process is sometimes referred to as **Normalizing** or **Standardizing** our data. The basic formula is: $C = \frac{A}{B}$, where A is our variable of interest, B is our confounding variable, and C is our new derived ratio. There are many circumstances where we might need to do this. One common example is population density (see figure). Another key example is housing affordability.

3.11 Summary of Data Types

3.12 Spatial is Special

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

You might encounter the phrase "Spatial is special" in your time studying GIS. Spatial data is the foundation of Geographic Information Science, it is what distinguishes GIS from the broader field of data science. This was succinctly summarized by Waldo Tobler in The First Law of Geography: - *"Everything is*

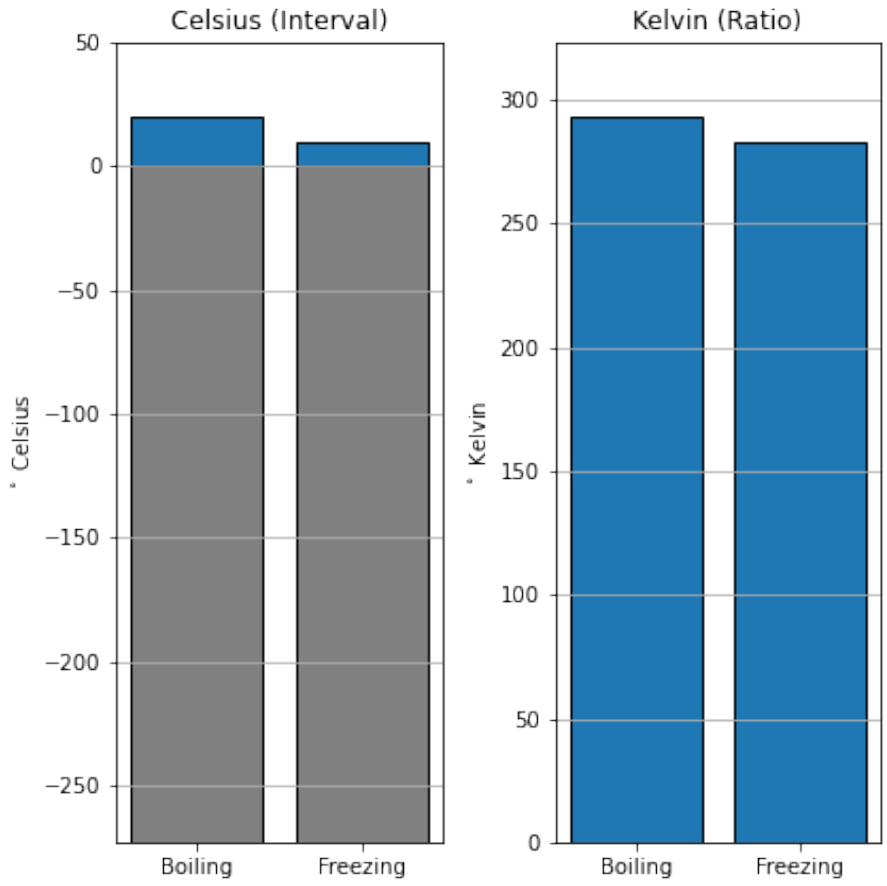


Figure 3.9: The ratio between two temperatures in Celsius is not meaningful, 20°C is not "twice" as warm as 10°C. Kelvin's zero point is fixed to absolute zero, the "absence" of temperature. So we can calculate the ratio, 293.15°K is 1.035 times warmer than 283.15°K.

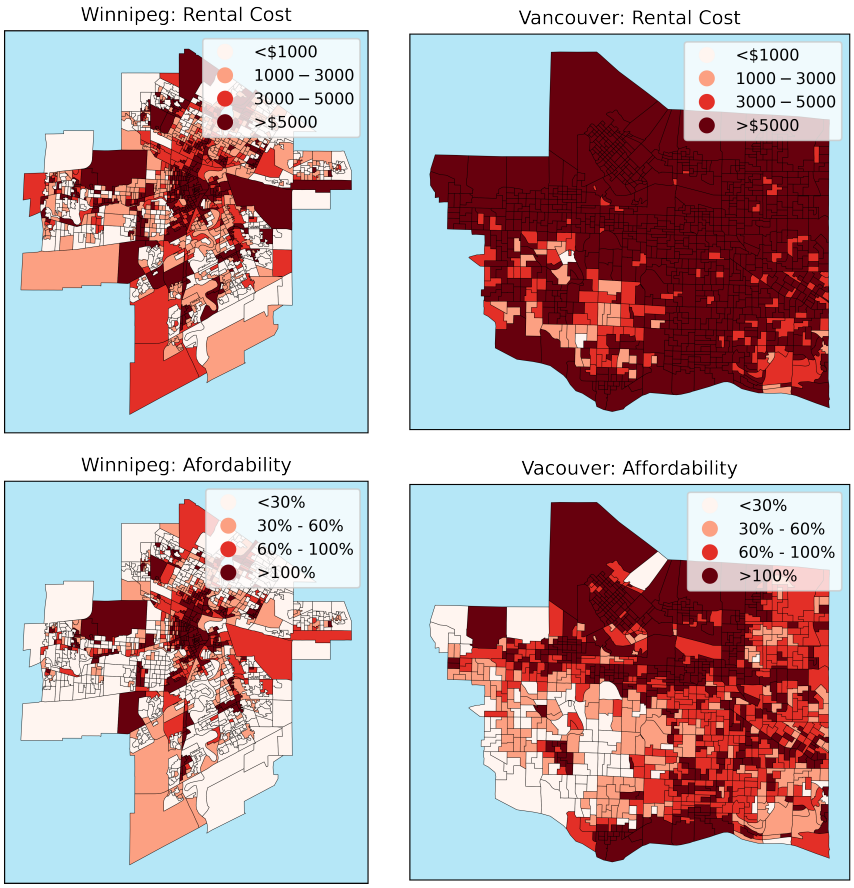


Figure 3.10: Figure needs to be re-worked as the data from simply analytics is wrong...

Operation	Nominal	Ordinal	Interval	Ratio
Equality	✔	✔	✔	✔
Order		✔	✔	✔
Add / subtract			✔	✔
Multiply / divide				✔
Mode	✔	✔	✔	✔
Median		✔	✔	✔
Arithmetic mean			✔	✔

Figure 3.11: From my lectures. IDK the original source, but it'd be easy enough to just "recreate" my own.

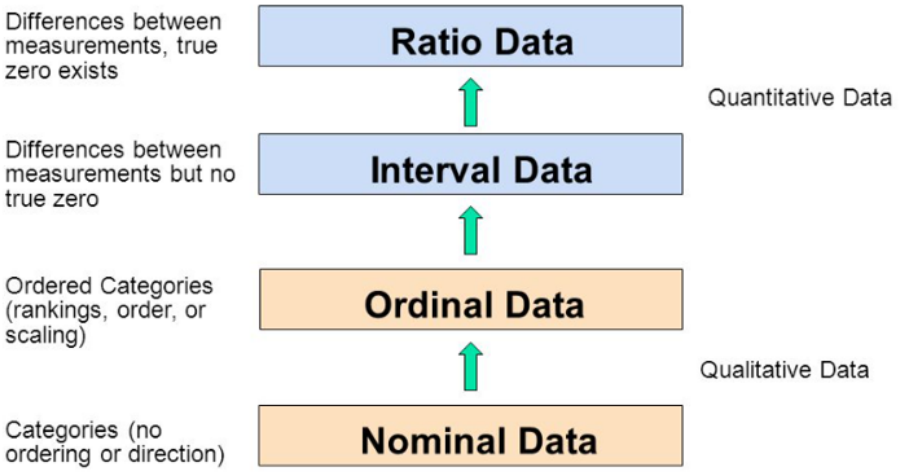


Figure 3.12: From my lectures. IDK the original source, but it'd be easy enough to just "recreate" my own.

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

This might seem obvious: people interact more if they live in the same city, orca pods in different areas develop different dialects, A hemlock on Vancouver Island are more related to its neighbors than to a hemlock in the New Brunswick. Generally, near things are more related to one another, but it **does not guarantee similarity**. Downtown Vancouver averages 40 cm of snow/year, but the ski resort on Grouse Mountain 15 km north gets over 9 m. These locations are impacted by the same storm systems, but the 1200 m elevation difference causes vastly different quantities and different types of precipitation.

The measure of similarity between objects across space called **spatial autocorrelation**. Spatial autocorrelation allows us to make some a key assumptions when representing spatial data. We don't have to measure a phenomena everywhere in order to represent it adequately. We only need to measure it at specific locations or over regular intervals. If point A is in dense forest, it is likely point B 10 m away is also in a dense forest. We don't have to get the location of every tree in the forest. Instead, we can look at the average presence of trees over a larger area.

3.13 Spatial Data Models

As discussed in the previous chapter, spatial data is three-dimensional, though we usually project it into two-dimensions for simplicity. Because of the unique transformations that must be applied to spatial data, it must be treated and represented differently than the non-spatial data that describe **what** is happen-

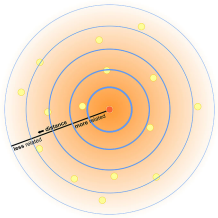


Figure 3.13: Visualization of Tobler's First Law. [8]

ing and **when**. We can't simply put all of our data into a spreadsheet and start analyzing it. We have to use **Spatial Data Models** to organize our data and link our spatial and non-spatial data. Spatial data models store geographic data in a systematic way so that we can effectively display, query, edit, and analyze our data within a GIS.

There are two main types of spatial data models: the **Raster** and **Vector** models. The raster data model represents spatial data as grid of cells, and each cell has one non-spatial attribute associated with it. The vector data model represents spatial data as either points, lines, or polygons that are each linked to one or more non-spatial attributes. These two models represent the world in fundamentally different ways. One is not inherently better than the other, but they are better suited for different circumstances. The choice of which model to use is often dictated by three main factors:

1) The type of phenomena we are trying to represent.
2) The scale at which we plan to analyze our data.
3) How we plan to use the data.

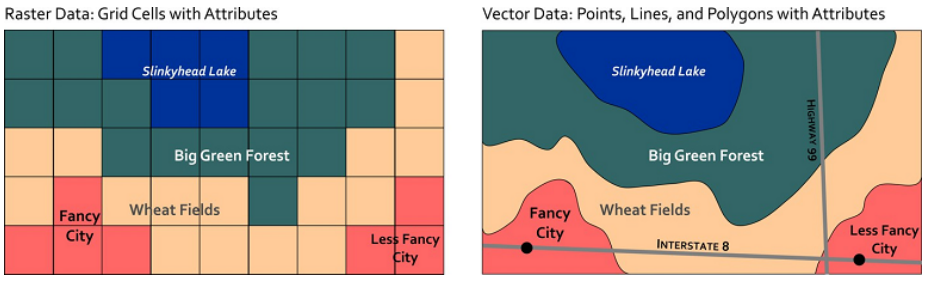


Figure 3.14: Representing space in the raster model vs. the vector model [8]

3.14 Raster Data Model

The raster data model represents a phenomena across space as a gridded set of cell (or pixels). The cell size determines the **Resolution** of the raster image, that is the smallest feature we can resolve with the raster. A 10 m resolution

raster has cells that are 10 x 10 m (100 m2), a 2 m resolution has cells that are 2 x 2 m (4 m2). Along with the cell size, the number of rows and columns dictates the bounds (or extent) of a raster image. A raster with a 2 m cell size, 15 rows, and 10 columns, will cover an area of 30 m x 20 m (600 m2). Because of the full coverage within their bounds, raster data models are very well suited for representing **continuous phenomena** where cell values correspond to measured (or estimated) value at specific location. In GIS, rasters are commonly encountered as: satellite and drone imagery, elevation models, climate data, model outputs, and scanned maps.

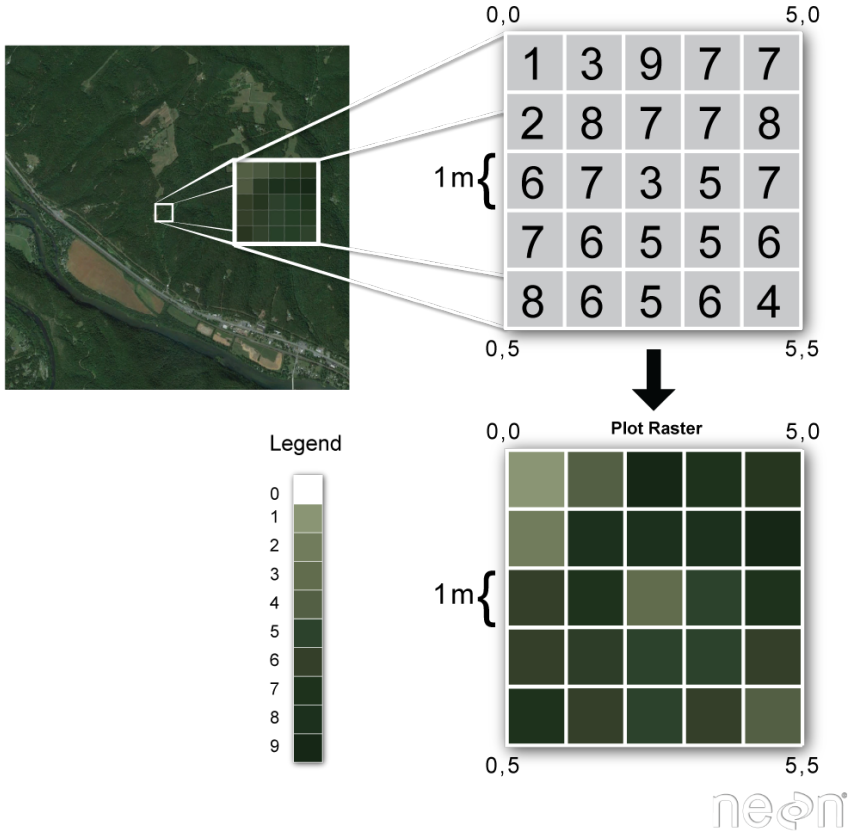


Figure 3.15: I think this one is pretty good, I use it in my lecture notes. [8]

The value of a pixel can be quantitative (e.g. elevation) or qualitative (e.g. land use). Each pixel/cell can only have a single value associated with it. Multiple bands can be combined to store or more information, as is done with a RGB color photograph. Algebraic expressions can also be performed quickly and efficiently with raster layers a inputs. This is known as raster overlay, and is one of the key advantages to raster data. If layer A = Average July Temperature and

layer B = Average January Temperature, then A – B will give us the Average Temperature Range across the rasters domain.

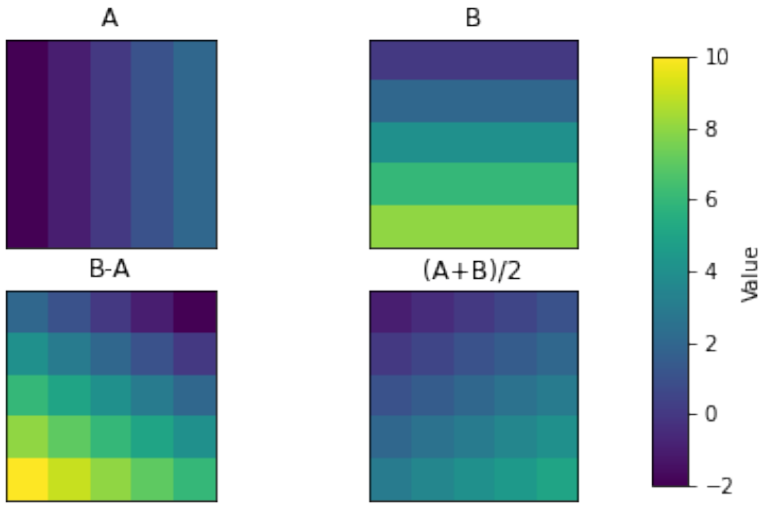


Figure 3.16: Raster math illustration

Rasters data relies on Spatial Autocorrelation and The First Law of Geography, the model assumes that *all areas* within a given cell are equally represented by the cell value. Depending on the resolution of the raster and the scale of the task at hand, this may or may not be an effective assumption. If you are trying to represent the coastline of Nova Scotia, 100 m or even 1 km resolution cells will likely suffice (see figure). However, 10 km cells severely degrade the quality of the representation. At a 100 km cell size, the province is indistinguishable.

3.15 File Types

Raster data can come in many different formats. One of the most common/functional is the GeoTIFF format which has the extension .tif. A .tif file stores metadata or attributes about the file as embedded tif tags. For instance, your camera might store a tag that describes the make and model of the camera or the date the photo was taken when it saves a .tif. A GeoTIFF is a standard .tif image format with additional spatial (georeferencing) information embedded in the file as tags. These tags should include the following raster metadata: * Extent * Resolution * Coordinate Reference System (CRS) - we will introduce this concept in a later episode * Values that represent missing data (NoDataValue) - we will introduce this concept in a later lesson.

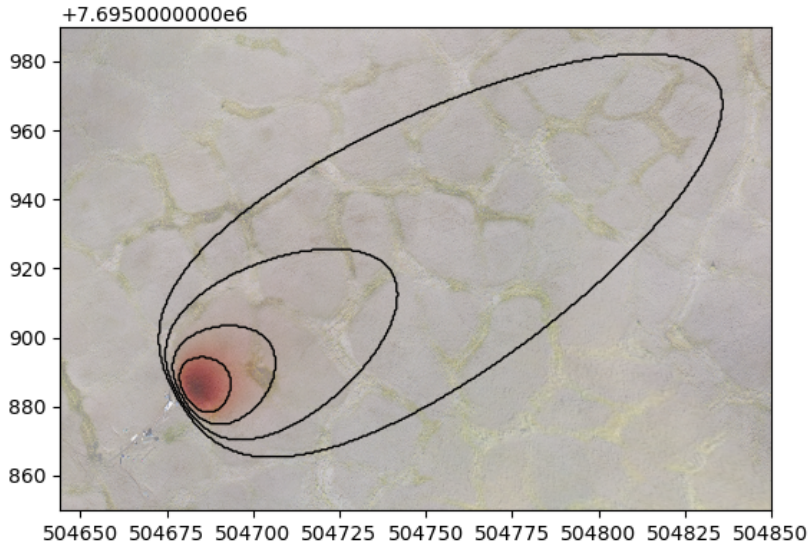


Figure 3.17: Raster Resolution

^ Adapted from [4]

Other file types you will likely encounter include: * 1) IMG - A proprietary image format commonly used by ESRI products * 2) JPEG2000 - A geospatial version of the common .jpg image type * 3) ASCII - An older human readable format (simple text file) with slower performance than the types listed above.

3.16 Vector Data

The vector data model is much more well suited to represent discrete phenomena than the raster data model. A vector feature is a representation of a discrete object as a set of x,y coordinate pairs (points) linked to set of descriptive attribute about that object. A vector feature's coordinates can consist of just one x,y pair to form a single point feature, or multiple points which can be connected to form lines or polygons (see figure). The non-spatial attribute data is typically stored in a **Tabular** format separate from the spatial data, and it is linked using an index. One of the key advantages of the vector model is the ability to store and retrieve many attributes them quickly.

Points are "zero-dimensional", they have no length, or width. A point feature is just an individual , coordinate pair representing a precise location, that has some linked attribute information. Points are great for representing a variety of objects, depending on the scale. Fire hydrants, light poles, and trees are suitable to be represented as points in almost any application. If you are making a map of mines in British Columbia, or cities across Canada, it's probably acceptable to just display them as points.

Lines are one-dimensional, they have length, but no width and thus no area. A line consists of two or more points. Every line must have start point and an end point, but they may also have any number of middle points, called vertices. A vertex is just any point where two or more lines meet. Lines are also great for representing a variety of objects, depending on the scale. Hiking trails, flight paths, coastlines, and power lines are suitable to be represented as lines in almost most applications. When making smaller scale maps, its often sufficient to represent rivers as lines, though at large scales we might elect to use a polygon.

Polygons are two-dimensional, they have both a length and width and therefore we can also calculate their area. All polygons consist of a set of at three or more points (vertices) connected by line segments called "edges" that connect to form an enclosed shape. All polygons form an enclosed shape, but some can also have"holes" (think doughnuts!), these holes are sometimes called interior rings. Each interior ring is a separate set vertices and edges that is wholly contained within the polygon and no two interior rings can overlap. Polygons are useful for representing many different objects depending: political boundaries boundaries, Köppen climate zones, lakes, continents, etc. At large scales they can represent things like buildings which we might choose to represent as points at smaller

scales.

Sometimes, a discrete object has multiple parts, that are spatially separated. In these circumstances, the vector model allows for multi-polygon, multi-line, or multi-point objects. A good example of when a multi-polygon would be useful is the StatsCanada provincial boundary file (see figure). Roads sometimes need to be stored as multi-lines as well, for example Highway 1 crosses the Georgia Straight from Vancouver to Nanaimo. If we want the to represent the entire Highway as one object, we need to use a multi-line.

Vector data also has a **Resolution** although it has a somewhat different definition in the context of the vector model.

3.17 File Types

–From [4] – Will adapt/adjust later

Like raster data, vector data can also come in many different formats. For this workshop, we will use the Shapefile format which has the extension .shp. A .shp file stores the geographic coordinates of each vertice in the vector, as well as metadata including:

```
Extent - the spatial extent of the shapefile (i.e. geographic area that the shapefile c
Object type - whether the shapefile includes points, lines, or polygons.
Coordinate reference system (CRS)
Other attributes - for example, a line shapefile that contains the locations of streams
```

Because the structure of points, lines, and polygons are different, each individual shapefile can only contain one vector type (all points, all lines or all polygons). You will not find a mixture of point, line and polygon objects in a single shapefile.

Simple text files are human readable file formats (.txt, .csv) that are suitable for storing point and attribute data. You will often encounter .txt or .csv files when working with weather data for instance (see Table). Coordinates (typically latitude and longitude) are stored in a text files along with the other attributes. We can bring this type of file into a GIS, but we need to convert the data to point features before we can display it.

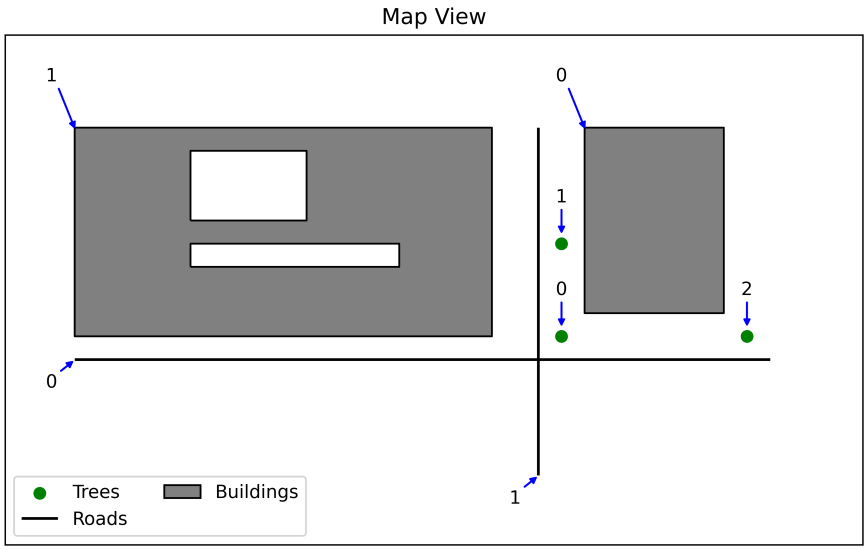
3.18 Canadian Weather Station File

Name	Province	Climate ID	Latitude (Decimal Degrees)	Longitude (Decimal Degrees)
ACTIVE PASS	BRITISH COLUMBIA	1010066	48.87	-123.28
ALBERT HEAD	BRITISH COLUMBIA	1010235	48.40	-123.48

Name	Province	Climate ID	Latitude (Decimal Degrees)	Longitude (Decimal Degrees)
BAMBERTON OCEAN CEMENT	BRITISH COLUMBIA	1010595	48.58	-123.52
BEAR CREEK	BRITISH COLUMBIA	1010720	48.50	-124.00
BEAVER LAKE	BRITISH COLUMBIA	1010774	48.50	-123.35
BECHER BAY	BRITISH COLUMBIA	1010780	48.33	-123.63
BRENTWOOD BAY 2	BRITISH COLUMBIA	1010960	48.60	-123.47
BRENTWOOD CLARKE ROAD	BRITISH COLUMBIA	1010961	48.57	-123.45
BRENTWOOD W SAANICH RD	BRITISH COLUMBIA	1010965	48.57	-123.43
CENTRAL SAANICH VEYANESS	BRITISH COLUMBIA	1011467	48.58	-123.42

GeoJSON

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "geometry": {
        "type": "Polygon",
        "coordinates": [
          [
            [100.0, 0.0], [101.0, 0.0], [101.0, 1.0],
            [100.0, 1.0], [100.0, 0.0]
          ]
        ]
      },
      "properties": {
        "prop0": "value0",
        "prop1": { "this": "that" }
      }
    }
```



Trees: Tabular View

index	Type	Age	Height	geometry
0	Fir	20	5.2	(11,11)
1	Cedar	120	58.4	(11,15)
2	Fir	31	8.9	(19,11)

Roads: Tabular View

index	Name	Type	geometry
0	Broadway	Avenue	(-10,10), (10,10), (20,10)
1	1st St.	Street	(10,5), (10,20)

Buildings: Tabular View

index	Address	geometry
0	100 Broadway	(12,20), (12,12), (18,12), (18,20), (12,20),
1	200 Broadway	(-10,20), (8,20), (8,11), (-10,11), (-10,20), [(-5,16), (0,16), (0,19), (-5,19), (-5,16)], [(-5,14), (4,14), (4,15), (-5,15), (-5,14)],

Figure 3.18: Vector objects (points, lines, or polygons) are stored along with any number of attribute. Point, line, and polygon data are typically stored in separate files. Each object in the map view is labeled with it's index. This value corresponds to the row in the attribute table and is useful for locating the attributes that correspond to the object.

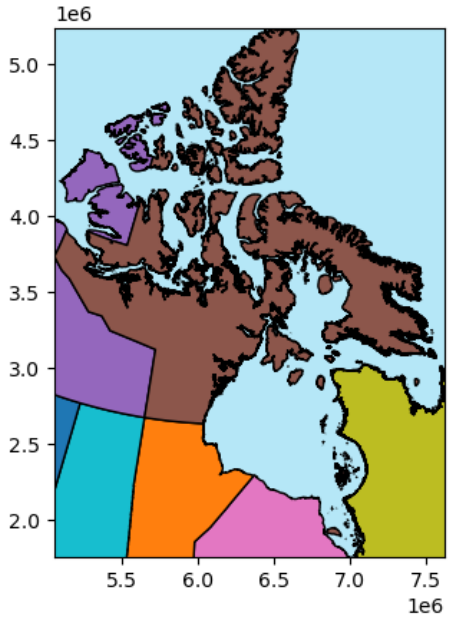


Figure 3.19: Nunavut is composed of many different islands. We don't need to represent every island as a separate object if we only want to

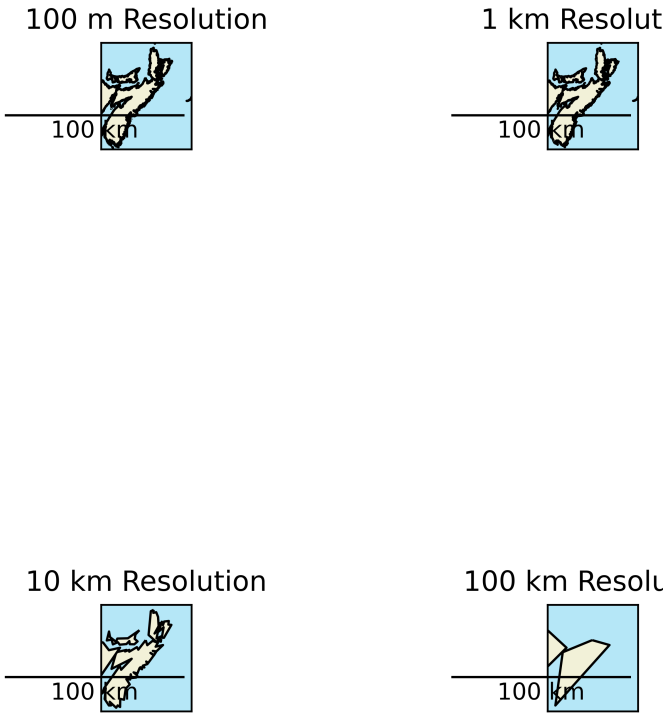


Figure 3.20: Vector resolution is defined a bit differently

```
  ]
}
```

3.19 Choice of Data Model

Vector data has some important advantages:

```
The geometry itself contains information about what the dataset creator thought was important
The geometry structures hold information in themselves - why choose point over polygon, for insta
Each geometry feature can carry multiple attributes instead of just one, e.g. a database of citie
Data storage can be very efficient compared to rasters
```

The downsides of vector data include:

```
potential loss of detail compared to raster
potential bias in datasets - what didn't get recorded?
Calculations involving multiple vector layers need to do math on the geometry as well as the attr
```

Vector datasets are in use in many industries besides geospatial fields. For instance, computer graphics are largely vector-based, although the data structures in use tend to join points using arcs and complex curves rather than straight lines. Computer-aided design (CAD) is also vector- based. The difference is that geospatial datasets are accompanied by information tying their features to real-world locations.

3.20 Case Study: Title of Case Study Here

You see textual case study content here

Police Involved Deaths in Canada

I'll flush this out in more depth later when I can dedicate some time to working on the actual project.Here is a link to the github org where I'll be hosting the data.

Data types used: 1) Nominal (Race) 2) Ratio (Counts & PKR) 3) Ordinal (Ranking) 4) Interval (PKDI)

Methods Discussed: 1) Normalization 2) Classification

Your browser does not support iframes

Data from City of Vancouver and licensed under the Open Government License - Vancouver

::::

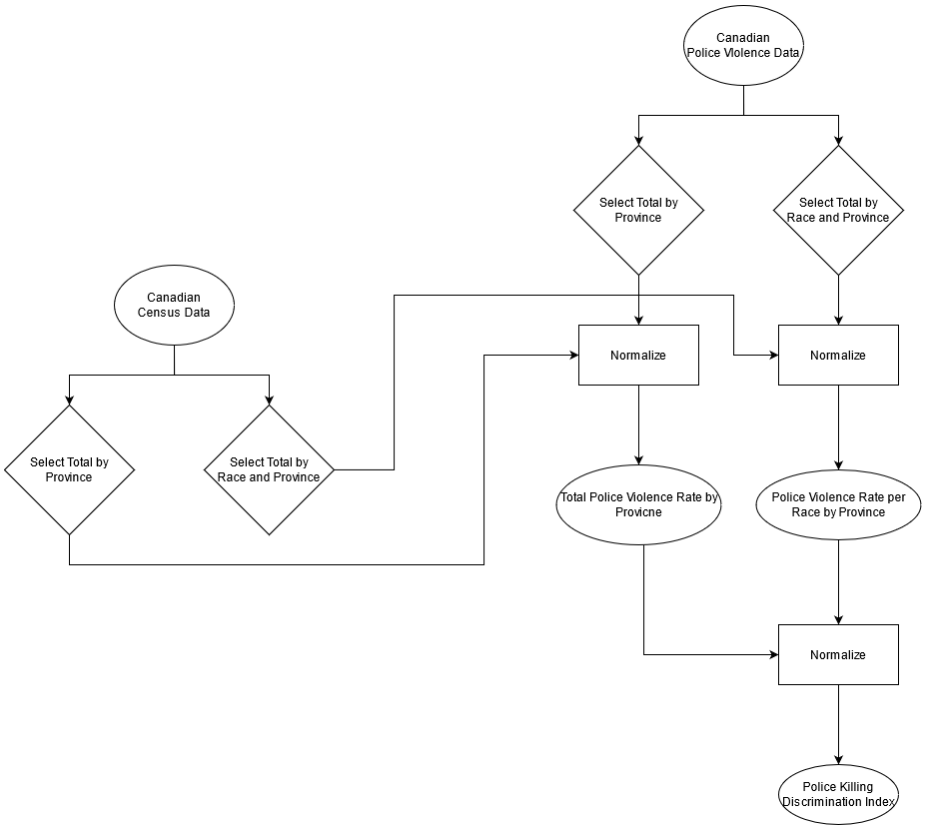


Figure 3.21: Rough flowchart draft

Your Turn!

I'll do some exercise building on the case study.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut in dolor nibh. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent et augue scelerisque, consectetur lorem eu, auctor lacus. Fusce metus leo, aliquet at velit eu, aliquam vehicula lacus. Donec libero mauris, pharetra sed tristique eu, gravida ac ex. Phasellus quis lectus lacus. Vivamus gravida eu nibh ac malesuada. Integer in libero pellentesque, tincidunt urna sed, feugiat risus. Sed at viverra magna. Sed sed neque sed purus malesuada auctor quis quis massa.

Call Out

This is a call out. Put some important concept or fact in here.

3.21 Summary

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut in dolor nibh. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent et augue scelerisque, consectetur lorem eu, auctor lacus. Fusce metus leo, aliquet at velit eu, aliquam vehicula lacus. Donec libero mauris, pharetra sed tristique eu, gravida ac ex. Phasellus quis lectus lacus. Vivamus gravida eu nibh ac malesuada. Integer in libero pellentesque, tincidunt urna sed, feugiat risus. Sed at viverra magna. Sed sed neque sed purus malesuada auctor quis quis massa.

Reflection Questions

1. Explain ipsum lorem.
2. Define ipsum lorem.
3. What is the role of ispum lorem?
4. How does ipsum lorem work?

Practice Questions

2. Given ipsum, solve for lorem.
3. Draw ipsum lorem.

Ensure all inline citations are properly referenced here.

Chapter 4

Collecting and Editing Data

The ability of a geomatician to answer research questions or produce a map or other visuals, rests, in part, on first finding the right data to do so. Geomaticians often spend much of their time finding, collecting, and editing data, yet the critical activity of finding data is often left as something that geomaticians are assumed to pick up along the way. This chapter addresses that gap by first introducing a range of possible data sources along with some theory, tips, and strategies to access them. We also address some common instances when data do not yet exist, and so we must create them. This chapter may be particualrly useful for students and researchers starting out on their spatial research projects, and for anyone interested in the rapidly changing data universe.

Learning Objectives

1. Become familiar with a wide range of spatial datasets and strategies to access them
2. Identify several sources of historical spatial information, including historical maps and aerial photos, and the steps required to analyze them as spatial information
3. …Editing data (Paul to change as desired write, or delete)
4. …GNSS (Paul to write)

Key Terms

aerial photography, area of interest, census, data repository, data request, georeferencing, GNSS, natural resource administrative data, historical collections, open data, orthophotos, spatial panel data.

4.1 Open data

Data is becoming increasingly easy to access thanks to the **open data** movement. The concept of **open data** suggests that governmental data should be available to anyone to use and, if desired, redistribute in any form without any copyright restriction (Kassen 2013) or, with minimal restrictions such as providing recognition.

Until recently, most government data were simply unavailable or could only be accessed by data request or by paying the government data provider. Countries around the world are moving to an open data model. For example, Britain is opening up its national geographic database (housed as the 'Ordnance Survey'). United States (US) has moved its data housed within the US Geological Survey into the public domain (USGS 2021). Canada has signed a Directive on Open Government, which promotes the proactive and ongoing release of government information. The province of British Columbia (BC) has just released all government LiDAR data under an open government license and many provinces and municipalities release data under similar licenses. Canada is also signatory of the Treaty of Open Skies, which is an international effort that encourages the sharing of aerial imagery to promote openness and transparency of each signatory nation's military forces and activities. Despite the tremendous momentum towards **open data**, many datasets are not yet fully open. The tips and strategies below will help locate both open and not so open datasets.

4.2 Finding Data

Here we introduce a network model to set a framework for finding data. Imagine that nearly all the data and information in the world is connected in some way through networks of information, composed of individuals, libraries, and institutions. The internet is an important component in this network, one we all use every day to answer questions. For example, me might ask Google: "what is the best lake in Canada to plan a summer holiday?" A common answer returned is 'Lake Louise, Alberta,' which is a stunning lake surrounded by tall Rocky Mountains, as well as hordes of tourists! If we asked this question to our friends – and maybe one happens to be an expert fisherman or fisherwoman - we may receive different answers including secret lakes that have not yet been discovered by tourists, or the best lake for fishing. Our friends can also consider our specific interests, suggest helpful resources (such as a lesser known forum on local fishing), and offer additional information about our query such as the best places on that lake to camp, where to fish on the lake, and what type of fishing gear to use. The point in this example is that there are different networks of information available to us, including formal networks of information organized on the internet and accessed by search engines as well as informal networks of individuals and experts who offer an additional strategy to connect us with the right information.

Data are becoming increasingly easy to discover through the use of **data repositories** (figure 4.1). Below we discuss the growing number (and centralization) of spatial data repositories, which can give access to academic, government non-governmental, international, and crowdsourced datasets. Here we introduce each type of repository and offer some hints at what environmental data can be discovered in each.

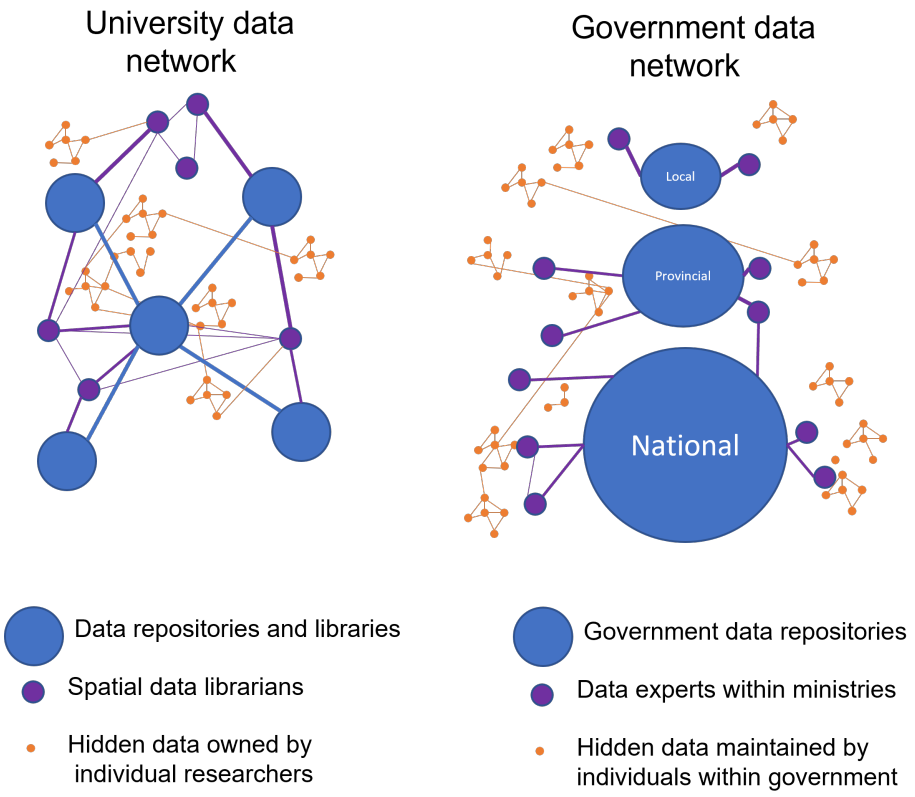


Figure 4.1: Figure 4.1 Envisioning university and government data networks. Data within each network is concentrated within data repositories, yet considerable data remains 'hidden' among individual researchers and silos of the Ministry, but can potentially be accessed by finding the right connections.

4.3 Data in Academia

Data librarians are particularly well connected and trained to help you navigate these repositories and contacting them can be a good starting point in your search. Nonetheless, considerable data is not yet published. Some of this unpublished data has been analyzed in previous research and its existence could be

discovered through a review of the academic literature. Other unpublished data remains essentially 'hidden,' only known about by individuals or small clusters of individual researchers who created those data. In such a case, your only possibility to find such data is through a combination of 'asking around' and reaching out to experts in the field. Once you know it exists, unpublished data could potentially be accessed through connecting with those researchers themselves, and requesting the data or inquiring about the possibility for a collaboration.

4.4 Government Data

Government data is also increasingly published in **data repositories**, specific to the level of government (figure 4.1). There are multiple levels to government, including municipalities (the smallest), provinces (or states), and nations (the largest), each of which often has its own **data repository**. Centralized repositories are becoming increasingly common and connect **open data** from all levels of Government. The Federated Research **data repository** is an aggregation of Canadian **open data** repositories, including municipal, provincial, and academic repositories. It includes a map-based search for datasets with location information tied to their metadata. In the US, geospatial data from federal, municipal, and state government repositories are being consolidated under [Data.gov](https://www.data.gov/](https://www.data.gov/).

Because not all repositories are yet connected by a centralized repository, one must search in teh correct repository. To do this, consider which government has jurisdiction over the specific subject area and geography of interest. For example, if you are interested in land use zoning and engineering features within a given city, this data is likely best provided by that individual city, either by finding it within a **data repository** or emailing the municipality with a data request (discussed below). In Canada, the provinces have jurisdiction over most natural resources and thus provincial government data repositories tend to provide data on natural resources, such as, water features, forests, wildlife, minerals, and topography. In British Columbia, for example, DataBC houses over a thousand datasets on natural resources, including forest cover mapping, natural disturbances, hunting statistics, administrative boundaries, and much more. Canada's **open data** portal provides data on fish as well as environmental conditions (e.g., water quality, air quality, historical weather, etc.), which is under federal jurisdiction. Hydrological flow and water quality monitoring is readily accessible across Canada through the Hydat database, which can be easily accessed through the R package called TidyHydat (Albers 2017).

Your Turn!

Try using a web search to find the government **open data** pages for your city, province/state, and nation. What kinds of data do you see (probably a lot!). Try searching for data within each of them that is related to your own research

interest.

4.5 Census Data

This section introduces the **census** at a cursory level before launching into the applied question of how to find **census** data for your spatial analysis, using the Census of Canada as an example.

Census generally refers to a complete count by government of a specific region's population by age, gender, language, income, housing and other demographic characteristics. **Census data** inform public policy, such as allocation of public funds, transportation network planning, and electoral area delineation. **Census data** also provide researchers with an opportunity to gain insight into the social and, to a lesser extent, environmental fabric of a country and are increasingly used in environmental and social-ecological research that aims to address social elements of environmental challenges (Tomscha et al. 2016, Biggs et al. 2021).

Census are typically conducted once every five years (e.g., Canada) or every 10 years (e.g., Unites States).

In addition to demographics, many nations survey information related to economics or specific industries, such as agriculture. For example, Canada's Census of Agriculture captures information on fertilizers, irrigation, livestock, farm types, and crop production across Canada. The Longform Census in Canada surveys additional questions but is only sent to a subset of the population, and the data from it are then estimated for the entire population.

A starting point to using **census data** in spatial analysis is to understand the geographic levels of census data, and then we address where the geography files and data can be downloaded.

4.6 Census of Canada Geographic Levels

To protect respondents' confidentiality, the individual data collected during **census** enumeration is obscured from the public. Thus, **census data** can only be accessed by researchers in the form of statistics aggregated to varying geographic levels. Knowing these geographic levels is key to accessing **census data**.

At the top of Figure 4.2 are Canada's provinces and territories, which are then divided into census divisions, which in turn are divided into census subdivisions. Census subdivisions correspond to municipalities, but also include Indian reserves, and 'unorganized areas.' These three areas (municipalities, Indian reserves, and unorganized areas) are also aggregated into census consolidated subdivisions, which offer a more consistent geographic unit for mapping large areas as compared to subdivisions themselves. Census subdivisions are divided

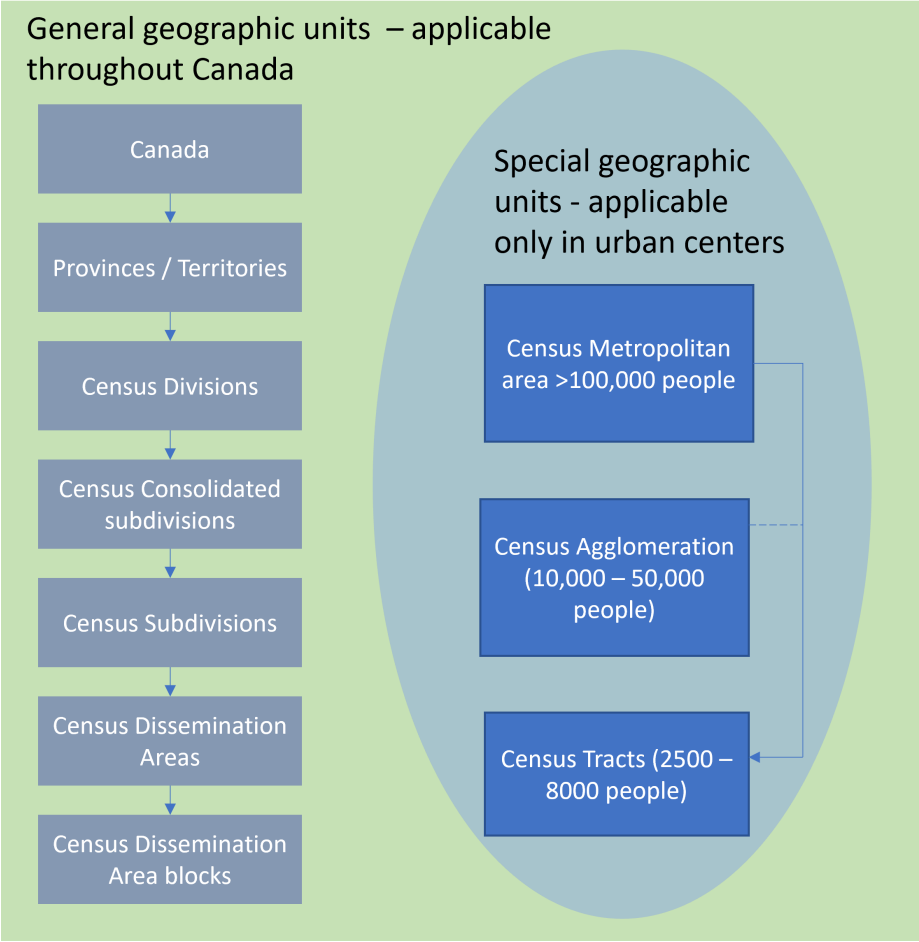


Figure 4.2: Figure 4.2 The geographic levels of the Census of Canada include general units (applicable everywhere throughout Canada) and also an additional layer for urban areas only.

into dissemination areas, composed of one or more 'dissemination area blocks' (generally, a city block bounded by roads on all sides).

In addition to these general geographies, which apply throughout Canada, special geographic units are implemented as an additional layer of aggregation for urban centers. A census metropolitan area" (CMA) is a grouping of census subdivisions comprising a large urban area and its surroundings. To become a CMA, an area must register an urban core population of at least 100,000 at the previous census. A census agglomeration (CA) is a smaller version of a CMA in which the urban core population at the previous census was greater than 10,000 but less than 100,000. CMA and CA are useful for making comparisons across cities. CMAs and CAs with a population greater than 50,000 are subdivided into census tracts which have populations ranging from 2,500 to 8,000 and are intended to be relatively homogeneous in their demographic identity (i.e., a local neighbourhood).

Using census data for geographic analysis typically involves first identifying the smallest spatial unit at which the data is available. Recall that to protect the privacy of respondents, some data is only available at higher geographic levels. Another consideration is that if you plan to compile multiple census years, the geographic boundaries have typically changed over time in response to how the landscapes and information needs have changed. This creates substantial (though, not insurmountable) additional work that limits how the data can be used, especially for finer spatial scale analysis. An example of changes in the geography of census divisions is seen for BC in figure 4.3

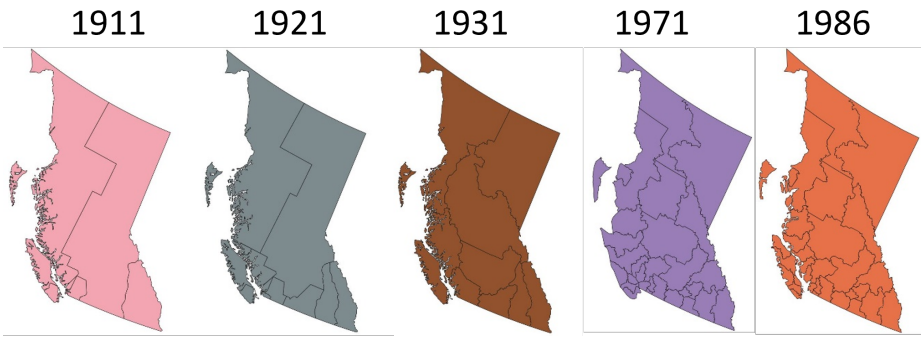


Figure 4.3: Figure 4.3 An example of how census boundaries have changed, showing changes in Census divisions for British Columbia from 1911 to 1986

Call Out

Spatial analysis will often want to work with the smallest geographic level available. The smallest geographic unit of the Canada **census** is the dissemination

block. Census tracts are also used frequently in spatial analysis but this geographic unit is only applicable to metropolitan areas.

4.7 Accessing Census Data

The geographic boundaries for the Census for each level can be downloaded as shapefiles for the 2016 **census** here.

The Canadian Socio-economic Information Management System Statistics Canada data portal provides access to the Census of Canada as well as the Census of Agriculture, Aboriginal Peoples Survey, and other government statistical datasets. You have the option to search by a vector or an area of interest. Students with access to CHASS Canadian Census Analyzer (students of University of Toronto as well as many other subscribing universities) can use CHASS to access additional statistical data, which they can aggregate to **census** geographic units of their choosing.

Your Turn!

Try this: Navigate to Canadian Socio-economic Information Management System Statistics Canada data portal and search a key word such as: "age." A list of available geographic levels should be present on the left side, allowing you to check which geographic levels you would like to retrieve the data for. What geographic levels are present for age and which is the smallest geographic level (refer to figure 4.3)? Now try searching the keyword: crop production. What is the smallest geographic level for crop production now?

4.8 Non-Governmental Organization Data

Many elements of the environment, such as biodiversity and large old trees, are not monitored by most governments. These knowledge gaps are sometimes filled by other organizations not associated with the government (i.e., non-governmental organizations) or by citizen science initiatives. For example, Pacific salmon have been a top conservation concern lacking data in western North America. An organization called the Pacific Salmon Foundation has collaborated with the help of First Nations and government to compile salmon information for BC so that the data can be readily viewed and downloaded for further analysis. Organizations such as the International Union for Conservation of Nature often synthesize and offer datasets that support their mandates such as monitoring species at risk and expanding protected areas.

4.9 Citizen Science

Citizen science describes activities where members of the general public contribute information and data to help generate new knowledge and informa-

tion (Lee et al. 2020). Citizen science has been used to fill in data gaps for widely distributed phenomenon that are otherwise difficult to gather. In addition to Open Street Map, which has created a free open geodatabase of the world, one of the most famous examples is a collective global effort to map the distribution of global bird species, which through an app called as [E-bird] (https://ebird.org/home) has generated nearly 1 billion bird observations as of 2021. Likewise, alpine wildlife are difficult for researchers to observe and are costly to study owing to the effort and risk associated with accessing alpine areas, yet may be frequently spotted by mountain climbers who venture into alpine areas during their recreational pursuits (Jackson et al. 2015). Citizen science is also used in fast-moving situations like natural disaster and to monitor long-term trends in the environment. For example, the British Columbia Big Tree Registry collates citizen science data on the locations of the largest trees in BC, thereby engaging citizens to help support policies to protect the largest trees in BC.

A useful starting point to check for citizen science datasets is Scistarter, which can be searched by keyword or location to identify citizen science projects around the globe. These datasets may be readily downloaded or downloaded through contacting the project leaders.

Your Turn!

If you were to start a citizen science project to capture environmental data to inform public policy, what kind of information would you try to capture?

4.10 International Data

Some research questions extend beyond borders. For example, oceans are primarily international and data on oceans can be searched through the Ocean Biodiversity Information System. A database on food production and timber is published by the United Nations Food and Agricultural Organization. Academic research that attempts to answer environmental problems at the global scale now often publishes their datasets for open use, such as the global tree canopy height map by Potapov et al. (2021).

4.11 Unpublished Data and the Data Request

Governments manage a wide variety of data, which is sometimes located in relatively siloed ministries and departments. Datasets that are not readily accessible online, may still exist and can potentially be retrieved through a data request to the appropriate government agency. In the spirit of **open data**, many governments are becoming increasingly responsive to data requests, but success of this approach often hinges on connecting with the right person that

may be able to help you. This requires networking.

While this textbook is primarily centered on technical skills, it is worth considering the old adage that "it is not what you know, but who know." Accessing data that is not readily available adds extra challenge but can reward you with new research and networking opportunities that can be highly beneficial for both parties. The data provider may benefit from the knowledge gained from your proposed research. They may be able to assist you with understanding the data, disseminating the final report, and even connecting you with job opportunities and other ways to continue your skill development.

When sending a data request or data query, always be respectful of their time, and be tactful. A data request template is provided below:

1. Dear … (person, or institution)
2. State your name and affiliation (e.g., University department and program/supervisor)
3. Briefly state your intended research or research aspiration (1-2 sentences)
4. State your **data inquiry** (e.g., do you know if x data exists?) or **data request** in bold text. Although you may not know exactly what you are looking for, try to be as specific as possible on the type of data you are requesting by describing. Give your geographical area of interest if known either descriptively, in a map, or as a shapefile.
5. Thank them for considering your request.
6. If you do not hear back from them within 1-2 weeks reply back with another, much shorter email (e.g., *I'd like to follow up and ask if someone in your office may be able to respond to the above data request?*)

Always be patient and remember that the individual you contacted is busy and may appreciate a reminder in case your first email slipped through.

4.12 Metadata

<to be written by Evan: at cursory level on how to collect and find metadata, why its important, how it can lead to other findings. >

4.13 Historical Data Collections

Historical data collections generally include any spatial datasource excluding satellite-based remote sensing that was produced prior to the widespread implementation of GIS in the mid 1990's. **Historical data** are typically not available as ready-to-use digital layers, and thus work is required up front to digitize them in preparation for spatial analysis.

Historical Datasets can be extremely valuable in environmental research because they extend our ability to observe how the environment has changed over

the longterm, potentially revealing vastly different landscapes and environmental conditions from those seen today. This insight can help remind us of levels of degradation or abundance that have become 'forgotten' by today's environmental managers, and can lead to surprising discoveries (McClenachan et al. 2015).

Although **historical datasets** can be very useful, they were often not collected for the intended purpose of being analyzed by future researchers. Data were often collected to serve the needs of the day, and were collected in a cost effective manner using tools and science that were available at that time. While this is less an issue for Census data, which has in some cases used relatively consistent survey questions through time, it complicates use of other datasets such as historical forest inventories, which have evolved their methods in step with technology and changing perceptions of how the forest ought to be monitored and valued. Thus, knowledge of how **historical data** were collected is sometimes required to accurately understand and interpret it. Overall, the process of locating, digitizing, and interpreting historical data can be a substantial portion of the work in a historical spatial analysis. In this section we cover historical **aerial photograph** collections, historical **natural resource administrative data** as well as **historical maps**.

4.14 Historical Aerial Photographs

The advent of **aerial photographs**, which are photographs of the Earth's surface taken from above (generally from an airplane), greatly improved mapping beginning in the 1930's and became the primary source of data for mapping land cover, timber volumes, topography, and national defense planning. Today, they offer a valuable tool for the unique spatial and temporal resolutions they offer. Temporally, **aerial photos** offer snapshots of landscapes that predate satellite-based remotely-sensed data by many decades (Morgan et al. 2017), which can help inform restoration targets and cumulative effects assessments (Harker et al. 2021). Aerial photos vary in their spatial resolution, but sometimes offer a surprisingly high spatial resolution that can be used to study fine-scale landscape attributes and their changes, such as stream courses (Little et al. 2013), fish habitat (Tomlinson et al. 2011), and soil hydrodynamics (Harker et al. 2021).

Using **aerial photographs** to track landscape change often requires first 'tying' them to the Earth to produce and orthophoto, a process discussed as it applies generally to image processing in Chapter 13 and discussed briefly here. An orthoimage is an aerial photograph or satellite imagery geometrically corrected so that the scale is uniform, such as in figure 4.4. Unlike orthoimages, the scale of ordinary aerial images varies across the image, due to the changing elevation of the terrain surface (among other things). The process of creating an orthoimage from an ordinary aerial image is called orthorectification. Photogrammetrists are the professionals who specialize in creating orthorectified aerial imagery, and in compiling geometrically-accurate vector data from aerial images.

Compare the map and photograph below. Both show the same gas pipeline, which passes through hilly terrain. Note the deformation of the pipeline route in the photo relative to the shape of the route on the topographic map. Only the topographic map is accurate here. The deformation in the photo is caused by relief displacement. The photo would not serve well on its own as a source for topographic mapping.

Figure 4.4: Figure 4.4 Example of how a linear feature can appear crooked in an aerial photograph that has not yet been orthorectified due to relief displacement.

Even in their un-orthorectified state, historical aerial photos can offer a powerful communication tool. They offer a window into historical landscapes that can be easily discerned and appreciated by viewers. Thus, even without orthorectification and performing spatial analysis, historical aerial photos can enrich a research report and other communications.

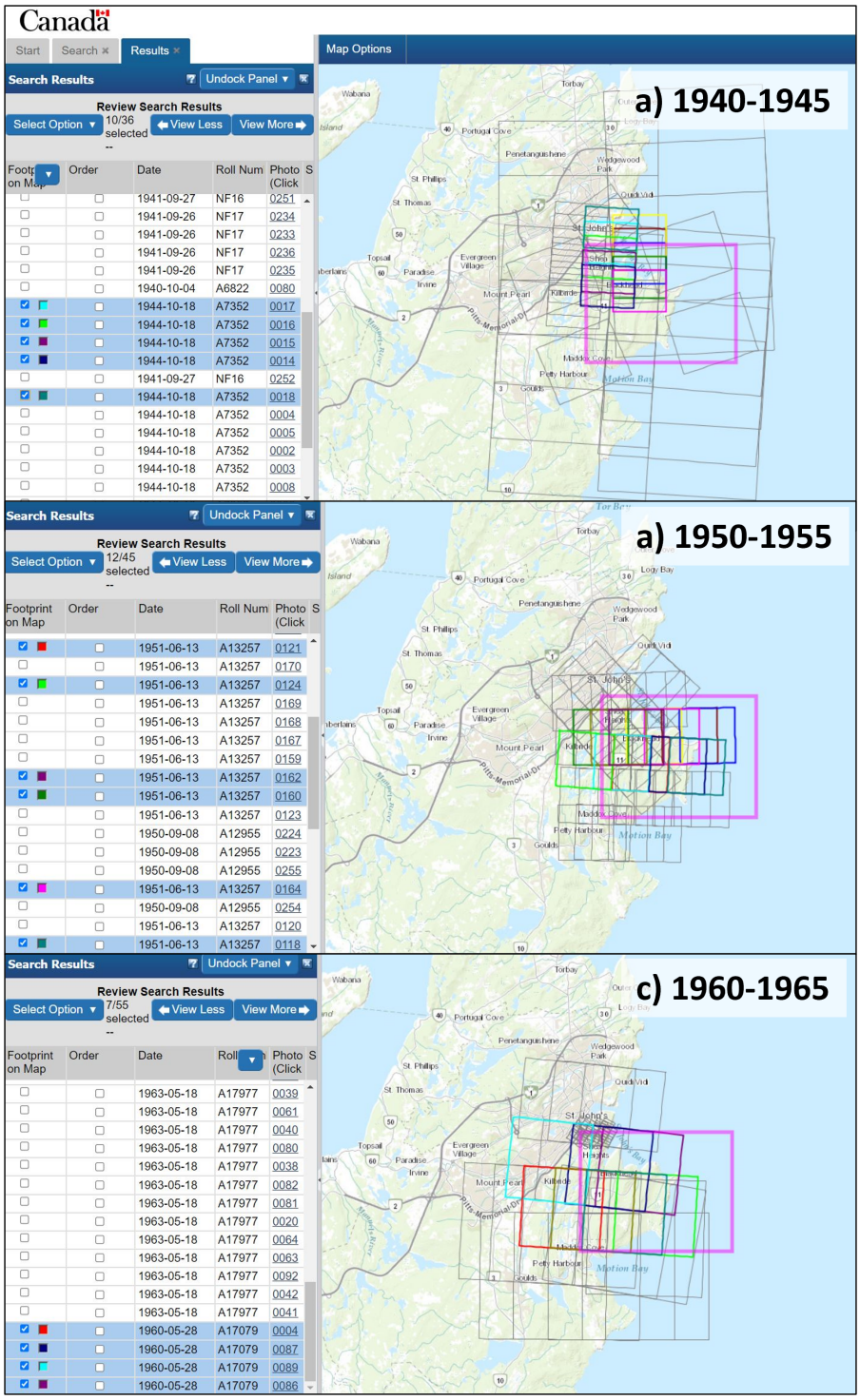
4.15 Accessing Historical Aerial Photograph Collections

Aerial photography missions involved capturing sequences of overlapping images along parallel flight paths. A flight path produces a 'roll' of numerous adjacent images that overlap. Flight paths tend to be here and there, but not necessarily exactly where you need them! Therefore, the first step is to determine the availability of historical photographs rolls for your timeframe and **area of interest**. Some collections can be searched relatively easily using a web-based GIS. For example, the Canada National Air Photo Library has a collection of roughly 6 million aerial photos some dating back to the 1920's, which can be searched using the Earth Observation Data Management System.

A search generally follows these steps:

1. Determine your **area of interest**.
2. Decide on the timeframe of interest.
3. Search via a GIS web map or paper flight line maps and examine which flight rolls cross over your timeframe and **area of interest**.

Figure 4.5 shows the results from an example search. In this example, the **area of interest** (large pink rectangle, figure 4.5) was set by navigating to the study site within the web map then setting the current extent as the **area of interest**. Here the extent is centered on the coastline between St. John's, Newfoundland and Cape Spear, the most easterly point in North America. We then searched for **aerial photographs** at three different timeframes: 1940-1945 (figure 4.5, panel A), 1950-1955 (figure 4.5, panel B), and 1960-1965 (figure 4.5, panel C). Indeed, aerial photos were found to be available at each period. The photos with smaller boxes (or foot prints) tend to have higher spatial resolution but cover less area. Assuming that fine spatial resolution is desired, the smallest photos have been selected in this example and could then be requested from the Library. Previews are often not available so we will not fully know the quality of the photos until we inspect them.


Canada
Start
Search
Results
Search Results
Undock Panel
Review Search Results
Select Option
10/36 selected
View Less
View More
Footprint on Map
Order
Date
Roll Num
Photo (Click
1941-09-27 NF16 0251
1941-09-26 NF17 0234
1941-09-26 NF17 0233
1941-09-26 NF17 0236
1941-09-26 NF17 0235
1940-10-04 A6822 0080
1944-10-18 A7352 0017
1944-10-18 A7352 0016
1944-10-18 A7352 0015
1944-10-18 A7352 0014
1941-09-27 NF16 0252
1944-10-18 A7352 0018
1944-10-18 A7352 0004
1944-10-18 A7352 0005
1944-10-18 A7352 0002
1944-10-18 A7352 0003
1944-10-18 A7352 0008
Map Options
a) 1940-1945
Search Results
Undock Panel
Review Search Results
Select Option
12/45 selected
View Less
View More
Footprint on Map
Order
Date
Roll Num
Photo (Click
1951-06-13 A13257 0121
1951-06-13 A13257 0170
1951-06-13 A13257 0124
1951-06-13 A13257 0169
1951-06-13 A13257 0168
1951-06-13 A13257 0167
1951-06-13 A13257 0159
1951-06-13 A13257 0162
1951-06-13 A13257 0160
1951-06-13 A13257 0123
1950-09-08 A12955 0224
1950-09-08 A12955 0223
1950-09-08 A12955 0255
1951-06-13 A13257 0164
1950-09-08 A12955 0254
1951-06-13 A13257 0120
1951-06-13 A13257 0118
a) 1950-1955
Search Results
Undock Panel
Review Search Results
Select Option
7/55 selected
View Less
View More
Footprint on Map
Order
Date
Roll
Photo (Click
1963-05-18 A17977 0039
1963-05-18 A17977 0061
1963-05-18 A17977 0040
1963-05-18 A17977 0080
1963-05-18 A17977 0038
1963-05-18 A17977 0082
1963-05-18 A17977 0081
1963-05-18 A17977 0020
1963-05-18 A17977 0064
1963-05-18 A17977 0063
1963-05-18 A17977 0092
1963-05-18 A17977 0042
1963-05-18 A17977 0041
1960-05-28 A17079 0004
1960-05-28 A17079 0087
1960-05-28 A17079 0089
1960-05-28 A17079 0086
c) 1960-1965

search for available rolls In either the paper or digital version display the extent of available photographs captured side by side.

Your Turn!

Go to the Canada Eartch Observation Data Management System and search for historical aerial photos in your chosen area of interest using the timeframes 1935-1950 and then 1950-1980. What is the oldest photo available?

If you searched but did not find anything helpful, don't be discouraged. The **area of interest** in the example of Cape Spear, Newfoundland, happens to be a strategic location for national defense so it not surprising that it has excellent coverage in the National Air Photo Library. In contrast, if you are interested in seeing an environmental feature such as historical forest cover in northern BC, recall that natural resources fall under the jurisdiction of provinces in Canada. Consequently, provinces may house aerial photo collections for your area. Some of these collections have been preserved by government or other institutions, such as the Geographic Information Center (GIC) at the University of British Columbia, which rescued a collection of 2.5 million aerial photos. These photos are available for researchers and commercial use. The GIC also maintains a list of other aerial photograph libraries, including for Alberta, Yukon, and the United States.

4.16 Natural Resource Administrative Data

Governments often conduct ecological and economic monitoring in their efforts to inform public policy and environmental management. Herein, this data is referred collectively to as natural resource administrative data. This data includes information collected during the process of administering natural resources use, such as to calculate fees, royalties, and licensing payments that the resource users must pay to the government for the use of public natural resources. Administering natural resources also requires monitoring data to spatially allocate harvest quotas on resources such as fish, big game, and timber. As opposed to remotely sensed data, this type of data often describes the actual amounts of natural resources available or used, and sometimes the number of users, who those users are, and what types of dependency they may have on the resources (e.g., their levels of income).

These data often come in a form called spatial panel data. Spatial panel data describe time series associated with particular spatial units (e.g., cities, wildlife management units, timber harvesting areas). Using spatial panel data typically requires:

1. downloading (or digitizing, if necessary) the statistical data as a spreadsheet
2. downloading the spatial geometry file

3. Linking the two files using an attribute join (chapter 5).

An example of a marvelous and yet relatively easy to use natural resource administrative data record is the BC big Game Hunting Statistics, which documents the number of large game hunted in BC by species, by hunter type (BC resident vs. non-resident hunter), and the effort (# days) that went into the hunts. This data can be made spatial by performing an attribute join with the BC Wildlife Management Units Layer. Attribute joins are discussed in chapter 5?).

Many natural resource administrative records are in digital form back to about 1980. Before that data often only exists in archival documents and must be digitized. Libraries are actively digitizing important archives, such as government annual reports, which are a rich source for natural resource administrative data.

4.17 Historical Maps

People have collected spatial information and mapped the world since long before GIS or aerial photos existed. Efforts are underway to preserve and digitize historical maps, and some collections are readily accessible. For example, insurance maps are maps made by insurance companies who mapped buildings, industrial complexes, and neighbourhoods to administer insurance policies since the late 1800's (e.g., for BC). Forest cover mapping became common in the early to mid 1900's (though, the early maps rarely survived) to estimate timber volumes. Natural disturbance mapping also became widespread in the early 1900's and considerable work has already been done to digitize and turn those data into readily usable forms (e.g., for wildfire and insect disturbance in BC). Land surveys dating back to the mid 1850's have also been used to systematically map historical forest cover, land ownership, and linear features such as roads (Tomscha et al. 2016).

Geographers recognize that all maps are subjective and **historical maps** are thus sometimes studied to understand how historical landscapes were perceived by society, revealing potential social biases and political orientations of who commissioned or created the map. This treads into the social sciences and humanities disciplines, which can offer additional and important ways to understand land management challenges today. For example, historical geographers have studied the history of fur trapline mapping because it offers insight into how First Nations traditional territories were ascribed into a form of information that could fit with the worldview of colonial governments (Iceton 2019). Thus understanding the transcription of these areas into maps which happened a century ago may help inform the complex spatial problem of how First Nations rights and titles to their traditional territories can be addressed in treaty negotiations and reconciliation.

4.18 Georeferencing Historical Maps

Although many types of data seem to come automatically **georeferenced**, such as photos taken from a modern mobile phone, other information must be first processed into a form that can be analyzed by the geomatician. This is especially true for any data captured prior to when gps became common in the 2000's. For example, decades and sometimes centuries of data exist in the form of herbaria, ship logs, and tree ring records that offer salient information on the spatial distribution of biodiversity and natural processes. This information cannot readily be brought into a GIS. The solution is **georeferencing**, which is a process to assign non-spatial information a spatial location (x and y coordinates) based on a coordinate system. Here we discuss **georeferencing** as it applies to historical maps. To supplement this section, general theory is provided about **georeferencing** aerial images in Chapter 13.

A common use case for georeferencing in landscape studies is when a historical map must be brought into GIS and overlaid with other data. Imagine you have a paper map and you use a desktop scanning device to scan it and save it as a digital image - this map depicts a particular area on Earth but there is no way for your computer to where and how on Earth to place this map (figure 4.6). In order to solve this problem, it is necessary to assign it geographic coordinate information so that GIS software can correctly align it with other georeferenced data.

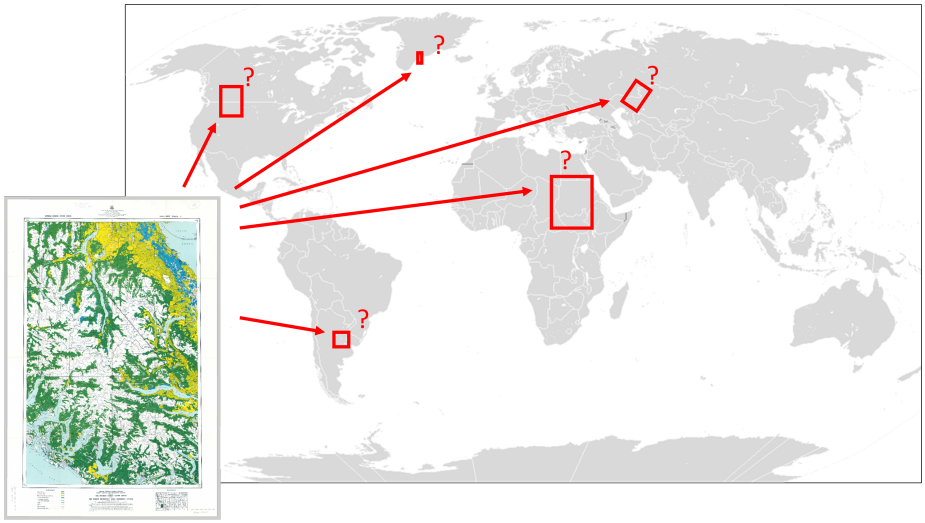


Figure 4.5: Figure 4.6 The need for **georeferencing** illustrated conceptually

Georeferencing is typically carried out using GIS software like QGIS, ArcMap, or ArcGIS Pro. The process of **georeferencing** varies slightly based on the GIS software you are using and the characteristics of the raster data you are

working with, but the case study below provides a generalized workflow to help learn the overall process. Two important aspects are placing control points and rubbersheeting.

Control points are the locations on the map that we will use to tie our historical map into a coordinate system. Control points should be spaced evenly across the the map. There must be at least 3 control points, but preferably more (e.g., >10). Control points should be spaced relatively evenly to obtain a good rendering. Two options are discussed for control points

4.19 Control Points on Maps with Grids or Graticule.

Large area maps (e.g., an entire country or province) typically have graticule, which depict lines of latitude and longitude, and smaller scale maps often have UTM grids. These grids or graticule may span across the map, or just be located along the corner or edges of a map. Such maps can often be georeferenced in a GIS by first setting the desired coordinate system and then toggling on the grid or graticule within the GIS. Control points can be placed on the scanned raster at the line intersections than tied to the grid toggled on in the GIS. Here is a [guide to **georeferencing** by map corners using QGIS] (https://guides.lib.utexas.edu/georeference-raster-data/qgis-georeference-by-map-corners)

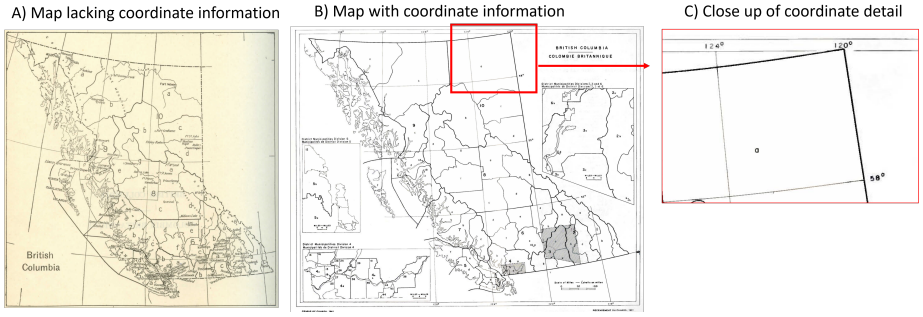


Figure 4.6: Figure 4.7 A comparison of A) a historical census map from 1931 with no graticule versus B) a 1961 census map with graticule representing latitude and longitude. Panel C) shows a close-up of the coordinate detail.

4.20 Grid and Graticule as Control Points

Not all maps have geographic coordinates on the map or along its corners (Panel A, figure 4.7). For such maps, control points must be placed on geographic features that can be linked to a base map that is already georeferenced and shows the locations of these features. Geographic features should be stable over

time. For example, an ideal geographic feature is an island or cape in the ocean, or a mountain top. Be aware that many features do change over time: rivers meander, lakes are sometimes flooded by dam construction, and houses or other landmarks can be moved. In urban areas, try to identify features that have not changed over time. If using roads, use the center of road intersections. Here is a guide to **georeferencing** by map features using QGIS

4.21 Rubbersheeting

Once the control points are set, a transformation is applied to mold the historical map as best as possible into GIS space. The practice of using georeferencing historical maps using control points and transformations is an example of rubber sheeting. In cartography, rubbersheeting refers to the process by which a layer is distorted to allow it to be seamlessly joined to an adjacent geographic layer of matching imagery. This is sometimes referred to as image-to-vector conflation. Often this has to be done when layers created from adjacent map sheets are joined together. Rubber-sheeting is necessary because the imagery and the vector data will rarely match up correctly due to various reasons, such as the angle at which the image was taken, the curvature of the surface of the earth, minor movements in the imaging platform (such as a satellite or aircraft), and other errors in the imagery. A variety of transformations can be used during rubber sheeting. You should test a few to see how they work then choose one, which appears to produce the most satisfactory results in terms of the visual fit and lowest amount of error, measured as the root mean square error (RMSE, discussed in chapter 13).

If you are rubber sheeting multiple maps, it may be beneficial to use a consistent transformation to facilitate writing up your methods and communicating your research.

4.22 Documenting Georeferencing

During the process of georeferencing you must document the number of control points and the root mean square error (RMSE). Although there are multiple sources of uncertainty in the spatial precision of a historical map, uncertainty should be characterized where possible to demonstrate rigour in your methods and for communicating uncertainty.

4.23 Data Transformations

To be written by Paul

4.24 Affine

4.25 Similarity

4.26 Projective

4.27 Reflection Questions

1. What are the key levels of Government where you live, and what kind of spatial data might each one manage?
2. What are two ways to find unpublished spatial data that is owned by a researcher?
3. What are the different types of data repositories where you can access spatial information?

4.28 Practice Questions

1. Try the case study on **georeferencing** a historical map. Record the number of control points placed, the RMSE, and the transformation use.
2. Draft a data request for a shapefile of bus routes as well as bus ridership statistics for the previous year in your hometown.

4.29 Summary

Data is becoming increasingly accessible thanks to the open data movement, but one must still need to know where to find it. The search for data, whether social, environmental, or economic in nature, is facilitated by data repositories as well as informal approaches such as networking with colleagues, consulting data librarians, and reaching out to experts in your subject area. When data does not exist, we can sometimes create it. Historical data such as aerial photos, natural resource administrative data, and historical maps must often by digitized into a form useable for spatial analysis. However, this effort can be worth while for researchers interested in history and for the unique information gained on social and ecological change.

GNSS and data transformations…

4.30 References

Statements on **open data**, aerial photos, and comparisons of government data among countries adapted partially from Biase). Georeferencing images and text adapted from University of Texas Libraries (2021). Rubber sheeting text is adapted from Wikipedia.

Albers S (2017). "tidyhydat: Extract and Tidy Canadian Hydrometric Data." *The Journal of Open Source Software*, 2(20). doi: 10.21105/joss.00511, http://dx.doi.org/10.21105/joss.00511.

Chen, K. (2002). An approach to linking remotely sensed data and areal census data. International Journal of Remote Sensing, 23 (1), 37-48

DeBiase, D. Nature of Geographic Information Systems. https://opentextbc.ca/natureofgeographicinformation/chapter/1-overview-2/

Biggs, R., de Vos, A., Preiser, R., Clements, H., Maciejewski, K., & Schlüter, M. (2021). The Routledge Handbook of Research Methods for Social-Ecological Systems (p. 526). Taylor & Francis

Iceton, G. (2019). "Many Families of Unseen Indians": Trapline Registration and Understandings of Aboriginal Title in the BC-Yukon Borderlands. BC Studies: The British Columbian Quarterly, (201), 67-91.

Jackson, M. M., Gergel, S. E., & Martin, K. (2015). Citizen science and field survey observations provide comparable results for mapping Vancouver Island White-tailed Ptarmigan (Lagopus leucura saxatilis) distributions. Biological Conservation, 181, 162-172

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. Government information quarterly, 30 (4), 508-513.

Lee, K. A., Lee, J. R., & Bell, P. (2020). A review of Citizen Science within the Earth Sciences: potential benefits and obstacles. Proceedings of the Geologists' Association.

Little, P. J., Richardson, J. S., & Alila, Y. (2013). Channel and landscape dynamics in the alluvial forest mosaic of the Carmanah River valley, British Columbia, Canada. Geomorphology, 202, 86-100

McCLENACHAN, L. O. R. E. N., Cooper, A. B., McKENZIE, M. G., & Drew, J. A. (2015). The importance of surprising results and best practices in historical ecology. BioScience, 65(9), 932-939

Morgan, J. L., Gergel, S. E., Ankerson, C., Tomscha, S. A., & Sutherland, I. J. (2017). Historical aerial photography for landscape analysis. In Learning Landscape Ecology (pp. 21-40). Springer, New York, NY

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., … & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165.

Rubbersheeting. (2020, June 11). In Wikipedia. https://en.wikipedia.org/wiki/Rubbersheeting

Tomlinson, Matthew J., et al. "Long-term changes in river–floodplain dynamics: implications for salmonid habitat in the Interior Columbia Basin, USA."

Ecological Applications 21.5 (2011): 1643-1658.

Tomscha, S. A., Sutherland, I. J., Renard, D., Gergel, S. E., Rhemtulla, J. M., Bennett, E. M., … & Clark, E. E. (2016). A guide to historical data sets for reconstructing ecosystem service change over time. *BioScience*, *66*(9), 747-762

University of Texas Libraries. (2020, July 26). Intro to Georeferencing. https://guides.lib.utexas.edu/georeference-raster-data

USGS "https://www.usgs.gov/information-policies-and-instructions/copyrights-and-credits%22;

Chapter 5

Relational Databases

You have almost certainly used a relational database in some form during your life, probably without even realizing it. Relational databases are foundational for information management in a GIS. In this chapter, we will look at the formal construction of relational databases, how they are used across a wide range of fields, and how we can use them to analyze spatial and aspatial information for environmental management.

Learning Objectives

1. Identify the purpose of Relational Database Management Systems in GIS
2. Describe the elements of relational databases
3. Practice applying relational algebra and Boolean logic to relations
4. Recognize the uses of different keys for joining and relating information
5. Understand how to query relational databases in order to extract or produce new information

Key Terms

Relational Database Management Systems, Tables, Relations, Rows, Tuples, Records, Columns, Attributes, Items, Structured Query Language, Boolean Logic, Relational Algebra, Entity-Relationship Model, Cartesian Product, Schema, Unary, Binary, Georelational Data Model, Domain, Symmetric Difference

5.1 Relational Database Management Systems

Suppose you have collected some data about some trees. You might have organized these data into a table, where each row represents a different plot, and each

column represents some quantitative or qualitative measure about each record. How do you *manage* these data in order to extract useful information from your trees? This is where Relational Database Management Systems can help. A **Relational Database Management System (RDBMS)** is a software that allows the user to interact with tabular data. The basic services provided by a RDBMS include storing, querying, and manipulating relational databases. We say the databases are **relational** because they are based on a relational model first developed by Edgar Codd in the 1970s at IBM. The relational model for database management is distinguished from non-relational models by the fact that data are stored in highly structured tables instead of some other format like documents. This distinction is important, because the vast majority of GIS software utilize the relational model for database management.

5.2 Relational Databases

Within a RDBMS, we find **relational databases**, which are highly structured tables comprised of rows and columns. In fact, a table in a relational database is called a **relation**, a row is a **tuple**, and a column is an **attribute**. Relational databases are a great way to store simple data structures that can be organized into a relation with tuples and attributes. When we say that a table or relation is "structured", we are referring to the fact that the data are organized according to a database **schema**, which is a set of constraints that ensure data integrity and consistency. For example, our set of trees likely all contain the same types of information and this can be easily organized into a relation. Suppose we measured the height, diameter at breast height (DBH), and species of each tree, then our relation would look like Figure **??**.

As you can see from the example above, there are two components to geospatial data: the tabular data containing tuples and attributes and the spatial data that contain the coordinate pairs for a projected or geographic coordinate system. This structure is known generically as the **georelational data model**. Many formats of geospatial data conform to the georelational data model, which stores a relation of tuples and attributes separately from another relation containing the geometry and coordinates. These two tables are then dynamically related to one another in a RDBMS using GIS software. You will almost never interact or see the relation that stores the geometry and coordinates of features contained in a relational database. Instead, the GIS software manages those files in the background for the purpose of displaying a set of features on a map, and you primarily interact with the tabular data stored in the relation of tuples and attributes.

The schema for the very simple example above would include the constraint and expectation that when we retrieve the height of a particular tree from the relation, it will be returned to us as an integer number and not a date. This logic is extended to all attributes so that types of values are never mixed and values are never unexpectedly changed by any database operation. That is to

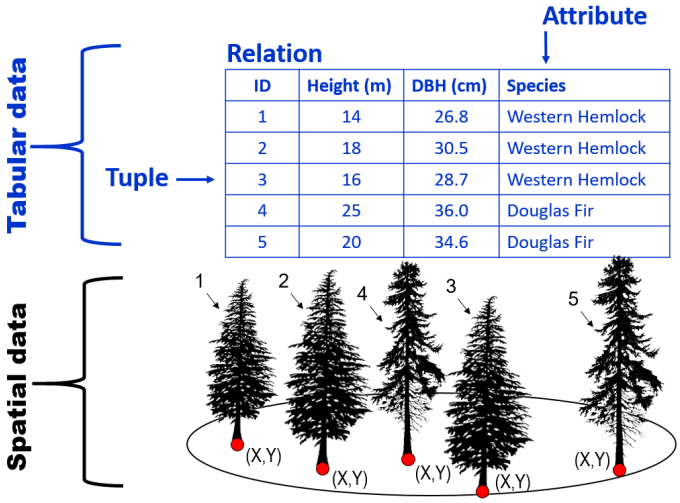
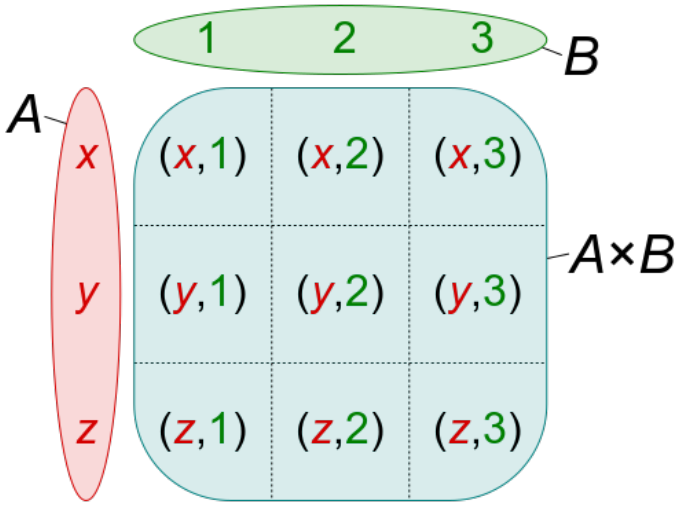


Figure 5.1: Tabular and spatial data are related by a Relational Database Management System (RDBMS) in a Geographic Information System (GIS). Images of Douglas-Fir and Western Hemlock trees by Natural Resources Canada, Canadian Forest Service, modified with permission.

say, we can and often do intentionally change values in a relation, but any new values must conform to the database schema for a particular attribute, which may also be constrained by a range and type of potential values, known as an attribute **domain**.

More formally, a relation R is a *subset* of two sets, A (tuples) and B (attributes). The product of these sets $A \times B$ is called the **Cartesian product**. In the same way that Cartesian coordinates are ordered pairs of values from two axes, the Cartesian product of two sets gives us an ordered pair of elements (a, b) from sets A and B, where a is an element in the set A, written as $a \in A$, and b is an element of set B, written as $b \in B$. Therefore, R is both the Cartesian product as well as any subset of $A \times B$.



\begin{figure}
\caption{Cartesian product A×B of A (tuples) and B (attributes). Image by Quartl, CC BY-SA 3.0.} \end{figure}

There are some important rules to be followed for organizing data into a relation:

1. Each tuple must share the same attributes as all the other tuples;
2. Each attribute has a unique name and is of the same *type* of data (i.e., integer, floating-point decimal, text, date, boolean, etc.);
3. The order of tuples and attributes can be rearranged without changing the meaning or integrity of the data;
4. Each value of an element in a relation (i.e., combination of tuple and attribute) is *logically* accessible; and
5. Each tuple is unique (i.e., no duplicate observations).

If any of the above rules are broken, then $R \neq A \times B$ and you are just looking at a plain-old table instead of a relation. In fact, Codd described a total of 13 rules for a RDBMS, but since this chapter is only a cursory introduction of RDBMS for GIS, you only need to be familiar with the five rules above. In this way, relational databases are comprised of relations that are highly structured by a schema, which allows the user to query, retrieve, update, and delete data using a RDBMS. At this point, you should understand that relational databases are highly structured so that we can apply logical expressions and languages to interact with the information contained within and between the relations. In the next two sections, we will look at how to apply two branches of mathematical logic to relations in order to extract useful information.

5.3 Relational Algebra

One of the fundamental jobs of a RDBMS is to apply relational algebra operations to relations stored in a relational database. Remember that we defined

a relation as $R = A \times B$ and that any subset of $A \times B$ is also a relation. This transitive property of relations combined with the fact that relations are just sets allows us to apply set algebra. In other words, relational algebra operations take one relation as input and produce a new relation as an output without modifying the input relation. This new output relation can then be used as an input to another operation because it is also a relation.

5.4 Selection

Selection is the simplest operation to understand and is probably the most-used in day-to-day GIS work. It does exactly what it sounds like, it retrieves a subset of a relation given some predicate or condition. For example, we could select all tree IDs from our relation R in Figure **??** that have a height greater than 20 m. This would yield tree ID=5. Formally, selection is expressed as $\sigma_{predicate}(R)$ and the example above would be written as $\sigma_{height>20}(R)$, which evaluates to the following:

ID	Height (m)	DBH (cm)	Species
4	25	36	Douglas-Fir

5.5 Projection

If selection is understood to operate on attributes to return tuples, then **projection** is an operation on tuples to return attributes. For example, suppose we are only interested in the height and DBH attributes for the trees. We would use projection to return this new subset of the relation. Formally, projection is expressed as $\Pi_{predicate}(R)$. Both projection and selection are referred to as **unary** operators because they only require a single relation as input. The example above would be expressed using the attributes that we want to preserve, so $\Pi_{height,dbh}(R)$, which evaluates to the following:

Height (m)	DBH (cm)
14	26.8
18	30.5
16	28.7
25	36.0
20	34.6

At this point, it is important to emphasize the case of $\Pi_{species}(R)$, which evaluates to:

Species
Western Hemlock
Douglas-Fir

Recall that the output of a relational algebra operation is also a relation. Remember the rule that a relation cannot have any duplicate tuples? Well, in the case of a 1-dimensional relation where we only have one attribute and several tuples, any duplicate values for the tuples must be eliminated, leaving us with only the two unique values "Douglas-Fir" and "Western Hemlock" when we project R over $Species$. You should recognize now that this property of projection can be useful for identifying the unique values of any attribute, which is frequently needed when sorting through a relational database.

5.6 Rename

Rename is an operator that allows us to assign a variable name to a relational algebra expression. This has the benefit of making it simpler to track or reuse previous operations in complex relational database algebra. For example, let $S = \sigma_{height>20}(R)$, then $\Pi_{species}(S)$ evaluates to:

Species
Douglas-Fir

5.7 Set Union

Next we will introduce **binary** operators, that is, they take two relations as input. **Set Union** is one such operator that effectively appends one relation to another. The important rule for union is that both input relations must share the same number and type of attributes or "union compatible". Formally, set union is expressed as $S \cup T$ where S and T are the two input relations. You can think of set union as simply concatenating the tuples of the two relations together. In other words, the tuples of S are appended to the tuples of T to generate a new output relation. For example, suppose that we make two subsets of our relation R of trees:

$$S = \sigma_{height<20}(R)$$

$$T = \sigma_{height\geq 20}(R)$$

Then we can union these two relations back into our original relation R as $S \cup T$, which evaluates to:

ID	Height (m)	DBH (cm)	Species
1	14	26.8	Western Hemlock
2	18	30.5	Western Hemlock
3	16	28.7	Western Hemlock
4	25	36.0	Douglas-Fir
5	20	34.6	Douglas-Fir

Formally, this would all be expressed as $\sigma_{height<20}(R) \cup \sigma_{height\geq 20}(R)$ or $S \cup T$, which in this case is also just equivalent to R. You should see that the result of union is an inclusion of all tuples, so semantically a union can be read as "the tuples in relation S *or* the tuples in relation T".

5.8 Set Intersection

On the other hand, suppose that we want to define a new relation based on restricting the set of tuples that are in two different relations. This is known as **set intersection** and is formally expressed as $S \cap T$. Just like union, intersection also requires that the two relations be union compatible. Suppose we have two relations defined by subsetting height by < 25 m and > 15 m:

$$S = \sigma_{height<25}(R)$$

ID	Height (m)	DBH (cm)	Species
1	14	26.8	Western Hemlock
2	18	30.5	Western Hemlock
3	16	28.7	Western Hemlock
5	20	34.6	Douglas-Fir

$$T = \sigma_{height>15}(R)$$

ID	Height (m)	DBH (cm)	Species
2	18	30.5	Western Hemlock
3	16	28.7	Western Hemlock
4	25	36.0	Douglas-Fir
5	20	34.6	Douglas-Fir

There are 3 tuples that appear in both of these relations, so the intersection $S \cap T$ would evaluate to:

ID	Height (m)	DBH (cm)	Species
2	18	30.5	Western Hemlock
3	16	28.7	Western Hemlock
5	20	34.6	Douglas-Fir

Semantically, set intersection is read as "the tuples in relation S *and* the tuples in relation T".

5.9 Set Difference

Set difference returns the tuples that are unique in one relation relative to another relation, but both relations must be union compatible. Formally, difference is expressed as $S - T$, and just like mathematical subtraction, the order of relations in the set difference is important and non-commutative. For example, $\sigma_{height<25}(R) - \sigma_{height>15}(R)$ evaluates to:

ID	Height (m)	DBH (cm)	Species
1	14	26.8	Western Hemlock

and $\sigma_{height>15}(R) - \sigma_{height<25}(R)$ evaluates to:

ID	Height (m)	DBH (cm)	Species
4	25	36	Douglas-Fir

Semantically, you would read the set difference $S - T$ as "the tuples in relation S minus any of the same tuples in relation T".

5.10 Cartesian Product

So far, we have seen the cases of mathematical addition (set union) and subtraction (set difference), but we can also apply multiplication and division. Multiplication of two relations is simply known as the **Cartesian product**. In the same way that a set of tuples and attributes can be multiplied to create a relation $R = A \times B$, we can also multiply two relations together and they do not need to be union compatible. For example, if $S = \Pi_{height,dbh}(\sigma_{height<20}(R))$ evaluates to:

Height (m)	DBH (cm)
14	26.8
18	30.5
16	28.7

and $T = \Pi_{ID,Species}(\sigma_{dbh>34}(R))$ evaluates to:

ID	Species
4	Douglas-Fir
5	Douglas-Fir

then the Cartesian product of $S \times T$ evaluates to:

ID	Height (m)	DBH (cm)	Species
4	14	26.8	Douglas-Fir
5	18	30.5	Douglas-Fir
4	16	28.7	Douglas-Fir
5	14	26.8	Douglas-Fir
4	18	30.5	Douglas-Fir
5	16	28.7	Douglas-Fir

5.11 Set Divison

Finally, **set division** is an operation of division between two relations, and you can think of it semantically as, "all the values of an attribute in R that are found with the tuples of S." Set division is expressed as $S \div T = U$ and like the Cartesian product and set difference, set division is non-commutative, so the order of S and T changes the value of U.

For the next example of set division, we will introduce a new relation S, which is not a subset of R. Suppose, in addition to R, we have cataloged information about different tree species, some of which are in R (these data are a small sample of a full list of tree species codes commonly used in British Columbia, Canada):

Code	Species
AT	Trembling Aspen
BB	Balsam Fir
CW	Western Red Cedar
E	Birch
FD	Douglas-Fir
HW	Western Hemlock
YC	Yellow Cedar

Suppose we want to answer the question, *What are all the species codes that are present in our plot of trees?* We can answer this question by first projecting *Species* over R to give relation $T = \Pi_{species}(R)$:

Species
Western Hemlock
Douglas-Fir

Then, dividing S by T, $S \div T = U$, can be formally expanded to:

$$\Pi_{code}(S) - \Pi_{code}((\Pi_{code}(S) \times T) - S)$$

We read the first term $\Pi_{code}(S)$ as "the projection of the attributes of S that

are not in T". In our case, there is only one attribute in S not in T, which is $Code$, so $\Pi_{code}(S)$ evaluates to:

Code
AT
BB
CW
E
FD
HW
YC

Then, $\Pi_{code}(S) \times T$ is the Cartesian product of the previous projection and T, which yields a relation of all the combinations of T with the attributes in S that are not in T:

Code	Species
AT	Western Hemlock
BB	Western Hemlock
CW	Western Hemlock
E	Western Hemlock
FD	Western Hemlock
HW	Western Hemlock
YC	Western Hemlock
AT	Douglas-Fir
BB	Douglas-Fir
CW	Douglas-Fir
E	Douglas-Fir
FD	Douglas-Fir
HW	Douglas-Fir
YC	Douglas-Fir

Next, we take the set difference between the Cartesian product above and S, $(\Pi_{code}(S) \times T) - S$, which has the effect of removing the tuples already observed in S. This leaves us with a relation that has all the "incorrect" code-species combinations:

Code	Species
AT	Western Hemlock
BB	Western Hemlock
CW	Western Hemlock
E	Western Hemlock
FD	Western Hemlock
YC	Western Hemlock
AT	Douglas-Fir
BB	Douglas-Fir
CW	Douglas-Fir
E	Douglas-Fir
HW	Douglas-Fir
YC	Douglas-Fir

Next, we project *Code*, which again is the only attribute in S not in T, from the set difference above $\Pi_{code}((\Pi_{code}(S) \times T) - S)$, which yields:

Code
AT
BB
CW
E
FD
YC
HW

And finally, we take the set difference between $\Pi_{code}(S)$ and the projection above to obtain the code for the trees in our plot:

Code
FD
HW

You can think of set division as the inverse of a Cartesian product. However, just like division, the Cartesian product itself is non-commutative because it is a set of *ordered* pairs. If S contains a tuple that is not in T, then the Cartesian product of $S \times T$ has a different order than would be the case if both S and T were identical. As an example, $S \times T$ evaluates to:

Code	Species
FD	Western Hemlock
HW	Western Hemlock
FD	Douglas-Fir
HW	Douglas-Fir

and $T \times S$ evaluates to:

Species	Code
Western Hemlock	FD
Douglas-Fir	FD
Western Hemlock	HW
Douglas-Fir	HW

Therefore, we cannot simply rewrite $S \div T = U$ as $U \times T = S$, but we could express $U \div S = T$, which evaluates to T:

Species
Western Hemlock
Douglas-Fir

We have now considered the eight primary relational algebra operators (selection, projection, rename, set union, set intersection, set difference, Cartesian product, and set division) that can be applied to relations in a RDBMS. In the next section, we will look at another set of logical operators known as Boolean algebra, which give rise to logical languages for interacting with a RDBMS.

5.12 Boolean Algebra

Whenever we create and solve an arithmetic or relational algebra expression, we usually focus on the *value* of the output. In other words, $1 + 1$ evaluates to a value of 2. But we often need to evaluate the *truth* of a statement. For example, $1 + 1 = 2$ is a *true* statement and $1 + 1 = 1$ is a $false$ statement. **Boolean algebra** seeks to express mathematical expressions in terms of *truth values*. Boolean truth values are usually expressed as *true* or $false$, but it is also common in computer programming languages and GIS to see these encoded with values of 1 for *true* and 0 for $false$. Attributes can also take on Boolean values of *true* or $false$ as a data type. Boolean algebra uses equality and conditional operators, which we will consider next.

5.13 Equality Operators

You are probably already familiar with the basic equality operators used in Boolean algebra: - = "exactly equal to" (usually expressed with ==) - > "greater than" - $\geq$ "greater than or equal to" (usually expressed with >=) - < "less than" - $\leq$ "less than or equal to" (usually expressed with <=) - $\neq$ "not equal to" (usually expressed with ! = or <>)

All of the equality operators above evaluate to logical *true* or $false$ values. They are quite elementary, so we will not go into much detail except to show that these equality operators are the basis for forming more complex Boolean expressions.

Basic arithmetic expressions can also be applied to Boolean truth values and it can be helpful to rewrite Boolean values with values of 1 and 0:

- $true + true = 2$
- $true + false = 1$
- $false + false = 0$
- $true - true = 0$
- $true \div false = undefined$
- $false \div true = 0$
- $true \times true = 1 = true$
- $5 \times false = 0 = false$

Multiplication of Boolean values is a special case where the expression will always result in a Boolean value. That is, multiplying any combination of 1 and 0 will always return 1 or 0. In other words, the domain of the input $[0, 1]$ is equivalent to the domain of the output $[0, 1]$, which is a property that is frequently exploited in GIS in order to concatenate more complex expressions. For example, the statement $true \times (1 + 2 = 3) \times (4 > 3)$ can be rewritten as $1 \times 1 \times 1$ and evaluates to *true*, while $true \times (1 + 2 = 3) \times (4 < 3)$ can be rewritten as $1 \times 1 \times 0$ and evaluates to $false$.

Below are some examples of using equality operators and what they evaluate to:

$true = false$ can be rewritten as $1 = 0$, which is $false$. $true > false$ can be rewritten as $1 > 0$, which is *true*. $true \neq false$ is *true*. $1 + 1 = 1$ is $false$. $2 + 3 = 4 + 1$ is *true*.

Next, we will look at how to apply arithmetic and equality expressions to relations and evaluate them in Boolean terms. We have already seen how predicates allow us to do set selection with relational algebra. For example, we know that q evaluates to:

ID	Height (m)	DBH (cm)	Species
2	18	30.5	Western Hemlock
3	16	28.7	Western Hemlock
4	25	36.0	Douglas-Fir
5	20	34.6	Douglas-Fir

If we were to break this down in Boolean terms, the statement $height > 15$ applied to R returns the following Boolean values for each tuple:

ID	Height (m)	DBH (cm)	Species	Boolean
1	14	26.8	Western Hemlock	FALSE
2	18	30.5	Western Hemlock	TRUE
3	16	28.7	Western Hemlock	TRUE
4	25	36.0	Douglas-Fir	TRUE
5	20	34.6	Douglas-Fir	TRUE

For another example, consider that $(height > 15) \times (species = WesternHemlock)$ evaluates to:

ID	Height (m)	DBH (cm)	Species	Boolean
1	14	26.8	Western Hemlock	FALSE
2	18	30.5	Western Hemlock	TRUE
3	16	28.7	Western Hemlock	TRUE
4	25	36.0	Douglas-Fir	FALSE
5	20	34.6	Douglas-Fir	FALSE

We can also evaluate the equivalency between two attributes such as $height = dbh$, which evaluates to:

ID	Height (m)	DBH (cm)	Species	Boolean
1	14	26.8	Western Hemlock	FALSE
2	18	30.5	Western Hemlock	FALSE
3	16	28.7	Western Hemlock	FALSE
4	25	36.0	Douglas-Fir	FALSE
5	20	34.6	Douglas-Fir	FALSE

5.14 Conditional Operators

Now that we have a good understanding of equivalency operators, let us turn to consider conditional operators, which are also known as Boolean operators. Boolean operators are, in some ways, similar to some arithmetic operators except that they are based on natural language. There are three primary Boolean operators: AND, OR, XOR, and NOT. These operators are commonly used for database queries and with search engines, and indeed they form an important basis for query languages that are used to interact with an RDBMS.

Consider the statement $(height > 15) AND (species = WesternHemlock)$. This statement is equivalent to $(height > 15) \times (species = WesternHemlock)$ and evaluates to exactly what what we saw earlier:

ID	Height (m)	DBH (cm)	Species	Boolean
1	14	26.8	Western Hemlock	FALSE
2	18	30.5	Western Hemlock	TRUE
3	16	28.7	Western Hemlock	TRUE
4	25	36.0	Douglas-Fir	FALSE
5	20	34.6	Douglas-Fir	FALSE

Figure **??** illustrates what is going on here, we are only returning the tuples that evaluate *true* for both statements. Hence, Boolean AND is equivalent to multiplying two Boolean truth values together. You should also recognize that a Boolean AND is equivalent to what a set intersection $A \cap B$ achieves between two relations.

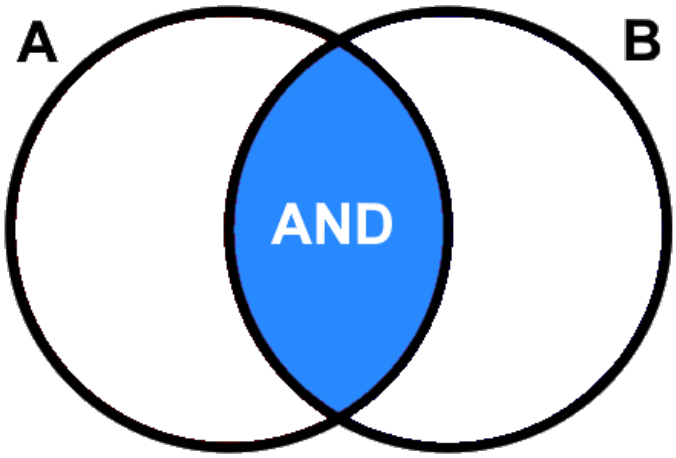


Figure 5.2: Boolean A AND B returns the area shaded blue. Pickell, CC-BY-SA-4.0.

If we do not want to be so restrictive, we could use Boolean OR such as $(height > 15) OR (species = WesternHemlock)$, which evaluates to:

ID	Height (m)	DBH (cm)	Species	Boolean
1	14	26.8	Western Hemlock	TRUE
2	18	30.5	Western Hemlock	TRUE
3	16	28.7	Western Hemlock	TRUE
4	25	36.0	Douglas-Fir	TRUE
5	20	34.6	Douglas-Fir	TRUE

Figure **??** illustrates the case of the Boolean OR. As you can see, it returns everything where either of the statements evaluate to $true$, regardless if the other statement is $false$. You should also recognize that a Boolean OR is equivalent to what a set union $A \cup B$ achieves between two relations.

Suppose we want to identify all the trees that are greater than 15 m, but not Western Hemlock. In this case, we would use the expression $(height > 15) NOT (species = WesternHemlock)$, which evaluates to:

ID	Height (m)	DBH (cm)	Species	Boolean
1	14	26.8	Western Hemlock	FALSE
2	18	30.5	Western Hemlock	FALSE
3	16	28.7	Western Hemlock	FALSE
4	25	36.0	Douglas-Fir	TRUE
5	20	34.6	Douglas-Fir	TRUE

Figure **??** illustrates how Boolean NOT essentially negates or inverts the state-

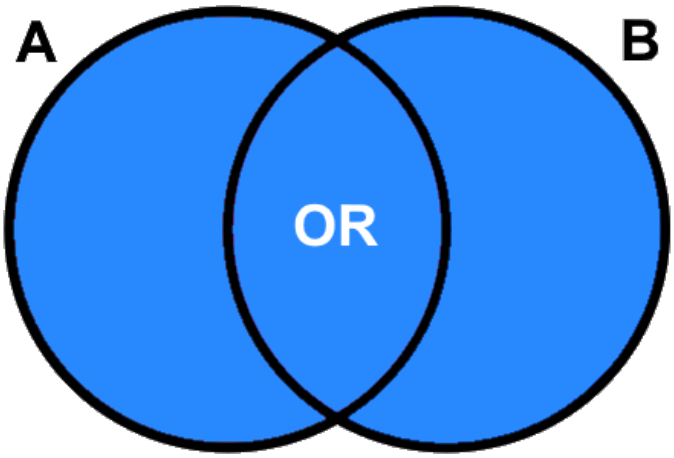


Figure 5.3: Boolean A OR B returns the area shaded blue. Pickell, CC-BY-SA-4.0.

ment that follows. In this case, instead of returning the Western Hemlock tuples, $NOT(species = WesternHemlock)$ returns "everything except" Western Hemlock, which is also equivalent to $species \neq WesternHemlock$.

Finally, the case of Boolean XOR returns any tuples that are not *true* for both statements, but are *true* individually. This is known as the **symmetric difference** and "e**X**clusive **OR**" because we are only returning the tuples that are exclusive based on both statements. For example, $(height > 15)XOR(species = WesternHemlock)$ evaluates to:

ID	Height (m)	DBH (cm)	Species	Boolean
1	14	26.8	Western Hemlock	TRUE
2	18	30.5	Western Hemlock	FALSE
3	16	28.7	Western Hemlock	FALSE
4	25	36.0	Douglas-Fir	TRUE
5	20	34.6	Douglas-Fir	TRUE

Figure **??** illustrates how Boolean XOR excludes all the tuples that evaluate to *true* for both statements. In the example above, tuples ID=2 and ID=3 are excluded because both of the statements for height and species are *true*.

5.15 Joining Relations

More often than not, information is stored in separate relations, even if that information is about the same features like lakes, forests, or cities. Remember that a relation cannot have any duplicate tuples. This rule encourages the

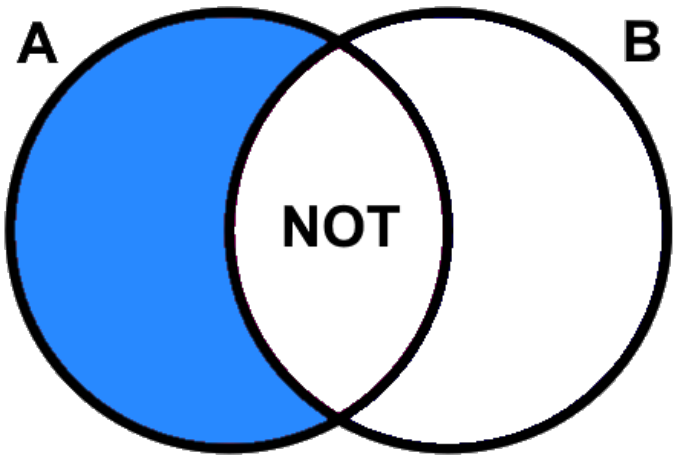


Figure 5.4: Boolean A NOT B returns the area shaded blue. Pickell, CC-BY-SA-4.0.

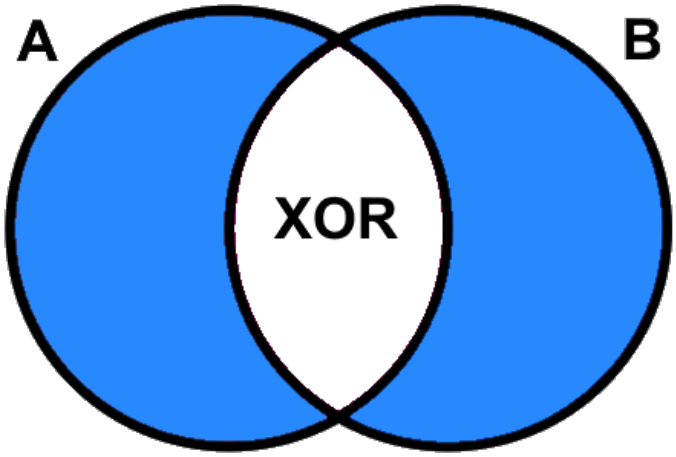


Figure 5.5: Boolean A XOR B returns the area shaded blue. Pickell, CC-BY-SA-4.0.

efficient storage and retrieval of information because information can be dynamically related as needed. For example, consider the overwhelming amount of information that is collected during a census. During the last census in 2016, there were over 14 million households in Canada. Can you imagine wielding a relation with 14 million tuples? These households can be segmented geographically by province, metropolitan areas, municipalities, and census subdivisions as well as by socioeconomic themes such as Indigenous peoples, age, sex, education, income, labour, housing, language, and others. Thus, those 14 million households can be divided up into many smaller relations, which can be accessed and summarized geographically and thematically. Since these relations represent different geographies or themes on the same set (i.e., households), we need to be more specific about how exactly two relations get combined if, for example, we want to combine themes with geographies. For this reason, we have joins.

5.16 Keys

Like the Cartesian product, joins are always binary operations, requiring two relations as input. While the Cartesian product combines relations by ordering all pairs of the elements from the two relations, we need a different method for correctly linking the tuples in relation R that correspond to the tuples of S. To do this, we rely on a common attribute called a **key**, which acts as an address between two relations. A **primary key** serves the purpose to identify the unique tuples in a relation and so it can be used to link other attribute information to those tuples. In a GIS, anytime that you create, copy or modify features such as points, lines or polygons, the newly created data layer (within the relational database) will be indexed with a primary key that counts from 1 to the number of features (tuples) n or from 0 to $n-1$. For example, ID in our relation R serves as the primary key. There are other attributes in R that also uniquely identify all the tuples, but why do you think $Height$ or DBH would be a poor operational choice as a primary key for a large field campaign?

While the primary key identifies the unique tuples in relation R, another key called the **foreign key**, serves to locate the same tuples in another relation S. In other words, a join is defined by a common attribute that is shared between two relations, the primary key in R and the foreign key in S. For example, $Species$ is a foreign key in R and a primary key in S. The case of joining two relations using a set of attributes instead of a single attribute requires a **composite key**. For example, suppose we have a spatial dataset of all the municipalities across Canada. Some of these municipalities will share the same name, though they are in different provinces. Richmond is a city in British Columbia, Ontario, and Quebec. If we need to join census data to these spatial features, we would need to use a composite key comprised of $CityName$ and $ProvinceName$.

5.17 Natural Join

A **natural join** restrictively joins two relations based on a set of common attributes. In this way, natural join is similar to a set intersection in that we are only combining tuples that share an attribute value and any tuples that do not share an attribute value in the other relation are dropped from the output. However, a natural join does not require that two relations be union compatible like a set intersection. Instead, the only requirement is that at least one attribute is shared between the two relations and has the same domain. Formally, natural join is expressed as $R \bowtie S$. and is sometimes referred to as an inner join. As an example, consider our example relations R and S:

R

ID	Height (m)	DBH (cm)	Species
1	14	26.8	Western Hemlock
2	18	30.5	Western Hemlock
3	16	28.7	Western Hemlock
4	25	36.0	Douglas-Fir
5	20	34.6	Douglas-Fir

S

Code	Species
AT	Trembling Aspen
BB	Balsam Fir
CW	Western Red Cedar
E	Birch
FD	Douglas-Fir
HW	Western Hemlock
YC	Yellow Cedar

The natural join $R \bowtie S$ evaluates to:

```
## Joining, by = "Species"
```

ID	Height (m)	DBH (cm)	Species	Code
1	14	26.8	Western Hemlock	HW
2	18	30.5	Western Hemlock	HW
3	16	28.7	Western Hemlock	HW
4	25	36.0	Douglas-Fir	FD
5	20	34.6	Douglas-Fir	FD

5.18 Outer Join

An **outer join** joins all the tuples of two relations based on a common attribute. The result is similar to a set union, except the input relations do not need to be union compatible. Formally, an outer join or sometimes called a full join is expressed as $R \ \ S$, which evaluates to:

```
## Joining, by = "Species"
```

ID	Height (m)	DBH (cm)	Species	Code
1	14	26.8	Western Hemlock	HW
2	18	30.5	Western Hemlock	HW
3	16	28.7	Western Hemlock	HW
4	25	36.0	Douglas-Fir	FD
5	20	34.6	Douglas-Fir	FD
NA	NA	NA	Trembling Aspen	AT
NA	NA	NA	Balsam Fir	BB
NA	NA	NA	Western Red Cedar	CW
NA	NA	NA	Birch	E
NA	NA	NA	Yellow Cedar	YC

5.19 Right and Left Outer Join

Sometimes, it may be desirable to join attributes or tuples from one relation, but not the other. For these cases, we can use either **right outer join** or **left outer join**. Formally, right outer join is expressed as $R \rtimes S$ and evaluates to:

```
## Joining, by = "Species"
```

ID	Height (m)	DBH (cm)	Species	Code
1	14	26.8	Western Hemlock	HW
2	18	30.5	Western Hemlock	HW
3	16	28.7	Western Hemlock	HW
4	25	36.0	Douglas-Fir	FD
5	20	34.6	Douglas-Fir	FD
NA	NA	NA	Trembling Aspen	AT
NA	NA	NA	Balsam Fir	BB
NA	NA	NA	Western Red Cedar	CW
NA	NA	NA	Birch	E
NA	NA	NA	Yellow Cedar	YC

Formally, left outer join is expressed as $R \ltimes S$ and evaluates to:

```
## Joining, by = "Species"
```

ID	Height (m)	DBH (cm)	Species	Code
1	14	26.8	Western Hemlock	HW
2	18	30.5	Western Hemlock	HW
3	16	28.7	Western Hemlock	HW
4	25	36.0	Douglas-Fir	FD
5	20	34.6	Douglas-Fir	FD

5.20 Theta Join

We can also join relations conditionally and without sharing a common attribute, which is known as a **theta join** and expressed as $R \bowtie_\theta S$. To understand how a theta join works, it is useful to recognize that $R \bowtie_\theta S = \sigma_\theta(R \times S)$. As you can see, a theta join is simply a selection of a Cartesian product where theta θ is the predicate. For example, $R \bowtie_{height>19} S$ evaluates to:

```
## Joining, by = "Species"
```

ID	Height (m)	DBH (cm)	Species	Code
4	25	36.0	Douglas-Fir	FD
5	20	34.6	Douglas-Fir	FD

5.21 Cardinality of Joins

Depending on the schema of the two relations being joined, the number of tuples joined from one relation to another can vary and is known as **cardinality**. In the simplest case, one tuple in R is joined to one tuple in S, and this cardinality is known as **one-to-one** usually expressed as 1:1. The natural join example above, $R \bowtie S$, is an example of **one-to-many** (1:M) or **many-to-one** (M:1) cardinality because one species tuple found in S corresponds to many species tuples in R. Finally, **many-to-many** (M:M) cardinality describes the case where there are multiple tuples in R that correspond to multiple tuples in S. An example of a many-to-many relationship might be many species of trees in R that correspond to many forest stands in S. In other words, a forest stand might be comprised of many species and any particular species might be found in many forest stands. Figure **??** illustrates how cardinality might emerge depending on the relational schema and problem at hand.

5.22 Structured Query Language

Throughout this chapter, we have seen the various ways that relations are manipulated through relational algebra, Boolean logic, and joins. Since a GIS relies on a RDBMS to interact with data, especially data in the attribute table, geomatics professionals literally need a language to programmatically execute relational algebra, joins, and the other functions of a RDBMS within the GIS software.

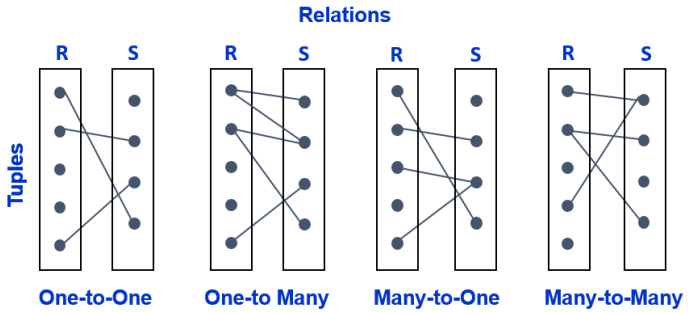


Figure 5.6: Cardinality of joins between relations R and S. Pickell, CC-BY-SA-4.0.

Such languages are known as query languages, each with its own syntax and use. By far, the most commonly used query language for RDBMS in GIS and across other systems is **Structured Query Language** abbreviated **SQL** and pronounced "sequel". SQL has five primary language elements: 1. Clauses state an action or operation; 2. Expressions evaluate to a value; 3. Predicates evaluate an expression using equivalency and Boolean operators; 4. Queries apply set selection on a predicate; and 5. Statements are the combination of all the elements above

SQL has numerous keywords, which are the actions that comprise a clause. It is beyond the scope of this textbook to describe all of them, but most of the SQL keywords are implemented within GIS software in other ways. For example, you would rarely need to programmatically `ADD` an attribute to a relation. Instead, you might click an "Add field" button within the GIS software you are using. Similarly, you might never programmatically `UPDATE` the value for a tuple because most GIS software will allow you to simply double-click a cell in the table and change the value. The primary action that is nearly always performed programmatically with proper SQL syntax is applying a query, which is what we will focus on for this section.

SQL queries are fundamental for implementing set selection $\sigma_{predicate}$ and they look like this:

```
SELECT attributes
FROM relation
WHERE predicate;
```

The entire form above is a statement, which is enclosed by a semi-colon at the end. The statement is comprised of three clauses using the keywords: `SELECT`,`FROM`, and `WHERE`. The `SELECT attributes` clause defines which attributes of the relation will be returned. You should recognize that this is the equivalent of applying a set projection $\Pi_{attributes}$ to the entire set selection statement. You can specify attributes by name (e.g., `SELECT Species`), but it

is more common to return all of the attributes of the relation with an asterisk like `SELECT *`. The `FROM relation` clause defines which relation the selection is performed on. Keeping in mind that a RDBMS is comprised of many relations and at any given time you may have several different data sources open in your GIS software, the `FROM` keyword helps to clarify exactly which relation contains the attributes defined by the `SELECT` clause. Finally, the `WHERE predicate` clause defines the predicate that will be evaluated for the set selection, and this is where the magic happens. Although this is the formal syntax for a SQL query, most GIS software will usually only require the user to define the predicate, so next we will look at how to construct different SQL queries on our relation R.

Suppose we want to select the trees that are greater than 15 m, like in our previous equivalency example of $\sigma_{height>15}(R)$. The SQL statement looks like this `SELECT * FROM R WHERE height > 15;`. If we only want to return the species for tree heights greater than 15 m, then the SQL statement looks like this `SELECT Species FROM R WHERE height > 15;` and evaluates to:

Species
Western Hemlock
Western Hemlock
Douglas-Fir
Douglas-Fir

The above SQL statement would be an example of $\Pi_{species}(\sigma_{height>15}(R))$. In SQL, the multiplication symbol has the arithmetic meaning and cannot be used to concatenate two predicates. For this reason, we have the Boolean operators for evaluating multiple predicates. For example, $(height > 15) \times (species = WesternHemlock)$ would be written in SQL as `SELECT * FROM R WHERE height > 15 AND species="Western Hemlock";`. Our previous example of using Boolean NOT in SQL would be written as `SELECT * FROM R WHERE NOT species="Western Hemlock";`. These are all relatively simple examples, but it common to create more complicated queries that use several Boolean operators. Note here how the species value in the expression above is in quotation marks `"Western Hemlock"` because the data type of the species attribute is a *string*. By contrast, the height value in the previous expression is simply an *integer number*. It is important to emphasize at this point that the only equivalency operator that can be used with string data type attributes is =. In other words, `"Western Hemlock">"Douglas-Fir"` is illogical, cannot be evaluated, and will return an error.

If you combine two or more Boolean operators into one statement, then they are evaluated in SQL according to the following precedence: 1. Anything enclosed within parentheses `()` 2. `NOT` 3. `AND` 3. `OR` For example, `SELECT * FROM R WHERE dbh < 30 AND species="Douglas-Fir" OR species="Western Hemlock";` would evaluate to:

ID	Height (m)	DBH (cm)	Species
1	14	26.8	Western Hemlock
2	18	30.5	Western Hemlock
3	16	28.7	Western Hemlock

But if we want the `OR` to be evaluated before the `AND`, then we need to use parentheses like `SELECT * FROM R WHERE dbh < 30 AND (species="Douglas-Fir" OR species="Western Hemlock");`, which evaluates to:

ID	Height (m)	DBH (cm)	Species
1	14	26.8	Western Hemlock
3	16	28.7	Western Hemlock

You may notice the case of XOR conspicuously missing from the order above and this is because SQL does not natively implement the XOR operator. If you want to evaluate the exclusive OR example used in the previous section, $(height > 15) XOR (species = WesternHemlock)$, then you would construct a SQL statement like this `SELECT * FROM R WHERE (height > 15 OR species="Western Hemlock") AND NOT (species="Western Hemlock" AND height > 15);`. As you can see, SQL queries can quickly get complex and involve many Boolean operators, so it is important to understand operator precedence and whenever in doubt, you can always use parentheses to override any precedence rules. You should also recognize that there are many ways to write complex statements to achieve your desired selection and you should always prefer the simplest statement possible.

Finally, a very common query that involves returning all tuples that match an attribute value in a list of values can be applied using the `IN` operator in SQL. For example, suppose we want to select all conifer tree species (codes: BB, CW, FD, HW, and YC) from S below:

Code	Species
AT	Trembling Aspen
BB	Balsam Fir
CW	Western Red Cedar
E	Birch
FD	Douglas-Fir
HW	Western Hemlock
YC	Yellow Cedar

Your natural reaction to this problem might be to write a long SQL statement like `SELECT * FROM S WHERE code="BB" OR code="CW" OR code="FD" OR code="HW" OR code="YC";`. This is perfectly fine, but you can write this more economically with `IN` such as `SELECT * FROM S WHERE code IN("BB","CW","FD","HW","YC");`. Be aware that a common mistake is to write a long predicate using `OR` like `code="BB" OR "CW" OR "FD" OR "HW"`

`OR "YC"`, but this is incorrect syntax in SQL. Remember that each side of an `OR` or `AND` operator is an *expression* that evaluates to a Boolean truth value. So `code="BB" OR "CW"` will return an error because `"CW"` alone cannot be evaluated to a Boolean truth value.

5.23 Case Study: Combining Socioeconomic and Vegetation Information for Assessing Population Vulnerability

Case Study Author: Taelynn Lam (CC BY 4.0. unless otherwise indicated), University of British Columbia, Master of Geomatics for Environmental Management graduate, 2021

Vegetation diversity in urban landscapes is important to support urban forest biodiversity and residents' mental health. The aim of this case study is to link together socioeconomic data and vegetation information to identify areas to prioritize intervention in the City of Vancouver. The Canadian Index of Multiple Deprivation (CIMD) data has four dimensions of population vulnerability scores and we will aggregate these scores to obtain an overall vulnerability score for each dissemination area (DA) in Vancouver. We will compute the vegetation diversity score using street trees data and vegetation type cover richness data[1] and then use query to identify priority areas.

5.24 Join

The raw CIMD tabular data includes the DA code and the corresponding vulnerability scores (table 1). In order to visualize these scores on a map, we will need to relate these scores to spatial data that include the information about the DA polygons and the coordinate pairs associated with each DA. Hence, a polygon shapefile of the DAs in Vancouver[2] is obtained, and its attributes are shown in Table 2.

The tabular data of the CIMD scores are related to the DAs polygon by the DA code. The cardinality of the relationship between these two tables is one-to-one as each DA is described by one set of the CIMD scores. To join the CIMD scores to the Vancouver DA polygons, we would use the `PRCDDA` attribute in the CIMD table as the foreign key to perform a join on the `DAUID` attribute in the Vancouver DA polygon relation. Now that the CIMD scores are joined to the Vancouver DA polygon attribute table, we can create choropleth maps to display the vulnerability scores of the DAs (Figure **??**).

[1] Obtained from reclassifying Land Cover Classification 2014 - 2m Raster to one vegetation class and five vegetation classes and counted the number of vegetation type cover classes using the `Zonal Histogram` Tool.

[2] Extracted by clipping the Canada-wide dissemination areas boundary to the City of Vancouver's municipality boundary.

Table 5.1: An excerpt of the CIMD data table.

PRCDDA	Province	DA population	Ethno-cultural composition quintiles	Ethno-cultur
59010123	British Columbia	434	1	
59010124	British Columbia	559	1	
59010125	British Columbia	522	2	
59010126	British Columbia	671	2	
59010127	British Columbia	319	1	
59010128	British Columbia	545	3	

Table 5.2: An excerpt of the Vancouver DA polygon shapefile attributes.

	DAUID
0	59150727
1	59150728
2	59150729
3	59150730
4	59150731
5	59150732

5.25 Calculation

Suppose we would like to calculate the overall vulnerability score for each DA. We would first name a new field (e.g., "aggregate_score"), set the data type to double (to allow negative values and values with decimal places), and then enter the mathematical expression to specify the calculation to sum the four dimensions of CIMD scores and divide it by four to obtain an averaged vulnerability score for each DA. Using similar steps, we could apply a min-max normalization to transform this overall vulnerability score to a range between 0 and 1 to allow for a quick interpretation of the score. The formula is as follows: $\frac{(X-X_{min})}{(X_{max}-X_{min})}$.

Using what you have learned, join the street tree data and the vegetation type cover richness data to the Vancouver DA attribute table and to compute a vegetation diversity score. The street trees data shows the number of unique street tree species at a DA. Make sure you apply a min-max normalization to obtain the street tree diversity score. The vegetation diversity score can be computed by averaging the normalized scores of the two vegetation data.

Figure **??** shows a map of the vegetation diversity score at the DA level in Vancouver. The vegetation diversity score and the normalized aggregated vulnerability score are linked to each DA and can be viewed as you hover o-ver the DA.

Table 5.3: An excerpt of the attribute table after the joins and calculations.

	DAUID	Aggregated scores	Normalized aggregated scores	Species count	Street tree diversity	Vegeta
0	59150727	0.043	0.125	18	0.212	
1	59150728	0.039	0.125	20	0.238	
2	59150729	0.116	0.137	20	0.238	
3	59150730	-0.103	0.101	24	0.288	
4	59150731	-0.336	0.063	29	0.350	
5	59150732	-0.095	0.103	33	0.400	

Table 5.4: Query result shows five records matched the priority area requirements.

	DAUID	Normalized aggregated scores	Street tree diversity	Vegetation richness	Vegetation diversity
217	59150755	0.505	0.150	0.8	
222	59150760	0.590	0.075	0.6	
223	59150761	0.510	0.112	0.6	
527	59153187	0.522	0.125	0.8	
528	59153188	1.000	0.138	0.8	

5.26 Query

Areas with a higher proportion of vulnerable populations and less variety of vegetation to support resident's mental wellbeing are more in need for intervention. Supposed we define the priority area as DAs with a normalized aggregated vulnerability score greater than or equal to 0.5 and a vegetation diversity score less than 0.5. We could use the `Select By Attributes` tool to identify these priority areas by entering the appropriate query expression.

Remember This?

Models are abstractions of reality and help us understand and communicate complex ideas.

5.27 Summary

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut in dolor nibh. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent et augue scelerisque, consectetur lorem eu, auctor lacus. Fusce metus leo, aliquet at velit eu, aliquam vehicula lacus. Donec libero mauris, pharetra sed tristique eu, gravida ac ex. Phasellus quis lectus lacus. Vivamus gravida eu nibh ac malesuada. Integer in

libero pellentesque, tincidunt urna sed, feugiat risus. Sed at viverra magna. Sed sed neque sed purus malesuada auctor quis quis massa.

Reflection Questions

1. Explain ipsum lorem.
2. Define ipsum lorem.
3. What is the role of ispum lorem?
4. How does ipsum lorem work?

Practice Questions

2. Given ipsum, solve for lorem.
3. Draw ipsum lorem.

Recommended Readings

Ensure all inline citations are properly referenced here.

Chapter 6

Overlay and Proximity Analysis

Introduction here.

Learning Objectives

1.

Key Terms

Overlay, Union, Intersect, Identity, Difference, Symmetrical Difference, Buffer, Near Distance, Thiessan Polygons

6.1 Cartographic Modelling

6.2 Geoprocessing

6.3 Capability Modelling

6.4 Suitability modelling

6.5 Overlay Methods

6.6 Attribute Transfer

6.7 Boolean Algebra

6.8 Spatial Join

6.9 Clip

6.10 Intersect

6.11 Line Intersection

6.12 Union

6.13 Identity

6.14 Erase

6.15 Split

6.16 Symmetrical Difference

6.17 Update

6.18 Proximity Methods

6.19 Buffer

Attribute-dependent buffer

6.20 Near Distance

6.21 Thiessen Polygons

6.22 Summary

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut in dolor nibh. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent et augue scelerisque, consectetur lorem eu, auctor lacus. Fusce metus leo, aliquet at velit eu, aliquam vehicula lacus. Donec libero mauris, pharetra sed tristique eu, gravida ac ex. Phasellus quis lectus lacus. Vivamus gravida eu nibh ac malesuada. Integer in libero pellentesque, tincidunt urna sed, feugiat risus. Sed at viverra magna. Sed sed neque sed purus malesuada auctor quis quis massa.

Reflection Questions

1. Explain ipsum lorem.
2. Define ipsum lorem.
3. What is the role of ispum lorem?
4. How does ipsum lorem work?

Practice Questions

2. Given ipsum, solve for lorem.
3. Draw ipsum lorem.

Recommended Readings

Ensure all inline citations are properly referenced here.

Chapter 7

Topology and Geocoding

Frequently, we need spatial data to behave and relate and in specific and predictable ways. Many types of analyses may expect spatial data to be represented and interact in a standard form. In this chapter, we will extend our knowledge of data models using topology, which unlocks many advanced spatial analyses. We will look at a specific example of an analysis that requires topology, geocoding, which will be a convenient segue into network analysis discussed in the following chapter.

Learning Objectives

1. Understand the role of topology in governing data behaviour and data organization
2. Recognize some examples and uses of 2D and 3D topologies
3. Understand the role of bounding a set of points from triangulation and convex hulls
4. Synthesize the process of geocoding
5. Practice geocoding addresses and reverse geocoding addresses to other coordinate systems

Key Terms

Vertex, Node, Pseudonode, Dangle, Planar Topology, Non-Planar Topology, Geocoding, Adjacency, Overlap, Connect, Inside, Reverse Geocoding, Singlepart, Multipart, Holes, Delaunay Triangulation, Thiessen Polygons, Voronoi Diagram, Centroid, Convex Hull, Convex Alpha Hull, Multipatch

7.1 Topology

Topology describes the relationships of spatial data. This is a very broad definition that encompasses the wide range of possible arrangements of spatial data in practice. If we drill down into this concept, topology is really what allows us undertake specific types of analysis that requires or expects spatial data to behave in a certain way. If you think about the feature geometries that we have at our disposal, then there are no fewer than nine combinations of how these geometries can interact as illustrated in Figure **??** below.

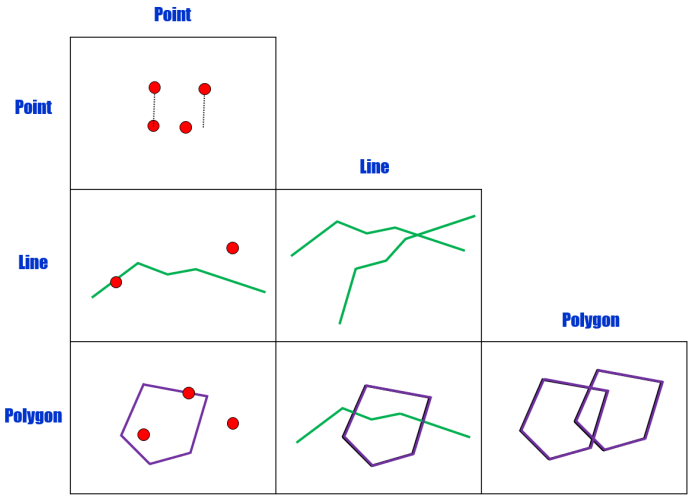


Figure 7.1: Grid showing all the combinations of how point, line and polygon geometries can interact. Pickell, CC-BY-SA-4.0.

It is important to recognize that there may be cases where we "expect" that a given combination of features will conform to a specific interaction. For example, the provinces and territories of Canada are typically represented as polygons that share adjacent boundaries. That is, the adjacent boundaries shared between any two provinces or territories cannot logically overlap as this representation (model) would contravene the legal definitions of the provinces and territories. In another case, a human technician may erroneously digitize a road that crosses another road without indicating that the two roads share an intersection, which could have consequences for how traffic flow can be modeled between the two roads (i.e., intersection with traffic light versus overpass). These are both examples of situations where topology is needed. Topology applies logic to define how features are expected to relate to other features in order to conform to knowledge systems like legal definitions of land and traffic flow. In short, topology ensures data integrity for other types of analysis.

7.2 Planar vs. Non-Planar Topology

In the context of topology, **planar** refers to the concept that all vertices of feature vector geometry are mapped onto the same plane. So in a planar worldview, all lines and polygons share coincident vertices. For example, if two polygons overlap, then the overlapping area forms a new polygon with a boundary of vertices defined by the union of the two other polygons. Also, if two lines overlap, then the two lines are divided into four new segments and a new vertex is formed at the intersection. In other words, planar topology does not allow polygons or lines to lay "underneath" or "on top" of another line or polygon and feature geometries must always be distinct.

On the other hand, **non-planar** topology is the concept that vertices of feature vector geometry can be mapped to different planes. It is important to emphasize here that when we are talking about planes that we are not referring to projected coordinate systems. It is generally assumed that any two spatial data layers containing feature geometries are interacting within the same projected coordinate system. Non-planar topology allows for other knowledge systems to be represented in spatial data. The case where a pipeline runs underneath a river or a territory that was traditionally used by several Indigenous peoples (Figure **??**) are examples of valid non-planar topology.

Non-planar topology of 36 indigenous territories overlapping Vancouver Island, British Columbia. Data from contributors at native-land.ca, CC0.

7.3 Implementing Planar Topology

Implementing planar topology involves defining specific rules for how features should relate to one another given some analytical context. This process also requires that the spatial data are housed a relational database or data model that supports topology. In other words, topology is enforced only by data models that support topological rules. When a topological rule is violated, the relational database identifies the contravening features and displays them on the map and in the attribute table. Then, it is up to an analyst to decide how the error should be corrected. For example, some errors like intersecting lines can automatically be split at the intersection while overlapping polygons might need to be manually edited to reflect the correct adjacency. Thus, the process of applying topology is first to work within a data model that supports topology, then choose the topological rules that reinforce a particular knowledge system, and finally to inspect and decide how to deal with any contraventions. Since planar topology is only supported by certain data models, and some data models are proprietary to certain software, the exact topological rules that can be implemented in a GIS are mostly dependent on the software that you are using. Instead of examining a specific GIS software package, we will discuss the "fundamental" planar topological relationships that are common across nearly all implementations of topology. (If you want to know more about how topology

is implemented within specific data models, skip ahead to the "Data models supporting planar topology" section.)

So far, we have seen that there are six ways to combine feature geometries (points, lines, and polygons). We can extend this understanding to include at least six different ways that they can relate to one another: adjacent; overlap; intersect; connect; cover; and inside. Some of these relationships can be modeled *between* two different spatial layers (e.g., two point layers) or *within* a single spatial layer. In the following sections, we will look at different planar topological rules that apply both between and within feature geometries.

7.4 Adjacency and Overlap

There are times when we need to ensure that two polygons are **adjacent** to one another by sharing a common edge. If two polygons are not adjacent to one another, then a gap, known as a **sliver**, exists between them or they must **overlap**. Consider the case where we are mapping land covers. If we have a formal scheme that describes all possible land covers, then we expect that a map of land covers will have perfect adjacency between all polygons so that there are no areas that are not mapped (i.e., slivers) and that no area has multiple, overlapping land covers. Since lines are also 2-dimensional, lines can overlap other lines. Depending on the context, a topological rule may be needed to promote or prevent this relationship. For example, if you are modeling bus routes, then one road might support several different routes.

[figure of sliver] [figure of overlap]

Some examples of adjacency and overlap topological rules: - Polygons within the same layer must not have gaps - Polygons within the same layer must not overlap - Polygons must not overlap other polygons - Lines must not overlap other lines - Lines must not self-overlap

7.5 Intersect and Connect

As we have seen from Chapter 3, lines are often used to represent phenomena that flow, so intersection and connection are important concepts for these representations. Important to understanding how connection and intersection work in planar topology, we need to understand that lines are comprised of a set of vertices and nodes. A **node** is simply the terminating vertex in a set of vertices for a line. For example, suppose the line segment A has a set of vertices, $[[1,0],[1,3],[1,5]]$. Then the nodes for A are $[1,0]$ and $[1,5]$ (Figure **??**). Since nodes define the end points of a line segment, they are key to enforcing connection rules. We will look at network analysis in more detail in the next chapter. For now, let us consider two different networks that can help us conceptualize some fundamental line topology using nodes.

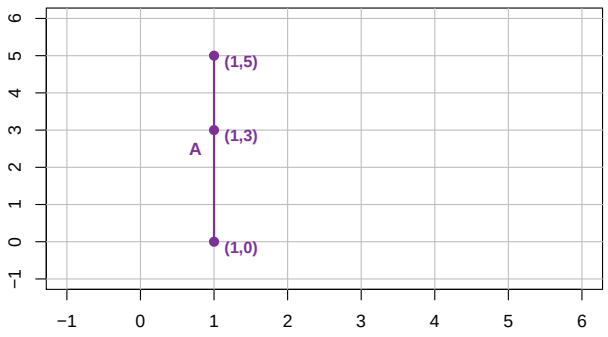


Figure 7.2: Lines are always comprised of two nodes. Line A shown here has nodes at [1,0] and [1,5]. Pickell, CC-BY-SA-4.0.

A network of streams and rivers is based on the hydrological knowledge system that explains how water moves over a terrain surface. In both theory and practice, we know that water flows from higher elevations to lower elevations with limited exceptions. Thus, we expect that streams will connect with other streams and continue to flow towards some outlet such as an ocean. **Connection** refers to the fact that the endpoint node of one stream will fall somewhere on another stream segment. Where two line segments come together, it is possible for one segment A to "undershoot" the other segment B, resulting in the end node of segment A appropriately named a **dangle** Figure **??**) and a loss of connection.

Dangles are the opposite case to **intersections**, which occur when two line segments cross each other. With planar topology, intersections must be modeled with a shared node representing the intersection location. For example, suppose line segment B has a set of vertices, $[[0, 1], [2, 1], [4, 1]]$. If line segments A (defined above) and B are mapped together with non-planar topology, then they will intersect at $[1, 2]$, which is not a vertex represented in either segment (Figure **??**).

Thus, the intersection of A and B with planar topology would yield four new segments: $C = [[0, 2], [1, 2]]$, $D = [[1, 2], [1, 3], [1, 5]]$, $E = [[1, 2], [2, 2], [4, 2]]$, and $F = [[1, 0], [1, 2]]$. Figure **??** illustrates how all four of these new segments share the same node $[1, 2]$ at the intersection of A and B.

As well, **pseudonodes** can occur when a node does not actually terminate a line segment at a junction, for example, between two streams or roads. In other words, a pseudonode is a node that is shared by two lines. Figure **??** illustrates

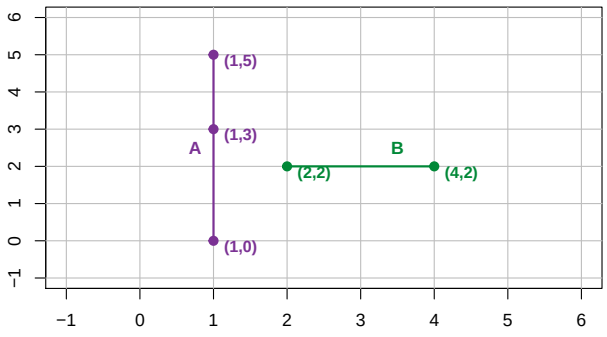


Figure 7.3: A dangle forms when a line (B) does not connect to another line (A). Pickell, CC-BY-SA-4.0.

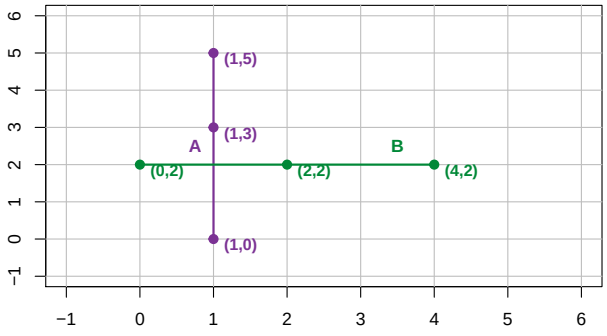


Figure 7.4: Line A mapped with Line B in non-planar topology. Pickell, CC-BY-SA-4.0.

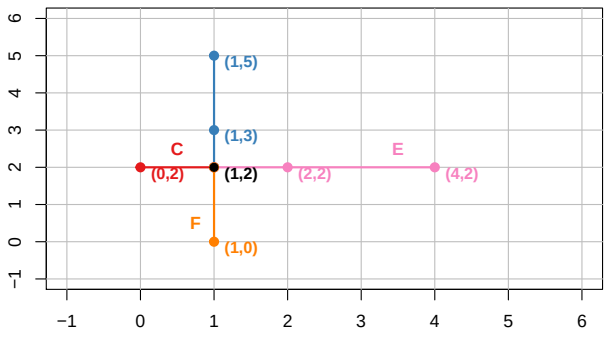


Figure 7.5: Line A mapped with Line B in planar topology yields segments C, D, E, and F. All segments share (1,2) as a node. Pickell, CC-BY-SA-4.0.

a pseudonode occurring at $[3, 5]$.

Some examples of intersection and connection topological rules: - Lines must not intersect other lines - Lines must intersect other lines - Lines must not self-intersect - Lines within a same layer must not self-intersect - Lines must not have dangles

7.6 Coincident and Disjoint

Point features can be either **coincident** or **disjoint** with other point features. Point features that need to be disjoint may be representing trees, mountain peaks, or any similar type of feature that would be expected to be discrete in geographic space. There are also instances where we might need one set of point features to be coincident with another such as field plots that are centered using a tree or other spatially-discrete feature on the landscape.

Some examples of coincident and disjoint topological rules: - Points must be disjoint with other points - Points must by coincident with other points

7.7 Cover

Cover refers to planar topology where a feature lays on or within another feature. For example, dams represented as point features must be covered by a line representing a river (Figure **??**). Similarly, lines representing rivers must be covered by polygons representing watersheds. As well, property parcel polygons must be covered by the municipal or regional tax authority polygon.

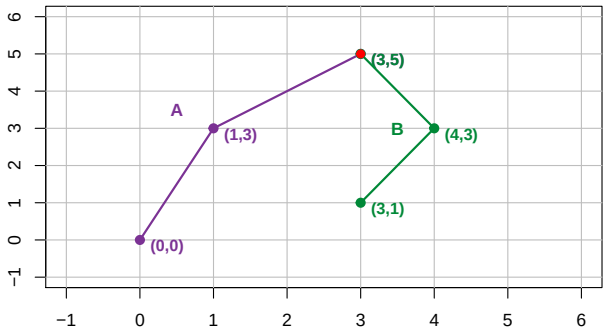


Figure 7.6: Lines A and B share a pseudonode at [3,5], indicated in red. Pickell, CC-BY-SA-4.0.

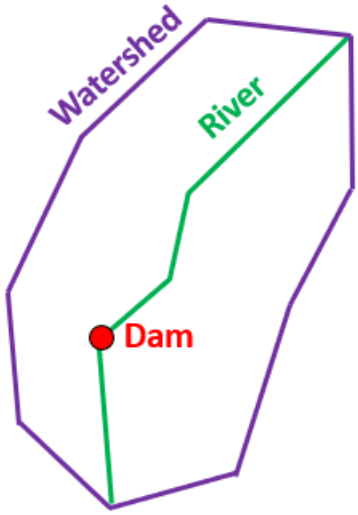


Figure 7.7: Topological relationship between dam (point) covered by a river (line), which is covered by a watershed (polygon). Pickell, CC-BY-SA-4.0.

Some examples of cover topological rules: - Point must be covered by a line - Point must be covered by a polygon - Line must be covered by a polygon - Polygon must be covered by a polygon

7.7.1 Multipart geometry

Sometimes we need to represent several points, lines or polygons as a collection, which is known as **multipart geometry**. Multipart geometry allow us to represent several disjoint and non-adjacent geometries as a single feature. In this way, we can assign attribute values to the collection of features rather than each geometry individually. The territorial boundary of Canada is a good example of an instance where a multipart geometry can be useful because all of the contiguous land and non-contiguous land (i.e., islands) can be represented and associated with a single feature in the attribute table. However, if the distinction of features is important, such as identifying the names of islands in the Haida Gwaii archipelago, then a singlepart geometry should be used (Figure **??**).

Singlepart geometry of the Haida Gwaii archipelago off the west coast of British Columbia, Canada. Hover over the islands to see the names. Animated figure can be viewed in the web browser version of the textbook. Polygon data from Statistics Canada and island placenames from Natural Resources Canada. Open Government License - Canada.

Although it is possible to convert from a multipart geometry into singlepart geometry, you need to carefully consider how your features should be represented in the attribute table. For example, if you will be undertaking calculations using area or perimeter of the constituent polygons that comprise a multipart geometry of Canada, then you will return a single value for all of Canada while singlepart geometry would return values for each individual polygon. As well, area calculations can vary between multipart and singlepart geometry. For example, approximately 27% of Canada's land area (including freshwater), is comprised of more than 52,000 islands, which is a statistic you could only calculate with singlepart geometry. Thus, the choice of representing a feature using singlepart or multipart geometry should be based on how the features will be used in your analysis (i.e., aggregated versus disaggregated).

7.7.2 Holes

When dealing with polygon features, **holes** may occur, which represent discontinuity of the interior polygon space. Imagine the case of a forested land cover that surrounds a lake. If we consider the forested land cover polygon on its own, then the polygon will have a hole where the lake exists (Figure **??**.

Topologically, holes in polygons imply that another polygon shares an adjacent boundary where the hole exists, for example, from the union of two layers (see Chapter 6). In our example, the lake would comprise its own polygon that

Figure 7.8: Conceptual forest land cover polygon that contains a lake causing a hole. Pickell, CC-BY-SA-4.0.

would completely fill the hole.

7.7.3 Delaunay triangulation

Delaunay triangulation is method for forming a triangle mesh over a set of points. The Delaunay triangulation method (Delaunay 1934) connects all points in a set such that no point in the set lays *within* a circumcircle formed by any of the triangles in the mesh. A circumcircle is a circle that passes through all the vertices of a cyclic polygon such as a triangle. In other words, the circumcircles are empty. To illustrate this, consider the four points in Figure **??**. There are only two circumcircles that can be formed from this set of points that ensures that no point lays within a circumcircle. The triangulation is then simply the lines connecting the three points that fall on any given circumcircle. One important property of the Delaunay triangulation is that the smallest angle in the resulting triangles is maximized from the circumcircle fitting, which minimizes sliver triangles that might from with very shallow angles.

Figure **??** shows a Delaunay triangulation for a set of 50 points. We can see that sliver triangles mostly occur on the edge of the extent of the points. Delaunay triangulations can be performed both in 2- and 3-dimensional Euclidean space and are therefore important for representing 3D surfaces as well as performing spatial estimation over 2D areas from a set of points.

7.7.4 Multipatch geometries

7.7.5 Thiessen polygons

Thiessen polygons are an implementation of a nearest neighbour algorithm in Euclidean space: given some set of input point features mapped on a plane, partition the plane into polygon areas that represent the nearest locations on the plane to those points. These resulting polygons are also sometimes referred to as proximal polygons, representing the proximal areas given some set of points. When Thiessen polygons are created for geographic data, the resulting

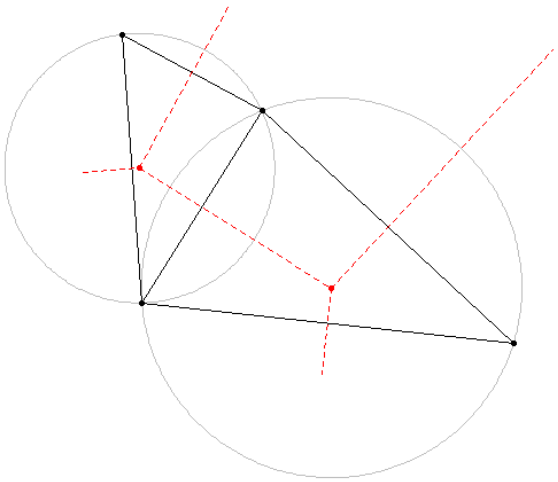


Figure 7.9: Delaunay triangulation of four points. Black lines show the triangulation, grey lines represent the circumcircles connecting the three points of each triangle, red points represent the centres of the circumcircles, and the red dotted lines show that connecting the centres of the circumcircles forms the Voronoi diagram. Pickell, CC-BY-SA-4.0.

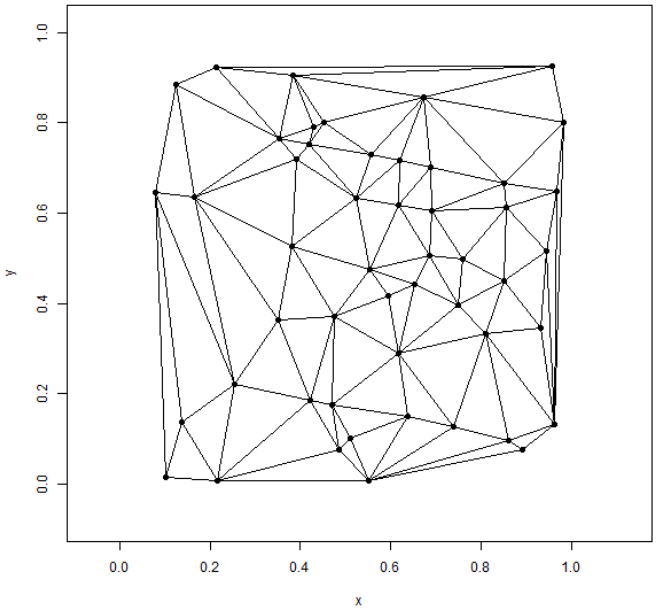


Figure 7.10: Delaunay triangulation of 50 random points. Pickell, CC-BY-SA-4.0.

diagrams are called **Voronoi diagrams** and sometimes referred to as Voronoi maps (Figure **??**. Voronoi maps have many uses such as partitioning geographic space into areas that are nearest to weather stations, airports, or cellular towers. Thiessen polygons can be intersected with other geographic data layers in a GIS using map algebra to efficiently solve proximal questions like, "what is the nearest X?" without having to search or calculate the exact distances of all nearby features, which can be computationally time-consuming (Okabe et al 2007).

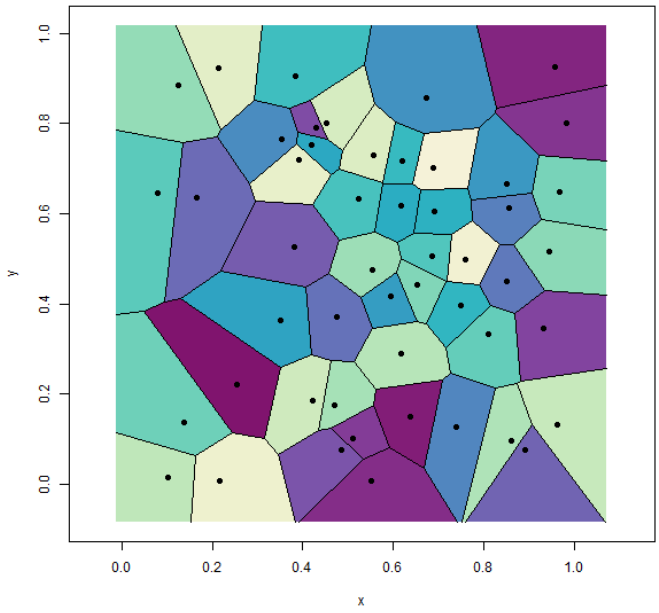


Figure 7.11: Thiessen polygons of 50 random points. Pickell, CC-BY-SA-4.0.

Thiessen polygons are a product of Delaunay triangulation described in the previous section. Figure **??** shows the relationship between the points, triangulation, circumcircles, and the Thiessen polygons. Connecting the circumcentres of the circumcircles produces the Voronoi diagram (Figure **??**).

7.7.6 Centroids

A **centroid** is a point that represents the geometric centre of a polygon. For convex polygons, the centroid will always lay within the polygon, but for concave polygons, the centroid may lay outside the polygon (Figure **??**. Circular polygons always have centroids that are equidistant to the boundary of the polygon (Figure **??**.

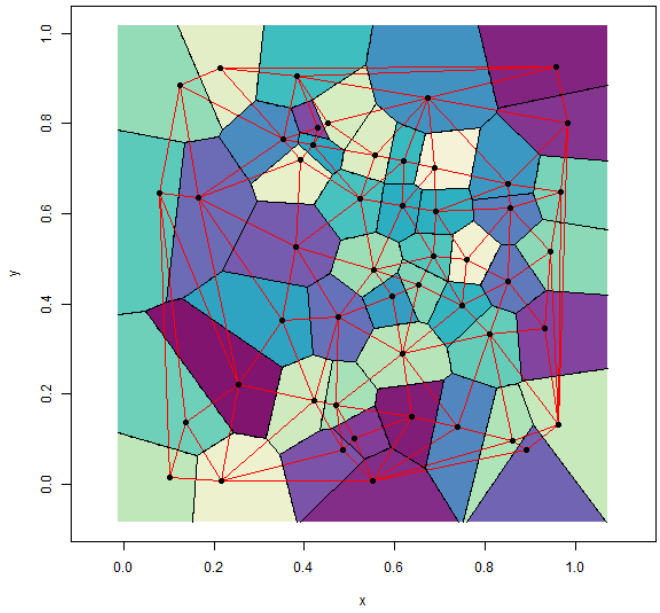


Figure 7.12: Delaunay triangulation (red lines) overlaid onto the Thiessen polygons. Pickell, CC-BY-SA-4.0.

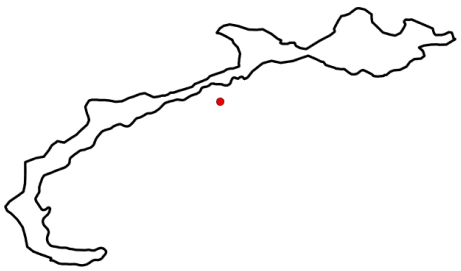


Figure 7.13: Concave polygon with the centroid (red dot) laying outside its boundary. Pickell, CC-BY-SA-4.0.

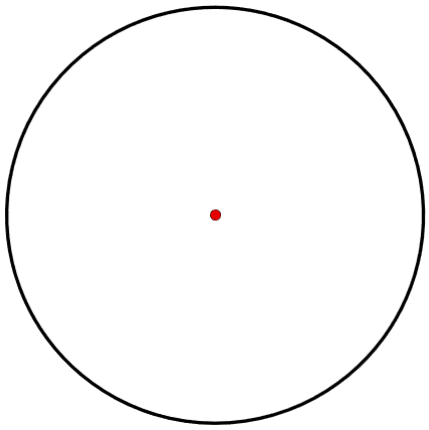


Figure 7.14: Circle polygon with the centroid (red dot) laying equidistant from the boundary of the polygon. Pickell, CC-BY-SA-4.0.

7.7.7 Convex hull

A **convex hull** is the smallest polygon that contains some set of points. It is sometimes also referred to as a "convex envelope" or "convex closure" because the perimeter of the polygon is formed by connecting the outermost points and closing or enveloping the remaining points. The convex hull is therefore the mathematical implementation of **topological closure**, where closure refers to the smallest closed set of points that contain the set of points. In practice, the convex hull is a bit like applying a rubber band around the outermost points so that the tension of the rubber band forms straight lines between the pairs of points in the closed set (Figure **??**). There are several algorithms for computing the convex hull, including Jarvis' March (Jarvis 1974), Graham's Scan (Graham 1972), quickhull (Barber et al 1996), and CudaHull (Stein et al 2012).

Convex hulls are easily drawn by hand and are used for identifying a natural boundary for a sample set of points. Formally, the calculation is

7.7.8 Convex alpha hulls and alpha shapes

Convex hulls can be generalized to the concave case, called **convex alpha hulls** or **-shapes (alpha shapes)**, by adjusting the maximum radius of the circumcircles through a parameter, alpha α. The objective of a convex alpha hull is to minimize the -shape formed by circumcircles of radius less than or equal to α. Similar to the Delaunay triangulation, the circumcircles must be *open*, meaning they contain no other points in the set. Defined in this way, the final -shape may not result in closure of the full set of points and can result in holes where the distance between points exceeds 2α. Surprisingly, -shapes are prone to not existing at all. For the case of $\alpha = 0$, applying circumcircles

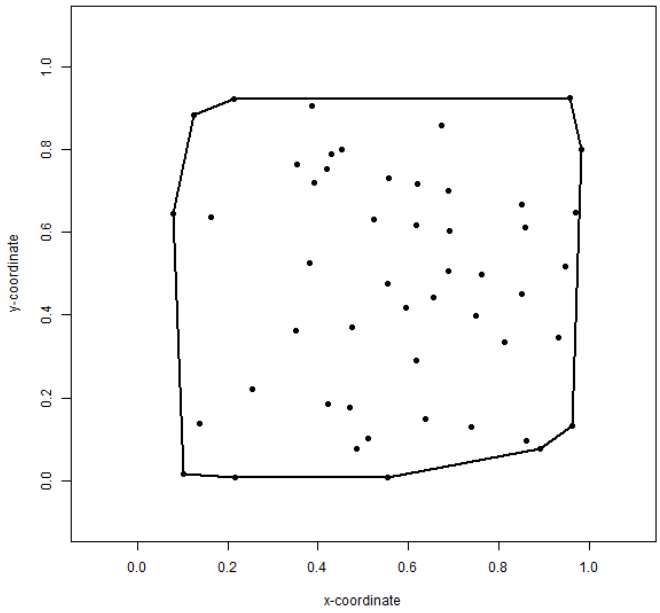


Figure 7.15: Convex hull formed by topological closure of the smallest closed set of points around the entire set of points. Arrangement of the points are the same as in the Thiessen polygons figure above. Pickell, CC-BY-SA-4.0.