ABOUT THIS BOOK

Publishing Information

Published by Scholarly Communications and Digital Publishing, WashU Libraries, Washington University in St. Louis, with funding from the DI2 Accelerator Award for Open Education Resources.

How to Use this Book

Introduction to Geospatial Thinking and Open Source GIS was developed for beginner GIS practitioners to learn foundational concepts and key skills to get started using geospatial data and GIS.

Part one introduces the concepts of geospatial thinking, how GIS works, coordinate reference systems and projection, effective communication, data management and mapping out a project. Part two provides opportunities to apply the concepts by conducting basic GIS actions such as adding data, changing projection, selecting data, exporting data, joining datasets, merging datasets, editing features, classifying data, and laying out maps, etc.

The activities are designed to be used with the open-source GIS software QGIS and with openly available datasets.

Navigating the book

This textbook has a table of contents to help you navigate through the book more easily. If using the online webbook, you can find the full table of contents on the book's homepage or by selecting "Contents" from the top menu when you are in a chapter. Inside a chapter, you can use the keyboard arrows to go to the previous or next chapter.

Searching

If using the online webbook, you can use the search bar in the top right corner to search the entire book for a key word or phrase. To search a specific chapter, open that chapter and use your browser's search feature by hitting **[Ctrl]** + **[F]** on your keyboard if using a Windows computer or **[Command]** + **[F]** if using a Mac computer.

The **[Ctrl]** + **[F]** and **[Command]** + **[F]** keys will also allow you to search a PDF, HTML, EPUB, and MOBI files if you are reading them on a computer.

If using an eBook app to read this textbook, the app should have a built-in search tool.

Acknowledgements

I would like to acknowledge the authors of the resources from whom I adapted (Briney, Campbell, Fang, Rosenfeld) or cited; all of the contributors to the QGIS software community over the years — a shining example of community-built, open-source tools; and to the catalog of free icons from the Noun Project.

…

To make this book happen, I would like to acknowledge and thank the Department of Scholarly Communications in the University Libraries at WashU. I'd like to especially call out Treasa Bane, Cadence Cordell, and Emily Stenberg for supporting me through the development. Thanks to WashU's Digital Intelligence and Innovation Accelerator who supported the OER program, namely Dan Maranan and Christy Potthast.

…

Immense, grateful shout out to the mentors who introduced me to geospatial data and GIS, Aaron Addison, Michael Frachetti, Cheryl Morton, and Melisa McLean. To the colleagues that encouraged me as I built my GIS skills, Cynthia Hudson Vitale, Cindy Traub, Melissa Vetter, Mollie Webb, and Bill Winston. And to the team of babysitters who watched my eldest son when I was a single mom, learning GIS for the first time, after hours, Jaleh Fazellian. Makiba Foster, and Evie Hemphill. Thanks also to Nadia Ghasedi, Rachel

Glick, and Kelley Wallace.

...

Finally I would like to acknowledge my little family, Ian, Ralph, and Theo who suffered late meals, distracted conversations, and my general grumpiness while I worked.

Licensing Information

Accessibility Statement

Accessibility features of the web version of this resource

The web version of **Introduction to Geospatial Thinking and Open GIS** has been designed with accessibility in mind by incorporating the following features:

- It has been optimized for people who use screen-reader technology.
- All content can be navigated using a keyboard
- Links, headings, and tables are formatted to work with screen readers
- Images have alt tags
- Videos have captions with transcripts available for download
- Information is not conveyed by color alone.
- There is an option to increase font size (see tab on top right of screen).

The web version of this resource follows all guidelines in the BCcampus Checklist for Accessibility.

Alternative Formats

In addition to the web version, this book is available in a number of file formats including PDF, EPUB (for eReaders), MOBI (for Kindles), and various editable files. This textbook is available in the following formats:

- **Online webbook**. You can read this textbook online on a computer or mobile device in one of the following browsers: Chrome, Firefox, Edge, and Safari.
- **PDF**. You can download this book as a PDF to read on a computer (Digital PDF) or print it out (Print PDF).
- **Mobile**. If you want to read this textbook on your phone or tablet, you can use the EPUB (eReader) or MOBI (Kindle) files.
- **HTML**. An HTML file can be opened in a browser. It has very little style so it doesn't look very nice, but some people might find it useful.

You can access the online webbook and download any of the formats for free here: Introduction to Geospatial Thinking and Open GIS. To download the book in a different format, look for the "Download this book" drop-down menu and select the file type you want.

Known accessibility issues and areas for improvement

While we strive to ensure that this resource is as accessible and usable as possible, we might not always get it right. Any issues we identify will be listed below. There are currently no known issues.

This statement was last updated on October 10, 2025.

Contact Us

Please contact Scholarly Communications and Digital Publishing, WashU Libraries, at digital@wumail.wustl.edu or the author Jennifer Moore, if you have any questions, comments, corrections, or concerns with the text.

An Introduction to Geospatial Thinking and Open Source GIS

AN INTRODUCTION TO GEOSPATIAL THINKING AND OPEN SOURCE GIS

JENNIFER MOORE

WashU Libraries
Saint Louis (Mo.)

Much of part 1 of this book was adapted (with significant updates) from previous works.

Chapters 1 – 4, and 7 were adapted, with significant changes to structure, prose, and images, from:

- Campbell, Jonathan. n.d. *Essentials of Geographic Information Systems*. 2011. Saylor Foundation. Accessed September 1, 2020. https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/.
- Fang, Yiping. n.d. "Spatial Thinking in Planning Practice: An Introduction to GIS." Open Textbook Library. Accessed August 17, 2020. https://open.umn.edu/opentextbooks/textbooks/234.
- Rosenfeld, Christine, and Nathan Burtch. n.d. *Human Geography*. Accessed September 30, 2025. https://viva.pressbooks.pub/humangeog/.

Chapter 5 was adapted to emphasize geospatial data considerations from:

- Briney, Kristin. n.d. *The Research Data Management*

Workbook. Accessed April 11, 2025.
https://caltechlibrary.github.io/RDMworkbook/.

CONTENTS

Part II. Part 2: Hands-On Workbook

INTRODUCTION: WHY GEOSPATIAL THINKING AND OPEN SOURCE GIS

Research methods are currently changing rapidly and will continue to change even more so as generative artificial intelligence (AI) further develops. Regardless of AI's progressive integration into geospatial research, it is still essential for humans to understand the fundamental geospatial principles, data, and functions. Otherwise, we will not be able to interpret or evaluate results.

The goal of this book is to provide a flexible, open, and timeless resource for geospatial thinking, and to open-source GIS. However, as I typed that, it became out of date.

What do I mean by flexible? In teaching, I take a problem-based approach with scaffolded topics and skills which increase in complexity as the class moves on. However, I also love all things modular, so by flexible I mean I want it to grow in complexity but also want each lesson to stand alone if teachers/learners want to focus on a given area.

What do I mean by open? It is my intention to use open tools, open data, and open materials to the greatest extent possible. Because some great research and scholarship have been published behind the wall, I'm going to reference those fine folks and cite them. Obviously if you have made it this far you realize you're reading (are you reading?) an open textbook and if you find anything useful, I encourage you to reuse it too.

What do I mean by timeless? To the greatest extent possible, I hope to develop the text focused on key principles and skills which persist even if software versions and datasets require updates. Although I believe this is a worthwhile effort, it's not possible to predict what our rapidly changing technology will bring, especially given the generative AI renaissance, which is likely to impact how we work geospatially. And although I truly believe learning geodesy concepts never go out of style, I also hope something useful persists from the exercises. In any case, because it's open and extensible, it can be reshaped at any time.

What's the difference between spatial thinking and geospatial thinking? These terms are often used interchangeably. Spatial thinking can refer to objects with multiple dimensions but are not necessarily tied to geography. Geospatial is intrinsically geographic.

PART I

PART 1: WHAT YOU SHOULD KNOW

What to Expect from Each Chapter

Chapter 1: Geospatial Thinking. This chapter will provide a high-level description of what it means to ask geospatial questions and provide examples of questions asked in various fields.

Chapter 2: In How GIS Works the elements of a GIS will be described, including data, hardware, tools, and software.

Chapter 3: Location, Location, Location describes the shape of the earth, measuring locations on the 3D globe and transforming those into a 2D plane.

Chapter 4: The Nature of Geospatial Data will unpack the data formats in which we encounter and store geospatial data.

Chapter 5: Geospatial Data Management is an essential chapter to introduce you to effective file management when dealing with messy data, as geospatial data can be.

Chapter 6: Project Planning. In this chapter there will be an overview of how to set up a project for success.

Chapter 7: Visually Representing Geospatial Thinking describes types of maps and how they are used. It covers the use of symbols and colors, and classification.

CHAPTER 1: GEOSPATIAL THINKING

Adapted from Campbell and Rosenfeld

Everything happens somewhere, that somewhere is known as "place." Examining where a thing happens may help us understand:

- What happened
- When it happened
- How it happened
- Why it happened
- The relationship and impact of a phenomenon and place
- What/where something might happen next.

Whether it is an outbreak of a highly contagious disease, the discovery of a new frog species, the path of a deadly tornado, or the nearest location of a supermarket, knowing something about where things happen is important to understanding and relating to society, nature, our locality, and to the world at large.

Functions of (Geo)spatial

Thinking

The National Research Council describes spatial thinking as the constructive amalgam of three elements: concepts of space, tools of representation, and processes of reasoning. And its functions are defined as:

- Description: capturing, preserving, and conveying the appearances and relationships among objects
- Analysis: function, enabling an understanding of the structure of objects
- Inference: generating arguments about spatial questions.

(National Research Council et al. 2006)

We might reframe these functions for geospatial thinking

- Descriptive: capturing and curating place and objects within it
- Analytical: understanding the structure and relationships of objects primarily in place, and in relationship to time, and context.
- Inferential: generating questions and answers about place and the objects within it.
- Communicative: creating purposeful, clean, and informative visual descriptions of place.

Asking Geospatial Questions

A geographic question seeks to understand place and how we relate to it. Such questions can be simple with a local focus (e.g., "Which way is the nearest hospital?") or more complex, with a more global perspective (e.g., "How is urbanization impacting biodiversity hotspots around the world?"). The thread that unifies such questions is geography. For instance, the question of "where?" is an essential part of the questions "Where is the nearest hospital?" and "Where are the biodiversity hotspots in relation to cities?" Being able to articulate questions clearly and to break them into manageable pieces are very valuable for geospatial problem solving.

John Snow's 1854 Cholera map

John Snow's Cholera Map is the most cited and earliest example of spatial thinking impacting public health. Cholera was considered a respiratory disease, but as a doctor particularly interested in the developing field of anesthesiology, John Snow, and his colleagues felt the symptoms Cholera victims displayed did not appear to be respiratory in nature. They believed it was a waterborne illness. Snow focused on Central London (*concept of space*) and mapped *(tools of representation*) the area including the wells and added incidents of Cholera deaths. Based on his map, it appeared the cholera epidemic's epicenter was the Broad Street well (*process of reasoning).*

Learning how to ask an effective question takes practice and is often more difficult than finding the answer itself. However, when we ask an effective question, problems are more easily solved and our understanding of the world is improved. There are five general types of geographic questions that we can ask. Each type of question is listed here and is also followed by a few examples (Nyerges 1991).

Questions about geographic location:

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

Questions about geographic distribution:

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

Questions about geographic

association (proximity):

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

Questions about geographic interaction (relationships):

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

Questions about geographic change:

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

Why Don't We Just Call It Geography?

Geospatial thinking is an interdisciplinary descendant of geography. Classic geography focused on understanding the shape of the earth and the location of things. As geography evolved over centuries from a purely descriptive and cartographic endeavor to a research-focused field, it became more concerned with complex spatial relationships. However, it's the applied interdisciplinarity which emerged in the mid-twentieth century that really helped the concept of geospatial thinking to take shape. The table below provides geographic topics which are frequently asked by experts from various areas of expertise, industries, and professions.

Field	Example Place Topics
Urban planners, traffic engineers, demographers	Understanding the commuting patterns between cities and suburbs (geographic interaction).
Biologists and botanists	Why one animal or plant species flourishes in one place and not another (geographic location/distribution).
Epidemiologists and public health	Where disease outbreaks occur and how, why, and where they spread (geographic change/interaction/location
Climatologists	The causes and consequences of global warming
Epidemiologists	Locating ground zero of a virulent disease outbreak
Archaeologists	Reconstructing an ancient site
Political consultants	Developing campaign strategies

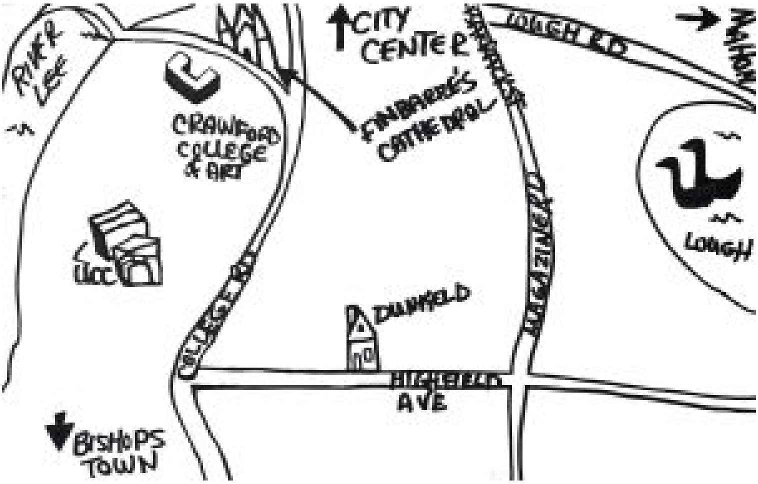
The Power and Responsibility of Mapping

Mapping is a way to "write the world" and therefore is a kind of representation. The art and science of creating maps is known as cartography. With cartographic representation, especially in trying to represent a 3D earth on a 2D surface, comes distortion, which shapes our geographical knowledge of the world and affects our perceptions of place. Maps have power. Historically, maps were created for and by those with political power. Maps were used to demarcate and claim ownership of land; a means of showing everyone that one has planted their proverbial flag to claim possession. As such, maps have an air of authority and of being "official" and "true." But all maps are made for specific purposes and can have both accidental and purposeful lies included. The accidental, small white lies of a map most often have to do with the creation of a representation and generalization.

Maps have scale and will show a larger geographic area than the size of the paper they are printed upon. Because the map is not at a 1:1 scale to the world, objects must be simplified or generalized to not be overly cluttered by the infinite complexity of the world. What is shown, and just as importantly not shown, are choices made by the cartographer. A map might show the locations of important socioeconomic features, like schools and mines, but not show features like landfills or prisons. Labeling maps and adding toponyms (names of places) is also a representative practice and reminds us of the power involved in the act of representing. Labels assign meaning to the area mapped here from the mapmaker's perspective, but do not necessarily reflect the meaning the land has for others. Colonial era maps often used the label *terra nullius*, a Latin term meaning "nobody's land" or "empty land." Maps with this label played a role in justifying colonialism, as colonizers used the supposed reality, as marked on a reliable map, that land was unoccupied and therefore up for grabs. Labels like this ignore and silence the presence of indigenous people and meanings of land held by non-mapmakers.

Mental Map

The purpose of this chapter is to increase our geographic awareness and to refine our geospatial thinking. Mental or cognitive maps are tools that we all use every day. As the name suggests, mental maps are maps of our environment stored in our brain. We rely on our mental maps to get from one place to another, to plan our daily activities, or to understand and situate events that we hear about from our friends, family, or the news. Mental maps also reflect the amount and extent of geographic knowledge and spatial awareness that we possess.



Example of a mental map

When mapping a place from memory, we learn from what is included and what is excluded, it tells us what's important to the map maker.

Exercises

Exercise 1: Create a Mental Map (10 mins)

Using a blank sheet of paper, draw a mental map of a place you know well to help a newcomer get around. Do not look at an existing map or any other references. When you finish your map show it to someone else and ask for their feedback on your geospatial thinking and curation.

CHAPTER 2: HOW GIS WORKS

Adapted from Campbell and Rosenfeld, respectively.

Geographic Information Systems (GIS) allows us to transform our geospatial thinking to answers by layering data and conducting complex analyses to produce in-depth understandings of place. A geographic information system is made up of computer software, a collection of computer hardware, services distributed and accessed via the internet, and tools, but GIS is also a science. It's used in research and business.

Data

Geospatial data refer to the real-world geographic objects of interest, such as streets, buildings, lakes, and countries, and their respective locations. In addition to location, each of these objects may also possess certain traits of interest, or attributes, such as a name, number of stories, depth, or population; GIS links the two types of data, raster and vector, together to create information and facilitate analysis.

These data can come from a range of sources. It can be collected through terrestrial surveying, or above ground by using remote sensing technologies or LiDAR. Remote sensing refers to collecting data from afar. This is typically achieved through cameras, videos, and sensors attached to aircraft including airplanes, satellites, and drones. It allows us to gather information about the Earth's surface and track landscape change over time.

Hardware

Computer, memory, storage, scanners, global positioning system (GPS) and other physical components, might make up hardware. The configuration of these hardware can vary widely in size. Although most of us work from a single terminal or laptop that leverages cloud services and tools, if the computer is situated on a network, the network and servers can also be considered an integral component of the GIS because it enables us to share data and information that the GIS uses as inputs and creates as outputs.

Tools

As a tool, GIS permits us to maintain, analyze, and share a wealth of data and information. GIS is actually a collection of tools that can be leveraged for a given geospatial question. From the relatively simple task of mapping the path of a hurricane to the more complex task of determining the most efficient garbage collection routes in a city, GIS is used across the public and private sectors. It's not just a tool to analyze data, it's also a tool to represent insights. Aside from traditional paper maps and presentations, through online and mobile mapping, navigation, and location-based services, GIS is personalizing and democratizing geospatial technology by bringing maps and mapping to the masses.

Software

GIS consists of a special type of computer program capable of storing, editing, processing, and presenting geospatial data and information such as maps. The most well-known GIS software provider is Environmental Systems Research Institute Inc. (Esri), which distributes a suite of software called ArcGIS, but there are many others including open-source, community-built software such as QGIS; these are both examples of desktop software with a graphical user interface (GUI). There are web-based software as well, such as ArcGIS Online (AGOL) and CartoDB. All GIS software, regardless of vendor, consists of a database management system that is capable of handling and integrating two types of data: geospatial data and attribute data. That said, it is also possible to work with geospatial data using programming languages such as R and Python; although on their own the capabilities are limited, combined with other tools, they can be very powerful.

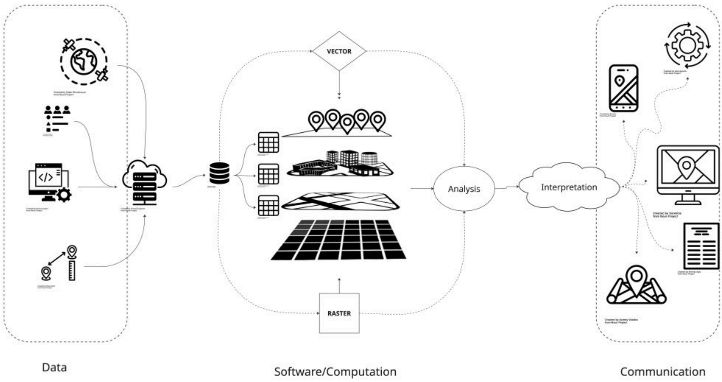


Illustration of GIS components

SOFTWARE	DESCRIPTION	PLATFOR
d3	Javascript library for data visualization, including maps	Code
Folium	Python library to prepare interactive maps for Leaflet	Code
Geemap		Code
Geopandas	Library to make working with geospatial data in python easier. GeoPandas extends the datatypes used by pandas to allow spatial operations on geometric types.	Code
GeoServer	Allows users to share and edit geospatial data.	Web ser
GRASS GIS	Geospatial data management, vector and raster manipulation	Deskto
gvSIG	Mapping and geoprocessing with a 3D rendering plugin	Deskto
ILWIS	(Integrated Land and Water Information System) Integrates image, vector and thematic data.	Deskto

IPyleaflet	Python library to build interactive maps combining jupyter notebooks and leaflet	Code
OpenJUMP	Formerly known as Java Unified Mapping Program (JUMPGIS), a Java based vector and raster GIS and programming framework.Current development continues under the OpenJUMP name	Deskto
Leaflet	Build interactive maps using javascript	Code
LiDar	Python library to visualize LiDAR and make DEM	Code
MapGuide Open Source	Runs on Linux or Windows, supports Apache and IIS web servers, and has APIs (PHP, .NET, Java, and JavaScript) for application development.	Web ser
Mapnik	C++/Python library for rendering – used by OpenStreetMap.	Web ser
MapServer	Written in C. Developed by the University of Minnesota.	Web ser
MapWindow GIS	Free desktop application with plugins and a programmer library	Deskto
QGIS	(previously known as Quantum GIS) Powerful cartographic and geospatial data processing tools with extensive plug	Deskto

RGISLib	Remote sensing Python library	Code
SAGA GIS	(System for Automated Geoscientific Analysis) Tools for environmental modeling, terrain analysis, and 3D mapping	Deskto
uDig	API and source code (Java) available.	Deskto

There many great open-source tools for mapping and working with geospatial data. The workbook will focus on QGIS which is a robust GIS platform with a GUI.

Exercises

Exercise 2: QGIS installation exercise

QGIS logo

Go to The QGIS web site and download the latest, stable version. Please note, the version you download may be more recent than the version used in this textbook, but you should be able to adapt. Remember, if you can't find the operation, you can always ask the internet. There are many, many resources to help you through.

CHAPTER 3: LOCATION, LOCATION, LOCATION

Adapted from Fang

Location-based analyses distinguished geography from other fields. Location is the key to accessing geospatial research. There are many ways to describe location.

Shape of the Earth

Illustration of the shape of the earth

The Earth's constant spinning causes it to bulge slightly along the equator, ruining its perfect spherical shape; it's a geoid with a slight pear shape; it is a little larger in the southern hemisphere and includes other bulges. The slightly oval nature of the Earth's geometric surface makes the terms ellipsoid and oblate spheroid more accurate in describing its shape, but they are not perfect terms either due to differences in material weights (for instance iron is denser than sedimentary deposits), and the movement of tectonic plates makes the Earth dynamic and constantly changing.

Absolute Location

This is precise position on the surface of the Earth, most often expressed as coordinate pairs (Latitude and Longitude). Global positioning systems (GPS) can determine absolute coordinates on Earth. GPS technology consists of a constellation of satellites that are orbiting the earth and constantly transmitting time signals. To determine a position, earth-based GPS units (e.g., handheld devices, car navigation systems, mobile phones) receive the signals from at least three of these satellites and use this information to triangulate a location. All GPS units use the geographic coordinate system (GCS) to report location. Location can also be expressed as an address. Addresses can be brought into a GIS through a process called geocoding, which uses an existing reference dataset, or a locator, to match an address to a location.

Nominal Location

Nominal location refers to named geographic features or place names; these locations can be cities, countries, administrative names (e.g., Siberia), or natural features (e.g., Rocky Mountains). Place names are organized in gazetteers, the most standardly used gazetteer is Geonames. Gazetteers can also be used as a locator to bring place names into a GIS.

Relative Location

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

Direction

Direction can be used a benchmark to establish relative location. South Poles serve as the geographic benchmarks for determining direction. Magnetic north (and south) refers to the point on the surface of the earth where the earth's magnetic fields converge. This is also the point to which magnetic compasses point. Note that magnetic north falls somewhere in northern Canada and is not geographically coincident with true north or the North Pole. Grid north simply refers to the northward direction that the grid lines of latitude and longitude on a map, called a graticule.

[1]

Geographic Coordinate System (Latitude, Longitude)

A geographic coordinate systems measure the spherical Earth based on a common datum. This measurement is expressed in coordinate pairs, latitude and longitude. Earth's center to a point on the Earth's surface.

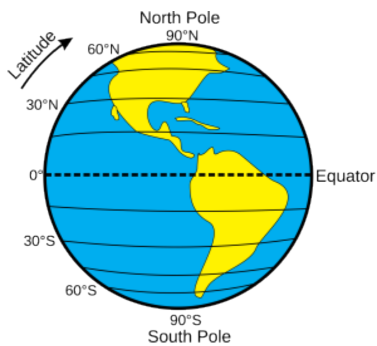


Illustration of latitude lines

Latitude

Across the spherical Earth, latitude lines ***(parallels)*** stretch horizontally from east to west, and they are parallel to each other, hence their alternative name, parallels. Latitude can be thought of as the lines that intersect the y-axis, and longitude as lines that intersect the x-axis.

1. Adapted from fang

Longitude

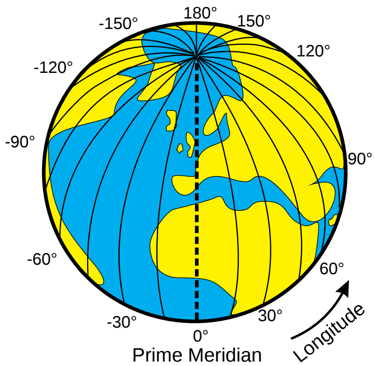


Illustration of longitude lines

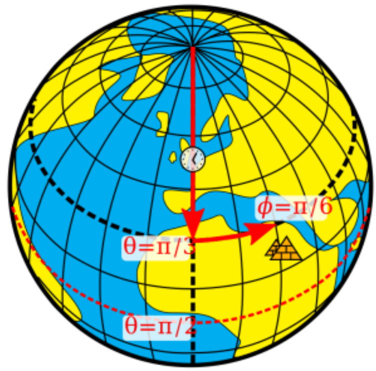
Longitude lines (***meridians***) stand vertically and stretch from the North Pole to the South Pole. The y axis is the prime meridian, which is a line running from pole to pole through Greenwich, England.

Together these "north to south" and "east to west" lines meet at perpendicular angles to form a graticule, a grid that encompasses the Earth. Just as the upper right quarter in the Cartesian coordinate system is positive for both x and y, latitude and longitude east of the prime meridian and north of the equator are both positive.

+ / -

Europe, Asia, and part of Africa – which have positive latitudes and longitudes – correspond to the upper right quarter of the Cartesian coordinate system. With the exception of some U.S. territories in the Pacific and the westernmost Aleutian Islands, all of the United States is north of the equator and west of the prime meridian, so all latitudes in the U.S are positive (or north) while almost all longitudes are negative (or west).

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



The point of intersection between the Prime Meridian and the Equator (0,0)

Datum

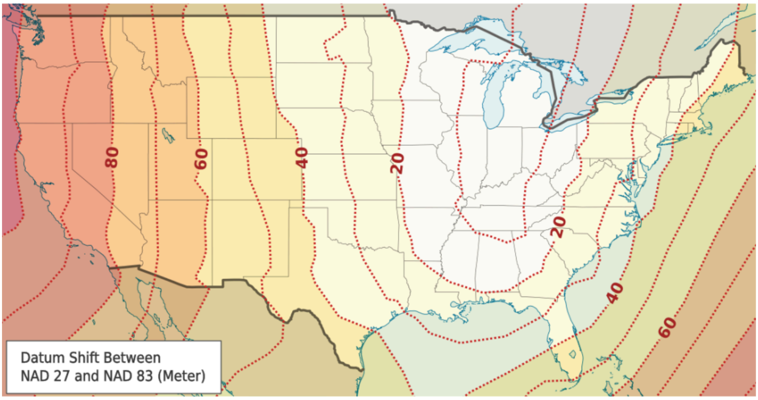


Illustration of datum the division between NAD27 and NAD83

All coordinate systems are measured relative to a datum. A datum defines the starting point from which coordinates are measured. Latitude and longitude coordinates, for example, are determined by their distance from the equator and the prime meridian that runs through Greenwich, England. But where exactly is the equator? And where exactly is the Prime Meridian? And how does the irregular shape of the Earth figure into our measurements? All of these issues are defined by the datum. Many different datums exist, but in the United States only three datums are commonly used.

- ***NAD 27:*** The North American Datum of 1927 (NAD27) uses a starting point at a base station in

Meades Ranch, Kansas and the Clarke Ellipsoid to calculate the shape of the Earth.

- ***NAD 83:*** Thanks to the advent of satellites, a better model later became available and resulted in the development of the North American Datum of 1983 (NAD83). Depending on one's location, coordinates obtained using NAD83 could be hundreds of meters away from coordinates obtained using NAD27.
- ***WGS 84***: A third datum with a starting point in Greenwich, England, the World Geodetic System of 1984 (WGS84) is identical to NAD83 for most practical purposes within the United States. The differences are only important when an extremely high degree of precision is needed. WGS84 is the default datum setting for almost all GPS devices and web maps. But most USGS topographic maps published up to 2009 use NAD27.

Projection (Transformation of Geographical Coordinates to Cartesian Coordinate Systems)

While the system of latitude and longitude provides a consistent referencing system for anywhere on the earth, to portray our information on maps or for making calculations, we need to transform these angular measures to Cartesian coordinates. These transformations amount to a mapping of geometric relationships expressed on the shell of a globe to an attenable surface — a mathematical problem.

Why project

Globes do not need projections, and even though they are the best way to depict the Earth's shape and to understand latitude and longitude, they are not practical for most applications that require maps. We most often experience 2D maps, whether physical or digital; reshaping a globe requires a mathematical reshaping of the Earth's 3-dimensions into a 2-dimensional surface.

Illustration of an orange representing a 3D earth unpeeled to represent a 2D map

To illustrate the concept of a map projection, imagine that we place a light bulb in the center of a translucent globe.

On the globe are outlines of the continents and the lines of longitude and latitude called the graticule. When we turn the light bulb on, the outline of the continents and the graticule will be "projected" as shadows on

Illustration of a light bulb in the center of the globe radiating light. Above the globe is a 2D world map which is situated within the rays of light. This illustration represents using projection to flatten the 3D globe into a 2D map.

the wall, ceiling, or any other nearby surface. This is what is meant by map "projection." The term projection implies that the ball-shaped net of parallels and meridians is transformed by casting its shadow upon some ‘at, or ‘attenable, surface. In fact, almost all map projection methods are mathematical equations. The projected graticule produced by projection equations in each category. Referring again to the previous example of a light bulb in the center of a globe, note that during the projection process, we can situate each surface in any number of ways. For example, surfaces can be tangential to the globe along the equator or poles, they can pass through or intersect the surface, and they can be oriented at any number of angles. As you might imagine, the appearance of the projected grid will change quite a lot depending on the type of surface it is projected onto, and how that surface is aligned with the globe. The following figures shows how these projections can vary.

Why Not to Project: Small Study Area

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

Projection Categories:

Within the realm of maps and mapping, there are three surfaces used for map projections (i.e., surfaces on which we project the shadows of the graticule). These surfaces categories are:

Cylindrical

- straight coordinate lines
- horizontal parallels crossing meridians at right angles
- all meridians are equally spaced
- the scale is consistent along each parallel

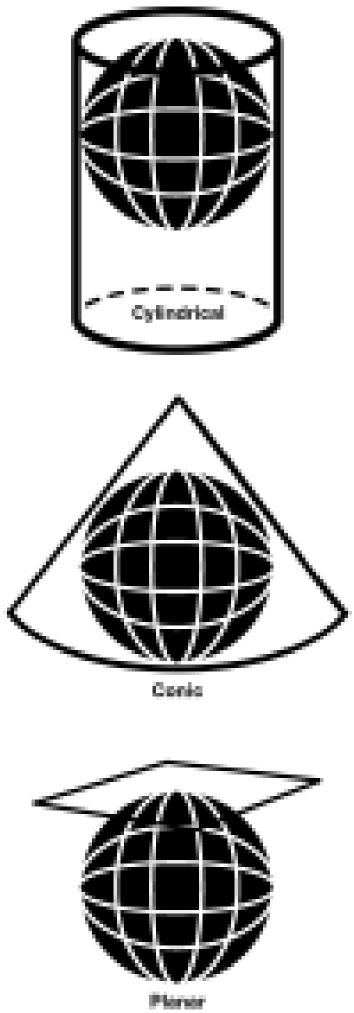


Illustration of projection categories

Conic

The cone is typically aligned with the globe such that its line of contact (tangency) coincides with a parallel in the mid-latitudes:

- meridians are equidistant & straight lines
- converge in locations regardless of presence of a pole
- parallels cross the meridians at right

angles
- constant measure of distortion throughout

Planar (Azimuthal)

The planar is centered upon a specific point and perspective and is frequently positioned tangent to the equator:

- the point can be based on the poles (stereographic), the equator (gnomic) or from a defined oblique point (orthographic).
- a point specifies the focused of projection identified by a central coordinate
- polar projections are most straightforward, and often cited
- meridian and parallels intersect at their true position
- most accurate at the focal point, patterns of distortion move out from it in straight line measurement creation "great circles".

Projection and Distortion

Flattening the globe cannot be done without introducing some error, and some distortion is unavoidable. Any projection has its area of least distortion. Projections can be shifted around to put this area of least distortion over the topographer's area of interest. Thus, any projection can have an unlimited number of variations or cases that determined by standard parallels or meridians that adjust the location of the high-accuracy part of the projection.

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

Within each projection family there are many different projections to choose from. All of them create distortion in one manner or another. Determining what needs to be preserved about your study area is the first step. ***Map projections are often classified by the characteristic they do not distort***. Usually only one property is preserved in a projection.

Classifications

- **Conformal** – preserves shape, and distorts area
- **Equal Area** – preserves: area, distorts shape, scale, or angle (bearing)
- **Equidistant** – preserves distances between certain points (but not all points) and distorts other distances

Conformal

Conformal map projections are used for navigational purposes due to the importance of maintaining a bearing or heading when traveling great distances. The cost of preserving bearings is that areas tend to be quite distorted in conformal map projections. Although shapes are more or less preserved over small areas, at small scales areas become wildly distorted.

The Mercator projection is an example of a conformal projection and is famous for distorting Greenland. As the name indicates, equal area or equivalent projections preserve the quality of area. Such projections are of particular use when accurate measures or comparisons of geographical distributions are necessary (e.g., deforestation, wetlands). To maintain true proportions in the surface of the earth, features sometimes become compressed or stretched depending on the orientation of the projection. Moreover, such projections distort distances as well as angular relationships.

Equal Area

This projection type is focused on preserving area at the cost of everything else. This means shape can often be distorted. Equal area is commonly used for thematic mapping. Their advantage is in most accurately representing distributions.

Azimuthal Equidistant projection (planar, orthographic). The focal point is Taipei in Taiwan. Projection is most accurate near to the focal point but becomes distorted as it circles out to the greater world.

- Great Circles:
    - **Red circle:** 10,000 km distance
    - **Purple circle**: 15,000 km distance
- Distorted Countries:
    - **Yellow** – Brazil
    - **Pink** – Argentina
    - **Gray** – Uruguay
    - **Orange** – Paraguay
    - **Green** – Chile
    - **Violet** – Bolivia

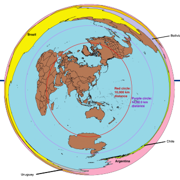


Illustration of Azimuthal Equidistant projection

The Albers (conic) projection is a common equal area projection using two standard parallels which ensures area is preserved between them. This projection is common in environmental and governmental maps.

Equidistant

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

- Azimuthal Equidistant is a common projection for mapping small areas.

Choosing the right projection

As noted earlier, there are theoretically an infinite number of map projections to choose from. One of the key considerations behind the choice of map projection are to reduce the amount of distortion. The geographical object being mapped and the respective scale at which the map will be constructed are also important factors to think about. For instance:

- Maps of the North and South Poles usually use planar or azimuthal projections, and conical projections are best suited for the middle latitude areas of the earth.
- Features that stretch east–west, such as the country of

Russia, are represented well with the standard cylindrical projection

- Countries which are oriented north–south (e.g., Chile, Norway) are better represented using a transverse projection.

If a map projection is unknown, sometimes it can be identified by working backward and examining closely the nature and orientation of the graticule (i.e., grid of latitude and longitude), as well as the varying degrees of distortion. Clearly, there are trade-offs made regarding distortion on every map.

- There are no hard-and-fast rules as to which distortions are more preferred over others.
- The selection of map projection largely depends on the purpose of the map.

Within the scope of GISs, knowing and understanding map projections is critical. For instance, to perform an overlay analysis, all map layers need to be in the same projection. If they are not, geographical features will not be aligned properly, and any analyses performed will be inaccurate and incorrect. If you want to conduct a measurement of land parcel size, you need to use a projection that does not distort area space. Most GISs include functions to assist in the identification of map projections, as well as to transform between projections to synchronize geospatial data. Despite the capabilities of technology, an awareness of the potential and pitfalls that surround map projections is essential.

On the Fly Projection

In a GIS the first dataset added to a project sets the coordinate system; subsequent layers are transformed to display in the projection of the first dataset. It should be noted that this is not a true reprojection, and any changes made to the dataset, or any comparisons made to other layers when it's displayed in an on-the-fly projection are likely to create errors. It's recommended that data be projected properly before editing or analyzing.

Although projecting a dataset was once incredibly difficult, today, we can project and reproject massive quantities of coordinates, transforming them backward and forward from Latitude and Longitude to overlay precisely with data that are stored in some other coordinate space. Automatic transformation of coordinate systems requires that datasets include machine-readable metadata. In about 2002, the makers of ArcMap made projection on the fly possible when it added one a new file to the schema of a shapefile; the .prj file, which contains the description of the projection of a shapefile, and if it exists, it is always copied with the shape file or elements that are exported from it. This is the machine-readable metadata that allowed ArcMap to know how to handle the dataset if any transformation (reprojection) is required;

including projection metadata is now standardly stored in the data. There are plenty of datasets that do not include such machine-readable metadata so it's important to know how to identify it when needed. Most GIS have the ability to project on-the-fly today.

Commonly used Projections

The Universal Transverse Mercator grid is commonly referred to as UTM and is based on the Transverse Mercator projection. Universal Transverse Mercator (UTM) is a coordinate system that largely covers the globe. The system reaches from 84 degrees north to 84 degrees south latitude, and it divides the Earth into 60 north-south oriented zones that are 6 degrees of longitude wide. The UTM system is not a single map projection.

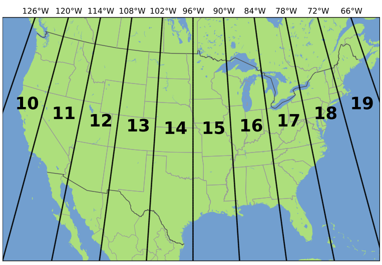


Illustration of UTM zones

- ***Transformation:*** The system instead 60 projections, and each uses a secant (a trigonometric function that for an acute angle is the ratio of the hypotenuse of a right triangle of which the angle is considered part and the leg adjacent to the angle) transverse Mercator projection in each zone.

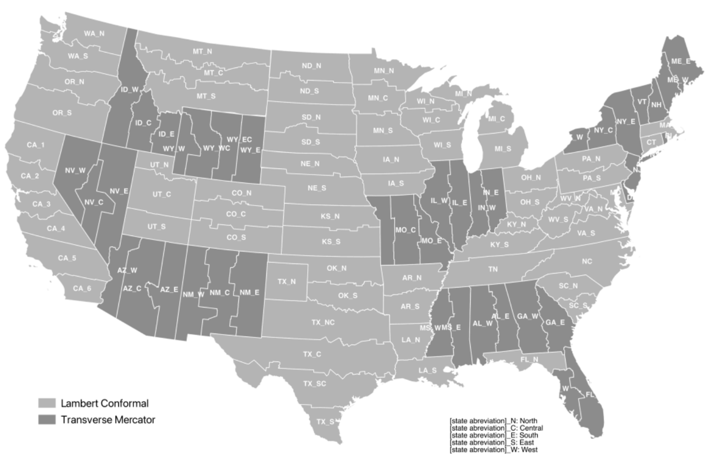
- ***Zones:*** The contiguous U.S. consists of 10 zones**:** UTM zones are numbered consecutively beginning with Zone 1. Zone 1 covers 180 degrees west longitude to 174 degrees west longitude (6 degrees of longitude) and includes the westernmost point of Alaska. Maine falls within Zone 16 because it lies between 84 degrees west and 90 degrees west. In each zone, coordinates are measured as northings and eastings in meters.
- ***Central Meridian:*** The central meridian in each zone is assigned an easting value of 500,000 meters. In Zone 16, the central meridian is 87 degrees west. One meter east of that central meridian is 500,001 meters easting.
- ***Northing/False Northing:*** In the Northern hemisphere, the equator is the zero baseline for Northings (Southern hemisphere uses a 10,000 km false Northing).
- ***Easting/False Easting:*** Each zone has an arbitrary central meridian of 500 km west of each zone's central meridian (called a false Easting) to insure positive Easting values and a central bisecting meridian.

State Plane Coordinate System

The State Plane Coordinate System is a series of separate systems, each covering a state, or a part of a state, and is only used in the United States. It is popular with some state and local governments due to its high accuracy, achieved using relatively small zones. State Plane began in 1933 with the North Carolina Coordinate System and in less than a year it had been copied in all the other states.

Accuracy: The system is designed to have a maximum linear error of 1 in 10,000 and is four times as accurate as the UTM system.

Zones: Like the UTM system, the State Plane system is based on zones. However, the 120 State Plane zones generally follow county boundaries (except in Alaska).



Modeled on map in Fang, 2014. Data sourced from "USA State Plane Zones NAD83." n.d. Accessed October 6, 2025. https://hub.arcgis.com/datasets/esri::usa-state-plane-zones-nad83/about.

State plane divisions, adapted from Fang using Esri data.

State Divisions: Given the State Plane system's desired level of accuracy, larger states are divided into multiple zones, such as the "Colorado North Zone."***Transverse Mercator:*** States with a long north-south axis (such as Idaho and Illinois) are mapped using a Transverse Mercator projection,

Lambert Conformal: States with a long east-west axis (such as Washington and Pennsylvania) are mapped using a Lambert Conformal projection.[2]

Central meridian: In either case, the projection's central meridian is generally run down the approximate center of the zone.

Positive Coordinates: A Cartesian coordinate system is created for each zone by establishing an origin at some distance (usually 2,000,000 feet) to the west of the zone's central meridian and some distance to the south of the zone's southernmost point. This ensures that all coordinates within the zone will be positive. The X-axis running through this origin runs east-west, and the Y-axis runs north-south.

Measurement: Distances from the origin are generally measured in feet but sometimes are in meters.

Eastings/Northings**:** X distances are typically called eastings (because they measure distances east of the origin) and Y distances are typically called northings (because they measure distances north of the origin).

2. https://gis.depaul.edu/shwang/teaching/geog258/Grid_files/image002.jpg

Exercises

Exercise 3a. Convert DMS to DD

At times we come upon datasets in which the latitude and longitude are recorded in degrees, minutes, seconds (DMS), but for GIS having coordinates in decimal degrees is preferred. Data this dataset and convert the DMS coordinates to DD.

Formula: Decimal degrees = Degrees + (Minutes/60) + (Seconds/3600)

Dataset: download geothink_sdms.csv

Exercises

Exercise 3b. Find Your Local Projected Coordinate System

Find the projected coordinate system most appropriate for your local area.

CHAPTER 4: THE NATURE OF GEOSPATIAL DATA

Adapted from Fang

How do we reduce the massive complexity of the Earth and its inhabitants so we can portray them in a GIS database and on a map? We do it by selecting the most relevant features (ignoring those we do not think are necessary for our specific research or project) and then generalizing the features we have selected. This is a geospatial data model, again there are two main types of geospatial data models, vector and raster.

Vector Data Models

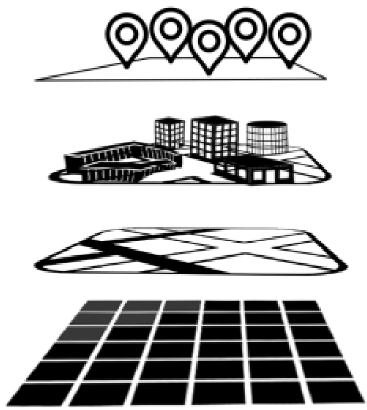


Illustration of points, lines, and polygon layers

Vector data models use points and their associated [X and Y] coordinate pairs to represent the vertices of spatial features.The features are the geometries we can see and the attributes (descriptive elements) of these features are stored in a separate database management system. The geospatial information and the attribute information for these models are linked via a simple identification number that is given to each feature on a map. There are three fundamental vector geometries in GIS:

1. points,
2. lines, and
3. polygons.

Geometries

Points

Points are zero-dimensional objects that contain only a single coordinate pair. Points are typically used to model singular, discrete features such as buildings, wells, power poles, sample locations, and so forth. Points have only the property of location. Other types of point features include the node and the vertex.

Specifically, a point is a stand-alone feature, while a node is a topological junction representing a common X, Y coordinate pair between intersecting lines (edges).

Vertices are defined as each bend along a line or polygon feature.

Lines

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

Polygons

Polygons are two-dimensional features created by multiple lines (edges) that loop back to create a "closed" feature. In the case of polygons, the first coordinate pair (point) on the first line segment is the same as the last coordinate pair on the last line segment. Polygons are used to represent features such as city boundaries, geologic formations, lakes, soil associations, vegetation communities, and so forth. They can also be used to represent buildings and parcels. Polygons have the properties of area and perimeter. Polygons are also called areas.

Vector Data Model Structures

Vector data models can be structured many ways. We will examine two of the more common data structures here.

Spaghetti Model (sans relationships)

The simplest vector data structure is called the spaghetti data model (Dangermond 1982). In the spaghetti model, each point, line, and/or polygon feature is represented as a string of X, Y coordinate pairs (or as a single X, Y coordinate pair in the case of a vector image with a single point) with no inherent rules. One could envision each line in this model to be a single strand of spaghetti that is formed into complex shapes by the addition of more and more strands of spaghetti.

It is notable that in this model, any polygons that lie adjacent to each other must be made up of their own lines or stands of spaghetti. In other words, each polygon must be uniquely defined by its own set of X,Y coordinate pairs, even if the adjacent polygons share the exact same boundary information. This creates some redundancies within the data model and therefore reduces efficiency.

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

Topological (relationships)

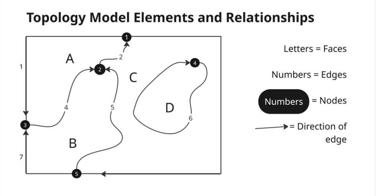


Illustration of a topological model

Topology in short is the definition of the relationships be between coincident geometry. In contrast to the spaghetti data model, the topological data model is characterized by the inclusion of topological information within the dataset, as the name implies. Topology is a set of rules that model the relationships between neighboring points, lines, and polygons and determines how they share geometry. Topology allows the computer to rapidly determine and analyze the spatial relationships of all its included features. In addition, topological information is important because it allows for efficient error detection within a vector dataset. For example, consider two adjacent polygons. In the spaghetti model, the shared boundary of two neighboring polygons is defined as two separate, identical lines. The inclusion of topology into the data model allows for a single line to represent this shared boundary with an explicit reference to denote which side of the line belongs with which polygon. Topology is also concerned with preserving spatial properties when the forms are bent, stretched, or placed under similar geometric transformations, which allows for more efficient projection and reprojection of map.

Three basic topological precepts that are necessary to understand the topological data model are outlined here.

First Topological Precepts

In the topological data model***, nodes are the intersection points*** where two or more edges (lines) meet. In the case of arc-node topology, arcs have both a from-node (i.e., starting node) indicating where the edge begins and a to-node (i.e., ending node) indicating where the edge ends. In addition, between each node pair is a line segment (edge), which has its own identification number and references both its from-node and to-node.

Second Topological Precept

The second basic topological precept is ***area definition***. Area definition states that an edge that connects to surround an area defines a polygon. Edges are used to construct polygons, and each edge is stored only once. This results in a reduction in the amount of data stored and ensures that adjacent polygon boundaries do not overlap.

Third Topological Precept

Contiguity, the third topological precept, is based on the concept that polygons that share a boundary are deemed adjacent. Specifically, polygon topology requires that all edges in a polygon have a direction (a from-node and a to-node), which allows adjacency information to be determined. Polygons that share an edge are deemed adjacent, or contiguous, and therefore the "left" and "right" side of each arc can be defined. This left and right polygon information is stored explicitly within the attribute information of the topological data model. The "universe polygon" is an essential component of polygon topology that represents the external area located outside of the study area.

Advantages of the Vector Model

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

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

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

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

Disadvantages of the Vector Model

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

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

Clipping of a Seurat painting demonstrating pointillism

The neoimpressionist artist, Georges Seurat, developed a painting technique referred to as "pointillism" Notably, the foundation of this technology predates computers and digital cameras by nearly a century. In the 1880s, which similarly relies on the amassing of small, monochromatic "dots" of ink that combine to form a larger image.

Raster Data Model

The raster data model is widely used in applications ranging far beyond geographic information systems (GISs). Most likely, you are already very familiar with this data model if you have any experience with digital photographs. The ubiquitous JPEG, BMP, and TIFF file formats (among others) are based on the raster data model. Take a moment to view your favorite digital image. If you zoom deeply into the image, you will notice that it is composed of an array of tiny square pixels (or picture elements). Each of these uniquely colored pixels, when viewed as a whole, combines to form a coherent image.

The raster data model consists of rows and columns of equally sized cells (pixels) interconnected to form a planar surface. These pixels are used as building blocks for creating points, lines, areas, networks, and surfaces. The contrast between raster and vector models reflect the 'pixelization' of a raster, which would be points, lines and polygons in a vector data model. These squares are typically reformed into rectangles of various dimensions if the data model is transformed from one projection to another (e.g., from State Plane coordinates to UTM (Universal Transverse Mercator) coordinates).

Because of the reliance on a uniform series of square pixels, the raster data model is referred to as a grid-based system. Each cell in a raster grid carries a single value, which represents the characteristic of the spatial phenomenon at a location denoted by its row and column. The data type for that cell value can be either integer or floating-point.

Resolution

The area covered by each pixel determines the spatial resolution of the raster model from which it is derived. The more area covered per pixel, the less accurate the associated data values. Specifically, resolution is determined by measuring one side of the square pixel.

10m resolution: A raster model with pixels representing

10m-by-10m (or 100 square meters) in the real world would be said to have a spatial resolution of 10m;

1km resolution: a raster model with pixels measuring 1km-by-1km (1 square kilometer) in the real world would be said to have a spatial resolution of 1km; and so forth.

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

Raster Requirements

Imagery employing the raster data model must exhibit several properties:

- Each pixel must hold at least one value, even if that data value is zero
- If no data are present for a given pixel, a data value

placeholder must be assigned to this grid cell.

- Often, an arbitrary, readily identifiable value (e.g., −9999) will be assigned to pixels for which there is no data value.
- A cell can hold any alphanumeric index that represents an attribute.
- In quantitative datasets, attribute assignation is fairly straight-forward. For example, if a raster image denotes elevation, the data values for each pixel would be some indication of elevation, usually in feet or meters.
- In qualitative datasets, data values are indices that necessarily refer to some predetermined translational rule. In the case of a land-use/land-cover raster graphic, the following rule may be applied:
    - 1 = grassland
    - 2 = agricultural
    - 3 = disturbed
    - so forth….
- Points and lines "move" to the center of the cell. As one might expect, if a 1 km resolution raster image contains a river or stream, the location of the actual waterway within the "river" pixel will be unclear.
- Therefore, there is a general assumption that all zero-dimensional (point) and one-dimensional (line) features will be located toward the center of the cell.
- As a corollary, the minimum width for any line feature must necessarily be one cell regardless of the actual

width of the feature.

- If it is not, the feature will not be represented in the image and will therefore be assumed to be absent.

Illustration of cell-by-cell raster encoding

Three (of many) models for encoding raster data from scratch:

Cell-by-cell raster encoding. This minimally intensive method encodes a raster by creating records for each cell value by row and column (This method could be thought of as a large spreadsheet wherein each cell of the spreadsheet represents a pixel in the raster image. This method is also referred to as "***exhaustive enumeration.***"

Run-length raster encoding. This method encodes cell values in runs of similarly valued pixels and can result in a highly compressed image file The run-length encoding method is useful in situations where large groups of neighboring pixels have similar values (e.g., discrete datasets such as land use/land cover or habitat suitability) and is less useful where neighboring pixel values vary widely (e.g., continuous datasets such as elevation or sea-surface temperatures).

Quad-tree raster encoding. This method divides a raster into a hierarchy of quadrants that are subdivided based on similarly valued pixels. The division of the raster stops when a quadrant is made entirely from cells of the same value. A quadrant that cannot be subdivided is called a "leaf node."

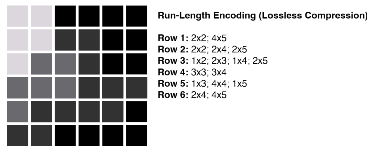
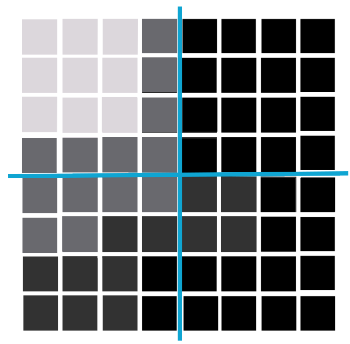


Illustration of run-length raster encoding



An illustration of a quad-tree raster.

Advantages of the Raster Model

Ubiquitous: Technology required to create raster graphics is inexpensive and ubiquitous. Nearly everyone currently owns some sort of raster image generator, namely a digital camera, and few cellular phones are sold today that don't include such functionality.

Satellite Data: A plethora of satellites are constantly beaming up-to-the-minute raster graphics to scientific facilities across the. These graphics are often posted online for private and/or public use, occasionally at no cost to the user.

Simplicity: Raster graphics are the relative simplicity of the underlying data structure. Each grid location represented in the raster image correlates to a single value (or series of values if attributes tables are included). This simple data structure may also help explain why it is relatively easy to perform overlay analyses on raster data.

Interpretation: The simplicity also lends itself to easy interpretation and maintenance of the graphics, relative to its vector counterpart.

Disadvantage of the Raster Model

Storage: Raster files are typically very large. Particularly in the case of raster images built from the cell-by-cell encoding methodology, the sheer number of values stored for a given dataset result in potentially enormous files. Any raster file that covers a large area and has somewhat finely resolved pixels will quickly reach hundreds of megabytes in size or more. These large files are only getting larger as the quantity and quality of raster datasets continues to keep pace with quantity and quality of computer resources and raster data collectors (e.g., digital cameras, satellites).

Scaling: The output images are less "pretty" than their vector counterparts. This is particularly noticeable when the raster images are enlarged. Depending on how far one zooms into a raster image, the details and coherence of that image will quickly be lost amid a pixelated sea of seemingly randomly colored grid.

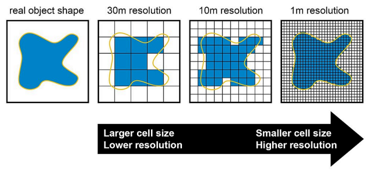


Illustration of a raster satellite image at different resolutions

Reprojection: Geometric transformations that arise during map reprojection efforts can cause problems for raster graphics and represent a third disadvantage to using the raster data model. Changing map projections will alter the size and shape of the original input layer and frequently result in the loss or addition of pixels (White 2006). These alterations will result in the perfect square pixels of the input layer taking on some alternate rhomboidal dimensions. However, the problem is larger than a simple reformation of the square pixel. Indeed, the reprojection of a raster image dataset from one projection to another brings change to pixel values that may, in turn, significantly alter the output information (Seong 2003).

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

Tabular Data

A database is a structured collection of data files. A database management system (DBMS) is a software package that allows for the creation, storage, maintenance, manipulation, and retrieval of large datasets that are distributed over one or more files. Database management normally refers to the management of tabular data in row and column format. Geospatial database management systems, alternatively, include the functionality of a DBMS but also contain specific geographic information about each data point such as identity, location, shape, and orientation.

Integrating this geographic information with the tabular attribute data of a classical DBMS provide users with powerful tools to visualize and answer the spatially explicit questions that arise in an increasingly technological society. Several types of database models exist, such as the flat, hierarchical, network, and relational models.

Hierarchical:

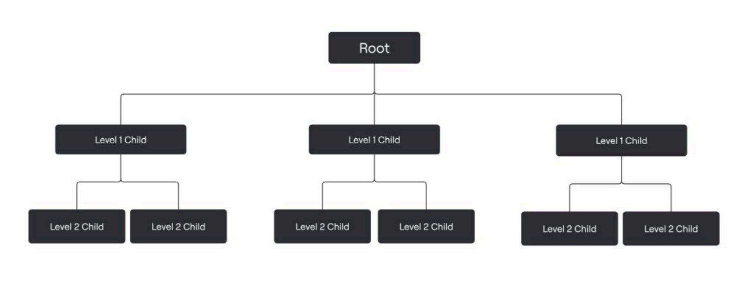


Illustration of a hierarchical database schematic

A hierarchical database is also a fairly simple model that organizes data into a "one-to-many" association across levels. Common examples of this model include phylogenetic trees for classification of plants and animals and familial genealogical trees showing parent-child relationships.

Network:

Network databases are similar to hierarchical databases, however, because they also support "many-to-many" relationships. This expanded capability allows greater search flexibility within the dataset and reduces potential redundancy of information. Alternatively, both the hierarchical and network models can become incredibly complex depending on the size of the databases and the number of interactions between the data points.

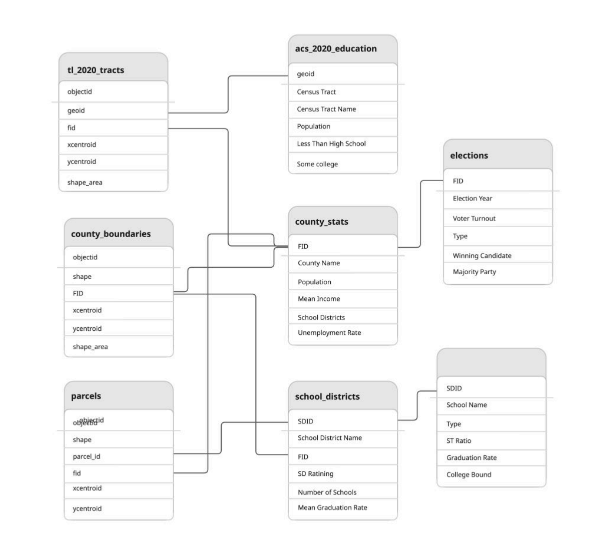


Illustration of a network database schematic

Relational Database Management Systems:

GIS software typically employs a relational database (Codd 1970). A relational database management system (RDBMS) is a collection of tables that are connected in such a way that that data can be accessed without reorganization of the tables. The tables are created such that each

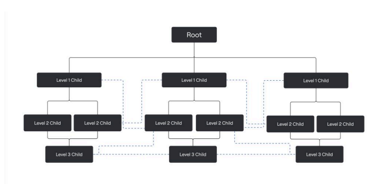


Illustration of a relational database schematic

- column represents a particular variable which describes

an object (attribute) (e.g., soil type, PIN number, last name, area)
- row contains a unique observation/instance of data for that column attribute (e.g., Delhi Sands Soils, 5555, Smith, 412.3 acres)
- each observation has a unique identifier

In the relational model, each table (not surprisingly called a relation) is linked to each other table via predetermined keys (Date 1995).

- The primary key represents the attribute (column) whose value uniquely identifies a particular record (row) in the relation (table).
- The primary key may not contain missing values as multiple missing values would represent nonunique entities that violate the basic rule of the primary key.
- The primary key corresponds to an identical attribute in a secondary table (and possibly third, fourth, fifth, etc.) called a foreign key. This results in all the information in the first table being directly related to the information in the second table via the primary and foreign keys, hence the term "relational" DBMS.

With these links in place, tables within the database can be kept very simple, resulting in minimal computation time and file complexity. This process can be repeated over many tables as long as each contains a foreign key that corresponds to another table's primary key.

The relational model has two primary advantages over the other database models described earlier.

1. Each table can now be separately prepared, maintained, and edited. This is particularly useful when one considers the potentially huge size of many of today's modern databases.
2. Tables may be maintained separately until the need for a particular query or analysis calls for the tables to be related. This creates a large degree of efficiency for processing of information within a given database.

It may become apparent to the reader that there is great potential for redundancy in this model as each table must contain an attribute that corresponds to an attribute in every other related table. Therefore, redundancy must actively be monitored and managed in a RDBMS. To accomplish this, a set of rules called *normal forms* have been developed (Codd 1970). [5] There are three basic normal forms.

- ***First Normal Form Violation*** refers to five conditions that must be met

    1. There is no sequence to the ordering of the rows.
    2. There is no sequence to the ordering of the

columns.

3. Each row is unique.
4. Every cell contains one and only one value.
5. All values in a column pertain to the same subject.

- ***The second normal form*** states that any column that is not a primary key must be dependent on the primary key. This reduces redundancy by eliminating the potential for multiple primary keys throughout multiple tables. This step often involves the creation of new tables to maintain normalization.
- ***The third normal form*** states that all nonprimary keys must depend on the primary key, while the primary key remains independent of all nonprimary keys. This form was wittily summed up by Kent (1983) [7] who quipped that all nonprimary keys "must provide a fact about the key, the whole key, and nothing but the key." Echoing this quote is the rejoinder: "so help me Codd" (personal communication with Foresman 1989).

Joins and Relates

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

Attribute Join

The join operation appends the fields of one table into a second table using an attribute or field that is common to both tables. This is commonly utilized to combine attribute information from one or more nonspatial data tables (i.e., information taken from reports or documents) with a spatially explicit GIS feature layer.

Requirements for a successful join

- Join fields must contain matching values
- Join fields must be the same data format (e.g., string, integer, etc.)
- Field headers must not contain spaces or special characters
- Field values are clean and standardized
- Field headers should not use reserved words (e.g., date," "day," "month," "table," "text," "user," "when," "where," "year," or "zone")
- Table is in an incompatible format

Spatial Join

A second type of join combines feature information based on spatial location and association rather than on common attributes.

- match each feature to the closest feature
- match each feature to the feature that it is part of
- match each feature to the feature that it intersects.

Relate

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

When to join or relate:

- Sometimes it can be unclear as to which operation one should use. Some considerations are:
- joins are most suitable for instances involving one-to-one or many-to-one relationships;
- joins are also advantageous because the data from the two tables are readily observable in the single output table;
- relates, on the other hand, are suitable for all table relationships (one-to-one, one-to-many, many-to-one, and many-to-many);
- relates can slow down computer access time if the tables are particularly large or spread out over remote locations.

Exercises

Exercise 4: Find the Errors in the Table

Review these tables with the ideas that you may like to join them in GIS. Identify 5 reasons these tables will NOT join.

objectid	geoid	name
1	29189215301	Census Tract 2153.01
2	29189215302	Census Tract 2153.02
3	29189215400	Census Tract 2154
4	29189215500	Census Tract 2155
5	29189215600	Census Tract 2156
6	29189215700	Census Tract 2157

geoid	Name of Area
1500000US295101011001	Block Group 1, Census Tract 1011
1500000US295101011002	Block Group 2, Census Tract 1011
1500000US295101012001	Block Group 1, Census Tract 1012
1500000US295101012002	Block Group 2, Census Tract 1012
1500000US295101012003	Block Group 3, Census Tract 1012
1500000US295101013001	Block Group 1, Census Tract 1013

CHAPTER 5: GEOSPATIAL DATA MANAGEMENT

Adapted from Kristin Briney's Data Management Workbook

Why do Data Management?

Most of us have spent time at some point digging through our computer to find a specific file that can't be located. It's incredibly frustrating and a waste of time and resources, especially if you end up recollecting missing data. The good news is that it is possible to avoid this situation entirely by strategically managing your data better.

Done well, data management means:

- always understanding what your data is and how you collected it even if the data is a year old
- always finding the file you need quickly
- never losing your data even if your hard drive crashes
- knowing what rights and responsibilities you have over your data
- knowing how and where to share your data to comply

with your funder's data sharing policy

- being able to pick up and easily reuse data from a past project

Geospatial data types:

Adapted from the DCN specialized curation curriculum

Geospatial data can be messy and therefore good data management is necessary to keep data functioning in additions to the reasons to manage data above.

This is a quick guide to common file types and the software they can be access in (not exhaustive).

Extension	***Description***	***QGI***
.shp .dbf .shx .sbx .sbn .prj .xml .cpg .qmd	shapefile	Y
.json .geojson	object files suitable for websites	Y
.kml .kmz	keyhole markup (Google)	Y
.csv	comma separated values	Y
.gdb	Geodatabase	Y

Shapefile

One of the most common vector File formats out there is the Esri Shapefile. The "shapefile" is, in fact, a set of related files with the same filename and different extensions. A shapefile can have anywhere between three and nine files.

- The primary file extensions are .SHP, .DBF, .SHX. These three files hold the geometry and location of the data (.SHP), the attributes for each location (.DBF), and an index tying the two together (.SHX).
- These three files must be present for the file to be read correctly in GIS software.
- There are a number of other extensions, however, that may also be present, containing information about things like the projection (.PRJ), metadata (.XML), and encoding (.CPG).

Geodatabase

Another extremely common file form is the Esri File Geodatabase. They are one of the most common kinds of geodatabase and somewhat complex to work with in QGIS.

- First it's important to note that if you look inside a folder with a .GDB extension, it is hard to make sense of the contents. There can be dozens or even hundreds of files, all with software-generated and non-descript filenames.
- These files are not particularly comprehensible to the human eye but can be interpreted when opened through

a GIS software. They should not be manipulated in any way outside of GIS software.
- When opening the example file geodatabase in QGIS, there are three data layers represented by that large number of files. Note: If there is raster data stored in a File Geodatabase it can only be accessed through ArcGIS.

Geojson

GeoJSON is a format for encoding a variety of geographic data structures using JavaScript Object Notation (JSON). A GeoJSON object may represent a region of space (a geometry), a spatially bounded entity (a feature), or a list of features (feature collection). GeoJSON supports:

- Point
- Line string (LineString)
- Polygon
- Multi Point (MultiPoint)
- Multi Line String (MultiLineString)
- Multi Polygon (MultiPolygon)
- Geometry Collection (GeometryCollection)

Geospatial data is defined broadly in this format; anything with qualities that are bounded in geographical space might be a feature regardless of structure. It uses geospatial standards to streamline the data for web applications.

KML

Keyhole Markup Language was developed for usage in Google Maps and Google Earth. Originating as a community standard, this standard defines an XML language focused on geographic visualization, including annotation of maps and images. It is used to encode and transport representations of geographic data for display in an earth browser. Geographic visualization includes not only the presentation of graphical data on the globe, but also the control of the user's navigation in the sense of where to go and where to look. https://www.ogc.org/publications/standard/kml/

Geotiff

One of the most common raster file formats is GeoTIFF, with file extension .TIF or .TIFF. This format is similar to other varieties of .TIF images, but with the important addition that it has information about where the data is located in the world. Sometimes this geospatial information is embedded within the primary file – in the header of the TIFF. Sometimes the information is stored in a separate file known as a "world file" (.TFW). There are a few other extensions that may also be present, storing supplemental content such as a lower resolution version of the data for faster display (.OVR) and metadata (.XML).

CSV

An extremely simple, and one of the most common formats for distributing geospatial information is just a spreadsheet organized by comma separate values (.CSV). If there are columns that contain geospatial information it can be displayed in a GIS. There are three key ways a csv file can be used in GIS:

1. If the table has a Latitude column and a Longitude column, the data can be displayed with a geographic coordinate system, most likely WGS84. After it's been displayed in can be saved and projected into an appropriate projection for the project.
2. If the table has address, zip code, or even place name information, a locator dataset can be used to geocode the information based on the locator reference dataset.
3. If the table has a key location field that is also present in a feature class, the table can be joined to the feature class. An example is a Census Tract table can be joined to a Census Tract Tigerline file based on a geoid.

File Organization System

Good file organization and naming are foundational data management practices, as they help you find files quickly when you need them.

Implementing a file organization system is the first step toward creating order for your research data. Well-organized files make it easier to find the data you need without spending lots of time searching your computer.

Everyone organizes their files slightly differently, but the actual organizational system is less important that having a place where all your files should logically go.

File Naming

File naming conventions are a simple way to add order to your files and help to find them later. Rich and descriptive file names make it easier to search for files, understand at a glance what they contain, and tell related files apart. This exercise guides researchers through the process of creating a file naming convention for a group of related files.

File naming rules:

- no spaces
- no special characters:
- special characters are used by systems for other purposes
- special characters include anything that is not alpha-numeric, except for an underscore, and sometimes a hyphen !@#$%^&*()+=.,~`, etc.
- a period is a special character reserved for parsing the file name from the extension type

- brief (fewer than 32 characters)
- meaningful: don't accept default names or create arbitrary names consistent
- dates should be in ISO 8601, YYYYMMDD
- versions should be alpha numeric (e.g., v1, v2, v3) rather than using words like "final"

Consistent pattern for file naming, for example:

- component 1: project abbreviation
- component 2: type of object
- component 3: responsible team member initials
- component 4: date
- component 5: version
- pattern: prj_type_person_date_v.ext
- example: rhzone_map_20270801_v3.pdf (22 characters)

Storage

Data needs to be stored and backed up, but it can be frustrating to pick these systems and ensure that they are working correctly. Back up means that it's in more than one location.

Local storage considerations:

- How much data will you have?
- Security; does your data need to be in HIPPAA compliant storage?
- Do you have access to a cloud storage platform (box, drive, dropbox, etc.)
- Who has access to the storage?
- Does it have automatic backups?

Documentation

Documentation has sometimes been called "a love letter to your future self" as it helps you remember important details about your research data. For project-level documentation we need a README.txt file and a data dictionary to ensure your documentation is of sufficient quality.

Readme file:

Data files living on a computer often need extra documentation for someone to understand what research they correspond to. In particular, it is useful to record the most basic project information and store it in the top-level folder of each research project. This can be done with a README.txt. The name, "README," indicates that the file conveys important information and the file type, TXT, can be opened by many different software programs, making the content maximally accessible.

What to include in the readme file:

- brief description
- time period of collection
- contributors/project team members
- where are the data stored
- how and where are they documented
- file organization and file descriptions
- what other information will help you or someone else understand the project

Data Dictionary:

Ideally, a spreadsheet is formatted with a row of variable names at the top, followed by rows of data going down. This makes it easier for data to be used in any data analysis software (interoperability is a good thing) but makes it impossible to document a spreadsheet within the file itself. For this reason, it's useful to create a data dictionary to describe the spreadsheet so that others can interpret the data. This exercise walks you through the major information you should record for each variable in the spreadsheet, adding up to a complete dictionary to accompany the spreadsheet file.

Question	Example
Variable name	site
Variable description	Two-letter abbreviation describing the name of the overall sit
Variable units	N/A
Relationship to other variables	Partner to variable "sampleNum," which together define the s that site). Related to variables "latitude" and "longitude," wh more specific than the larger site code.
Variable coding values and meanings	Coding values and meanings: BL = Badlands NP; DV = Deat Joshua Tree NP; ZN = Zion NP
Known issues with the data	Some Badlands samples were collected outside of the park bo for specific locations.
Anything else to know about the data?	Older data (pre-2013) used one-letter abbreviations for site co identification.

Writing a Living Data Management Plan (DMP)

A DMP describes how data will be actively managed during a project and may be updated whenever necessary to reflect current data practices. A living DMP is a useful touchstone for understanding where data lives, how it's labelled, how it moves through the research process, and who will oversee the data management.

What's in a Living DMP?

- **Short summary of the project**: Provide a short section that provides contextual information, often this includes:
    - Description of the problem
    - The audience and your relationship to them
    - Start-End Date
- **What is the collection-storage-wrangling-analysis pipeline?**
    - Source of data collection
    - Active storage location details
    - Active storage location organization plan
- **File naming convention?**
    - for raw files
    - for clean files
    - for analysis files
    - for deliverables
- **Documentation plan for active data and analysis methods**
- **Archival storage location details**

- **Project roles and responsibilities**
    - E.g. data manager: provides access to team members; ensures DMP is followed; provides templates for collection and documentation and reviews documentation
    - E.g data collector: using guidelines and templates created by data manager to collect and document data.
- **Any other relevant details to support the project**

Exercises

Exercise 5: Create a directory structure and file naming convention

Start with a directory you use often. Leverage the best practices above to create a logical direction structure and a filed naming convention.

Directory structure:

- Level 1:
- Level 2:
- Level 3:

File Naming convention

- Component 1
- Component 2
- Component 3
- Component 4
- Component 5

Example:

- Component 1: project abbreviation
- Component 2: type of thing
- Component 3: coverage data
- Component 4: modified by
- Component 5: modified date
- gisbc_map1_2020_jm_202302017

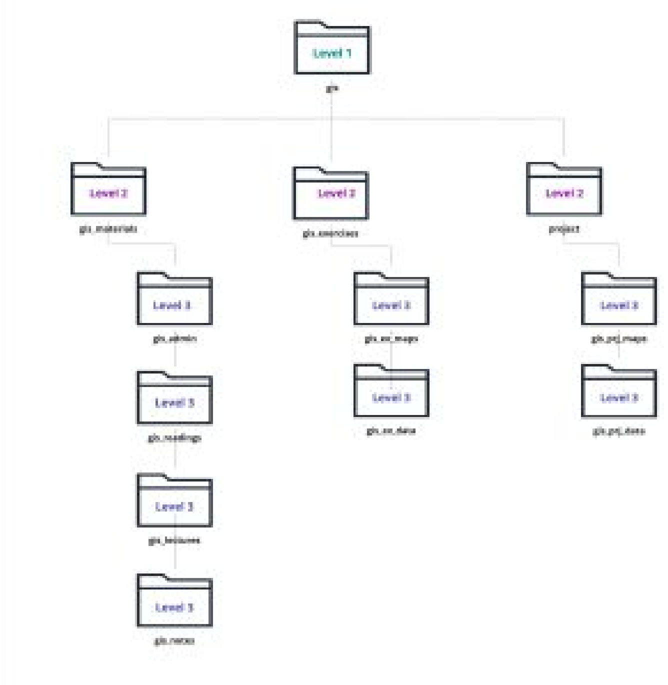


Illustration of a directory structure schematic

CHAPTER 6: PROJECT PLANNING

This project lifecycle is a useful framework for thinking about the stages of a project. The idea stage includes the development of the projects purpose. It might mean exploring existing literature and reusing data from your own work or the work of others, it may mean brainstorming with collaborators.

Creating a Project Plan

When starting a GIS project, it helps to begin by outlining your project, although you should also create a DMP [see Chapter 5], the project plan is at a higher level.

- ***Problem:*** Use geospatial thinking to identify and articulate the problem you hope to address. This often starts with an observation that you would like to explore. The problem descriptions include the intended study area, your assumptions, and your goals.
- ***Context:*** The context includes defining your audiences. Who cares about this problem and what is your relationship to them? How do you anticipate impacting

this audience and how do you imagine you could best present the results to them?

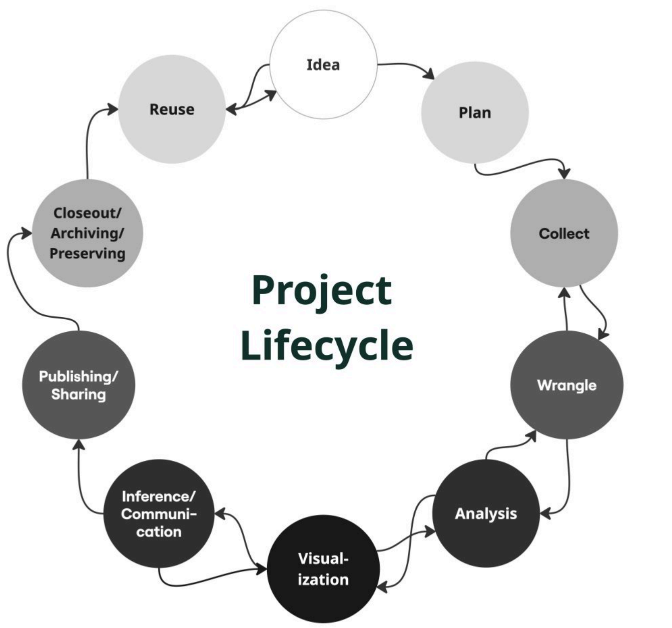


Illustration of a project lifecycle for data

- ***Data:*** When creating your project plan, you don't necessarily need to know exactly what data you'll use. However, it would be helpful to outline the expected extent of your study area, if you will collect primary data (data that you gather from observation), what secondary data (data collected by someone else) will you use where will you look for it, and how will you evaluate it to make sure it's reusable.

- FAIR is a helpful framework for evaluating secondary data.
  - Findable: it came from an authoritative data source, which you can cite
  - Accessible: the repository uses open protocols
  - Interoperable: the data and metadata are in open formats
  - Reusable: the data are well documented so that you understand details about the data and variables. This includes any licensing restrictions. As mentioned in the data management chapter, a README file and data dictionary is important when reusing data.
- Methods: This part of the plan describes what you want to explore at a high level. You may not know at this stage all the methods you'll need, but you can map out major steps you know you will need to make: find data, wrangle data, geocode, join, etc.
- Communicate: It's essential to think back to your audience and what sort of communication will be most impactful. For example, it could be through a series of printed maps or it could be an interactive online map.
- Challenges: You might know going into a project that there are issues you will have to address in the project process. Some examples might be that you

need to learn more about some aspect of the project, or that you aren't sure if you will be able to find the data you need, and so your plan B might be proxy data or a different line of discovery.

Creating a diagram or a story board for your project is a great way to work through the project plan.

Part 1, Exercise 6: Project Scenario

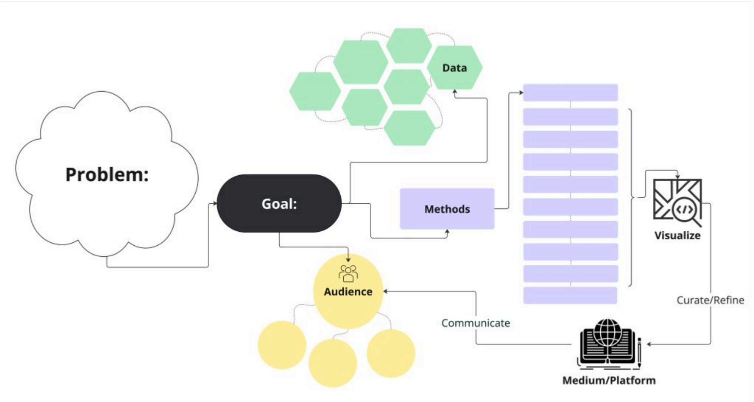


Illustration of a project diagram

Exercises

Exercise 6: Map out project problem you would like to plan on your own.

Brainstorm a geospatial thinking problem. Fill as a much of the of the plan as you are able. You can come back to this exercise as we work through Part 2.

You can use this diagram if it helps you.

CHAPTER 7: VISUALLY REPRESENTING GEOSPATIAL THINKING

The Lay of The Land or Mapping Basics: Lying with Maps

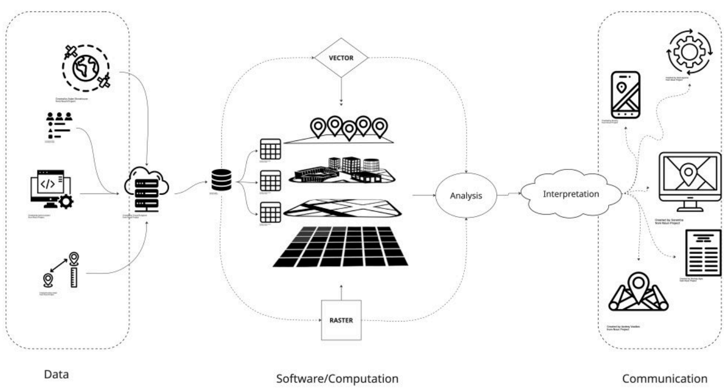


Illustration of a GIS workflow

Successfully communicating the results of geospatial thinking depends on how data are presented. This chapter will provide an overview of communicating geospatial results. As a reminder, the communication may or may not be a map.

Insights from geospatial thinking could also come in the form of a research article, a workflow, or a dashboard.

Whatever the medium, geospatial results can wield a powerful authority. Thematic mapping is an excellent example of how we can sway audiences with our communication choices. Mark Monmonier warns us in "How to Lie with Maps" not only to be careful in designing our own maps, but also to critically examine the maps of others with healthy skepticism. Maps are not alone; compelling data visualization can also tell powerful stories, or at least a side of it. For those who rely on data-driven decision making, it is especially important to critically evaluate what's presented.

GIS utilizes many concepts and themes from cartography, the formal study of maps and mapping. To become proficient with GIS, it helps to understand cartography. The first part of this chapter defines what a map is and describes a few key map types. Next, cartographic or mapping conventions are discussed with particular emphasis placed upon map scale, coordinate systems, and map projections. The chapter concludes with a discussion of the process of map abstraction as it relates to GISs. This chapter provides the foundations for working with, integrating, and making maps with GIS.

[1]

Maps are among the most compelling forms of information

1. Adapted from Campbell 2.1 Maps and Map Types

for several reasons. Maps are artistic. Maps are scientific. Maps preserve history. Maps clarify. Maps reveal the invisible. Maps inform the future. Regardless of the reason, maps capture the imagination of people around the world. As one of the most trusted forms of information, map makers and geographic information system (GIS) practitioners hold a considerable amount of power and influence (Wood 1992; Monmonier 1996).

So, what exactly is a map? As stated before, maps are tools of representation, notwithstanding the actual presentation medium of the map (e.g., our fleeting thoughts, paper, or digital display), maps communicate aspects of the world. For purposes of clarity, the three high-level map categories are reference (general purpose) maps, thematic maps, and the dynamic maps.

Reference Maps

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

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

Thematic Maps

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

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

When presented in hard-copy format, both reference and thematic maps are static or fixed representations of reality. Such permanence on the page suggests that geography and the things that we map are also in many ways fixed or constant. This is far from reality. Leveraging dynamic, interactive mapping elevates the information and provides the viewer an opportunity to engage with the data.

Dynamic Maps

Dynamic maps are changeable or interactive representations of place. It refers more to how maps are designed, delivered, and used than to the content of the map itself. Both reference and thematic maps can be dynamic in nature, and such maps are an integral component to any GIS. Dynamic maps encourage and sometimes require user interaction. Such interaction can include changing the scale or visible area by zooming in or zooming out, selecting which features or layers to include or to remove from a map (e.g., roads, imagery), or starting/stopping a map animation. The use of dynamic mapping has become ubiquitous and map users continue to demand more interactive map features and controls. As democratization of maps and mapping continues, the geographic awareness and map appreciation of map users will also increase. Therefore, it is of critical importance to understand the nature, form, and content of maps to support the changing needs, demands, and expectations of map users in the future.

Basics of Symbology

2

Color is an extremely powerful communication tool, but it has to be used appropriately, Color will greatly enhance and support map design, but if used carelessly, it can also detract from a mapping product. To use color properly, one must first understand color basics and consider the purpose of the map. In some cases, the use of color is not warranted. Grayscale maps can be just as effective as color.

Primary aspects

Color is a tricky subject in mapping and data visualization in general. Colors vary in hue as well as lightness. Monmonier has written in depth about color choice in How to Lie with Maps, and Dorris Scott has a helpful textbook on Digital Cartography [link to this: . For the purpose of this text, we will cover some basic rules.

2. Color Basics (adapted from Campbell)

The three primary aspects of color that must be addressed in cartography are hue, value, and saturation.

Hue is the dominant wavelength or color associated with a reflecting object.

Hue is the most basic component of color and includes red, blue, yellow, purple, and so forth. Value is the amount of white or black in the color.

Value is often synonymous with contrast.

Variations in the amount of value for a given hue result in varying degrees of lightness or darkness for that color. Lighter colors are said to possess high value, while dark colors possess low value. Monochrome colors are groups of colors with the same hue but with incremental variations in value. As seen in, variations in value will typically lead the viewer's eye from dark areas to light areas.

Saturation describes the intensity of color.

Full saturation results in pure colors, while low saturation colors approach gray. Variations in saturation yield different shades and tints. ***Shades*** are produced by blocking light, such as by an umbrella, tree, curtain, and so forth. Increasing the amount of shading results in grays and blacks. ***Tint*** is the opposite of shade and is produced by adding white to a color.

Tints and shades are particularly germane when using additive color models (see for more on additive color models). To maximize the interpretability of a map, use saturated colors to represent hierarchically prominent features and washed-out colors to represent background features.

Meaning

Color is particularly suited to convey meaning. For example, red is a strong color that evokes a passionate response in humans. Red has been shown to evoke physiological responses such as increasing the rate of respiration and raising blood pressure. Red is frequently associated with blood, war, violence, even love. On the other hand, blue is a color associated with calming effects. Associated with the sky or ocean, blue colors can actually assist in sleep and is therefore a recommended color for bedrooms. Too much blue, however, can result in a lapse from calming effects into feelings of depression (i.e., having the "blues").

Green is most commonly associated with life or nature (plants). The color green is certainly one of the most topical colors in today's society with commonplace references to green construction, the Green party, going green, and so forth. Green, however, can also represent envy and inexperience (e.g., the green-eyed monster, greenhorn). Brown is also a nature color but more as a representation of earth and stone. Brown can also imply dullness. Yellow is most commonly associated with sunshine and warmth, somewhat similar to red. Yellow can also represent cowardice (e.g., yellow-bellied).

Black, the absence of color, is possibly the most meaning-laden color in modern parlance. Even more than the others, the color black purports surprisingly strong positive and negative connotations. Black conveys mystery, elegance, and sophistication (e.g., a black-tie affair, in the black), while also conveying loss, evil, and negativity (e.g., blackout, black-hearted, black cloud, blacklist).

Clarification and Emphasis

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

Aesthetics

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

Abstraction

Color abstraction is an effective way to illustrate quantitative and qualitative data, particularly for thematic products such as choropleth maps. Here, colors are used solely to denote different values for a variable and may not have any rhyme or reason. shows a typical thematic map with abstract colors representing different countries.

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

Reserved Colors

There are certain feature/color combinations that are widely accepted.

Color	Example	Use
Blue		Water (streams, lakes, permanent snow fields and glaciers, et
Green		Forest and vegetation, parks etc.
White		A general lack of vegetation (or data)
Brown		Contour lines (elevation information)
Black		Man-made/cultural features (buildings, place names, bound
Red		Highways and major roads, Township/Range/Section infor
Pink		Urban areas
Purple		Reflect revisions to a map but is no longer used on maps in p

Data Classification

As discussed earlier, all maps are abstractions. This means they depict select information based on the question the geospatial problem addressed and the limits of display resolution, comparable limits of human visual acuity, and especially the limits imposed by the costs of collecting and processing detailed data.

Single Symbol

Single symbol is a method of stylizing data in a feature class will the same symbol shape and color. The character of the symbol can be customized to represent the feature class, but they are all like each other.

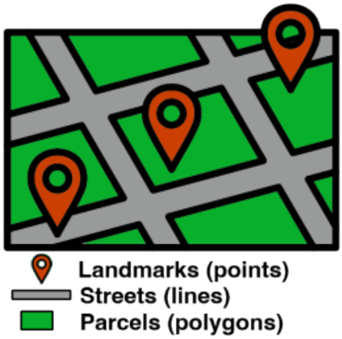


Illustration of single symbolization

Unique Values

Unique values uses categorical (ordinal) information to symbolize features using a specific variable, so that when values for that variable are the same, they can be symbolized the using the same symbol type or color.

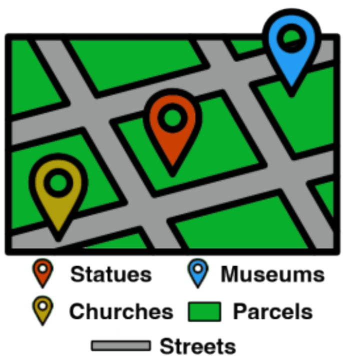


Illustration of unique symbolization

Graduated Colors

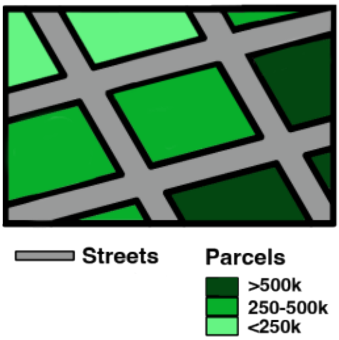


Illustration of graduated symbolization

Maps produced by this method are properly called choropleth maps. They are symbolized by a gradient representing numeric values which increase/decreases; generally high values are given darker shades, decreasing in shade as the numbers decrease. Because our ability to discriminate among colors is limited, attribute data values at the ratio or interval level are usually sorted into four to eight ordinal level categories. The categories are often referred to as classes. The number of classes may be adjusted, as can the method of parsing the values.

Parsing methods

Although our data values don't change for a variable in each dataset, the way those values are presented can suggest different meanings. It's important as map makers that we are transparent in how we have parsed the data for visualization

• ***Equal Interval*** : This method breaks values into consistent chunks. The chunks serve as milestones which provide a useful point of reference.

•***Quantiles*** : Divide values into five bins, each of which contains the same number of values. This is accomplished by varying the width or range of each class. Quantile is a general label for any grouping of rank ordered data divided into an equal number of entities.

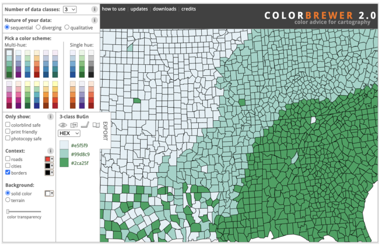
• ***Standard Deviation*** : This method finds the mean of a variable and then the distance from the mean for each value in that variable. Groupings are made based on that distance from the mean.

• ***Jenks Natural Breaks*** : This algorithm groups data values based on their likeness to each other. The number of classes will help determine what values are grouped.

Exercise

Visit Color Brewer and explore colorblind-safe ramps for each nature of map:

1. Sequential
2. Diverging
3. Qualitative



Screenshot from Colorbrewer

PART II

PART 2: HANDS-ON WORKBOOK

What to Expect from Each Chapter

Chapter 8: Planning Scenario focuses on describing the problem or enquiry, thinking through the goals, context, data collection,

Chapter 9: Collecting and Cleaning the Data will walk through finding and preparing data, for analysis.

Chapter 10: Adding Data and Symbolizing Data describes how to add data to QGIS to describe the project study area and basic symbology.

Chapter 11: Curating Your Dataset through Selecting Feature Selection describes several methods for isolating features of interest, using manual selection, selection by attributes, and selection by location.

Chapter 12. Basic Data Management focuses on joining tables to features, creating joins between features based on location, merging data, clipping data, and editing features.

Chapter 13. Thematic, Choropleth Maps walks through how to use data classification in practice to describe numeric data.

Chapter 14: Kernel Density Estimation (KDE) introduces using the KDE algorithm to understand the geographic magnitude of features in a study area.

Chapter 15. Buffer Analysis and Points in Polygons describes how to create a radius around features to understand what falls inside a distance.

Chapter 16. Communicating the Message describes more about symbology, labeling, layout and map essentials.

CHAPTER 8: PLANNING SCENARIO

Project Summary

Problem

I want to evaluate household demographics, zoning, and housing values in St. Louis County's inner-ring suburbs in the year 2010. I'll also look at other natural characteristics of areas, such as sinkholes and flooding.

Goal

Create an exploration between demographic, economic, and natural characteristics.

Context

As a resident of an inner-ring suburb of St. Louis, I would like to understand varying characteristics of the municipalities within.

DataCollection Pipeline

I will collect the following datasets:

- Census data for income and race (US Census Bureau)
- Parcel data for housing values (St. Louis County)
- Street centerlines (St. Louis County)
- Topographic data like contours and sink holes, flood plains (St. Louis County)

File Storage

- Active Storage: Local machine and cloud drive
- Backups in networked serves

Completed project diagram.

Analysis Methods

1. Find data, evaluate it for authenticity and appropriateness
2. Clean tabular data for use in GIS
3. Geocode address and/or display X,Y data.
4. Join tables to feature layers
5. Selection by attributes and location
6. Buffer features
7. Classify data
8. Kernel Density Estimation
9. Layout

Communication

- Create several maps that can viewed online or printed.

Challenges

- Finding the right data and processing it for use.

CHAPTER 9: COLLECTING AND CLEANING THE DATA

Getting the Data

Data are provided for the purpose of these exercises, but in this section, we will walk through what data we need and how to find it.

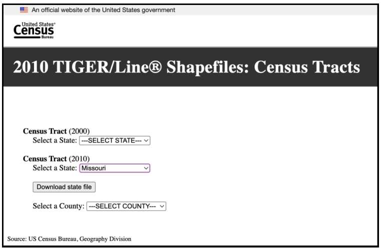
Data we will collect includes:

- Income (source: US Census)
- Housing characteristics (source: US Census)
- Demographics (source: US Census)
- Housing assessments (source: St. Louis County Open Data)
- Zoning information (source: St. Louis County Open Data)
- Schools (source: MSDIS)
- Floodways (source: St. Louis County Open Data)
- Sinkholes (source: MSDIS)
- Voting results (source: St. Louis County Elections)

- Bakeries (source: Data Axel)

Getting Census Data

The Census provides two types of data we need, geospatial and tabular. The exercises in this book will use Census Tract Level Data for St. Louis County.



Screenshot from Census Tiger/Line site

Getting Census Tract Tigerline Files

Go to the Census TIGER/Line Shapefiles page.

- Select year: 2010
- Select layer type: Census Tract
- Submit
- State: Missouri
    - County: St. Louis County
    - Download

The Census tables can be retrieved from the Census's Data and Maps Page. Census.gov provides the Decennial Census (10yr) as well as the American Community Survey (estimates between the decennial census).

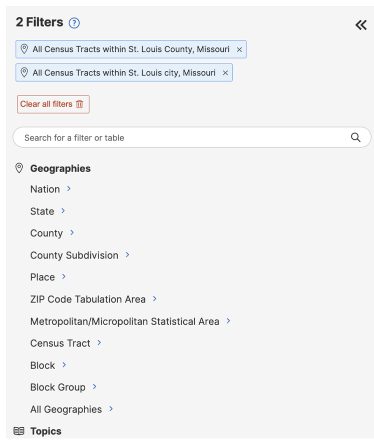
If you are unable to get historical data from the Census site, another useful tool is IPUMS NHGIS.

Using the **advanced search** in Census.gov, filter for the following.

1. Go to the Explore Census Data page and do a separate *Advanced Searches* for tables for housing, income, and race.

Geographies

- Census Tracts
  - Missouri
    - St. Louis County
      - All



Screenshot of filter from Census Data Portal

Census Tracts in St. Louis County Missouri

I recommend gathering these topic tables separately.

Topics

- Housing
  - Download 2010 table DPO2 (zip file)
  - Remove topic filters, keep geography filters

Topics

- Income and Poverty
  - Download 2010 table S1901 (zip file)
  - Remove topic filters, keep geography filters

Topics

- Race
  - Download 2010 table B02001 (zip file)

2. Getting Floodways

These are made available through St. Louis County Open Data Portal.

Search: flood

- Select Flood Hazard Areas, 1% Annual Chance
  - Select FEMA Floodways
  - Download shapefile dataset

3. Getting Sinkhole Data and School

Locations

MSDIS stands for Missouri Spatial Data Information Service, and it houses a variety of data sources for the state.

You can find schools by searching for schools.

- Download the .shp file

You can find sinkholes by searching for sinkholes

- Download the 2018 sinkholes .shp file.

4. Datasets can be found here.

Cleaning Data

Software for cleaning data:

First go to Open Refine and download software specific to your operating system (e.g. OpenRefine 3.9.3 for MacOS)

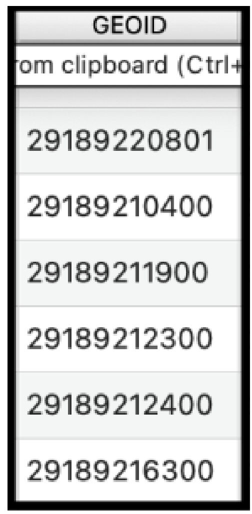
The cleaning we will need to do is on the Census tables. Our tables require a few basic actions:

1. Prepare the data so it will successfully join.
2. Ensure numerical columns will be read as numbers.
3. Remove any unnecessary information.
4. Document our changes

You will remember from earlier chapters that we can join tabular data to geospatial data. However, we can only accomplish this if we have a field that meets all the criteria for joining.

Requirements for a successful join

- Join fields must contain matching values (a unique identifier is an ideal value for joining)
- Join fields must be the same data format (e.g., string, integer, etc.)
- Field headers must not contain spaces or special characters
- Field values are clean and standardized
- Field headers should not use reserved words (e.g., date," "day," "month," "table," "text," "user," "when," "where," "year," or "zone")
- Table is in an incompatible format



GEOID
om clipboard (Ctrl+
29189220801
29189210400
29189211900
29189212300
29189212400
29189216300

Screenshot from QGIS application

Identifier values match:

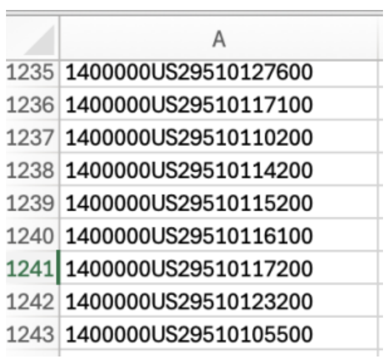
Our project will require us to join our Census tables to our Tigerline Files. The Census uses a field called Geoid to accomplish this.

A Tigerline GeoID looks like this: 29510127600; the id breaks down like this:

- 29 is the Census ID for the State of Missouri
- 510 is the County Number (in this case it represents the county of St. Louis City)
- 127600 is the Tract number
- On their own, none of these numbers are unique, but when amalgamated, it can be unique.

Census tables GeoID's often look like this when downloaded: 1400000US29510127600

1400000US in a national code. The rest of the id is the same as the id in the Tigerline file

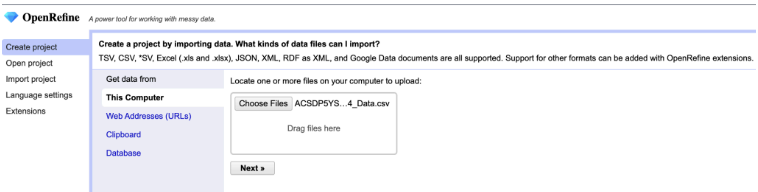


	A
1235	1400000US29510127600
1236	1400000US29510117100
1237	1400000US29510110200
1238	1400000US29510114200
1239	1400000US29510115200
1240	1400000US29510116100
1241	1400000US29510117200
1242	1400000US29510123200
1243	1400000US29510105500

Screenshot of Census table geoid in Excel app

Steps to clean the data:

1. Launch OpenRefine
2. Create Project
3. Upload, navigate to one of the downloaded Census tables
4. Next



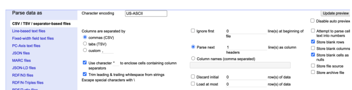
Screenshot of the OpenRefine upload screen

When prompted to parse the data you will be asked:

- **Columns are Separated By**: the column separation method, which is based on the type of file you are working with our data is in CSV format (comma separated values), so our separation method is a comma.
- **Use Character**: because sometimes your data will contain the character used for separation, we need to tell OpenRefine how to know when it's serving its separation function and when it's part of the value. We will accept the default (").
- **Trim leading and trailing white space from string**. Spaces are hard to clean out of a dataset because you can't really see them with your eyes when they're at the end of a value, but they can cause big problems for analyses. Open refine can remove the white when this is selected. We will select this box.
- **Ignore first line**: we want our first line treated as a header, so do not choose this option.
- **Parse next**: If our dataset has multiple headers we can

tell OpenRefine how many lines are reserved as headers. In our case we will choose one.

- **Column names**: if you have ignored the first line and want to provide column names you can list them here, for our purposes we will use the provided column names in the dataset, so we will not fill out this field.
- **Discard initial rows of data**: this can be helpful if there are rows of information preceding the dataset. For our purposes we will not use this.
- **Load at most**: if you don't want to load all the rows of the dataset you can indicate how many to load. For our purposes we will not be using this.
- **Disable auto preview**: If you don't need to see the parsing, check this. We will not check this

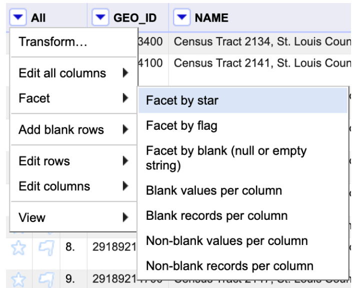


Screenshot of the OpenRefine parsing screen

- **Attempt to parse all text into number**: This can be useful if you know all the number character should. be text. In our case, we want our variable numbers to read as numbers, but our unique identifier, which is also numbers should read as a string, so we will not check this.
- **Store blank rows**: Unless there is a good reason, I would recommend removing blank rows because they will read as null value. We will deselect store blank rows.
- **Store blank columns**: Unless there is a good reason, I would recommend removing blank columns because

they will read as null value. We will deselect store blank columns.

- **Store file source:** This would store in OpenRefine, which is not necessary, but you want to document your source.



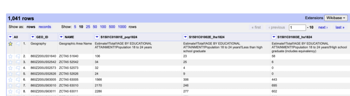
Screenshot of the OpenRefine star row action

- **Store file archive:** This would store in OpenRefine, which is not necessary, but you want to document your changes and keep a copy of the original file.

Remove Row 2

To clean up this data we need to remove row 2 so that we are working with number values only. To do this we will:

- Click the star symbol next to row two.
- Toggle down the "All" menu
- Choose "Facet by Star"
- Select the values that are "true"

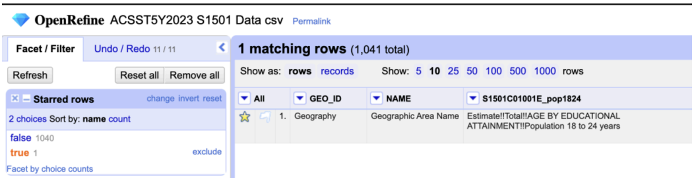


Screenshot of the OpenRefine create facet by star action

Split Column

Toggle down the "All"

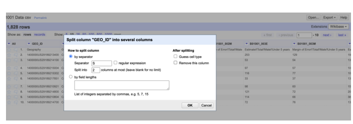
- Go to edit rows
- Remove matching rows
- Next step is to remove the extra numbers (1400000US) from the geoid.
- Toggle down in the geoid column, click "edit column"



Screenshot of the OpenRefine matching rows

Choose the option to "split into several columns"

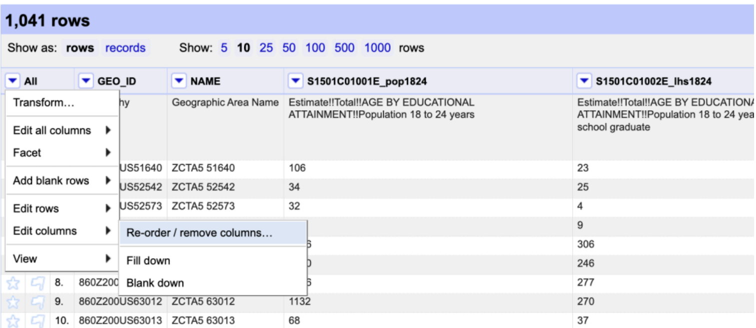
- How to split the column: by separator. We have to tell it what the separator, in our case the separator will be "S" because that's the last character in the string we want to remove.
- Split into 2 columns
- After splitting, "**deselect guess cell type**" because may guess number for our geoid and we want it to remain a string.
- "**Deselect remove this column**" because in case we need to redo it it's good to have it available.



Screenshot of the OpenRefine split columns dialog

If it parsed successfully, there are now three columns: the original column “geoid”, geoid1 has the content we want to remove (1400000US), and geoid2 has the tract id that will match our geospatial data.

Remove un-needed columns



Screenshot of OpenRefine reorder/split columns dialog

To remove one or two columns; you can toggle down to

- “Edit column”
  - select “remove column”

To remove several columns, toggle down the All column, Edit columns, “re-order/remove columns.” All the columns will be listed in the left pane, and the right pane will be empty.

You can drag and drop columns 1 by 1 **OR** you can select remove all columns, and all the columns will move to the left pane, then drag the ones you want to keep back into the left pane.

Using the metadata table for each of our tables, you can identify which columns to keep.

For the ACSST5Y2010_S1901, keep:

- GEO_ID_2
- S1901_C01_001E
- S1901_C02_001E
- S1901_C02_0012E
- S1901_C02_0013E

For the ACSDP5Y2010_DP04, keep

- GEO_ID_2
- DP04_0001E
- DP04_0045E
- DP04_0045PE
- DP04_0088E
- DP04_0088PE

For the ACSDT5Y2010_B02001, keep:

- B02001_001E
- B02001_002E
- B02001_003E
- B02001_004E

- B02001_005E
- B02001_006E
- B02001_007E
- B02001_008E
- B02001_009E
- B02001_0010E

Rename Columns

In each column, toggle down the column.

For the ACSST5Y2010_S1901, keep GEO_ID_2.

- S1901_C01_001E, change to totpop
- S1901_C02_001E, change to totfam
- S1901_C02_0012E, change to medfam
- S1901_C02_0013E, change to meanfam

For the ACSDP5Y2010_DP04, keep GEO_ID_2.

- DP04_0001E, change to totoccs
- DP04_0045E, change to estoo
- DP04_0045PE, change to prctoo
- DP04_0088E, change to medval

For the ACSDT5Y2010_B02001

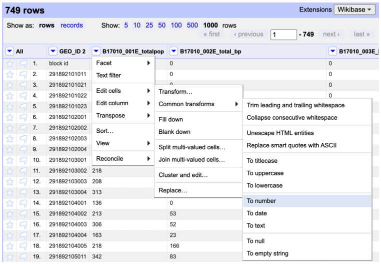
- B02001_001E, change to totpop
- B02001_002E, change to ewa

- B02001_003E, change to eba
- B02001_004E, change to eaiana
- B02001_005E, change to eaa
- B02001_006E, change to enhopa
- B02001_007E, change to esora
- B02001_008E, change to e2mra
- B02001_009E, change to e2mior
- B02001_0010E, change to em2reor3mr

Convert strings to Numbers

The variable columns are reading as strings because before row 2 was removed, there was text in those columns. Because this is a popular transform another option is to go to "**Common Transforms**" and select "**to number**".

After changing a few columns one-by-one, you might notice that there are A LOT of columns in this dataset. If you go to the "**All**" menu, choose "Transform", "To Number", you will get a checked list of columns. Deselect "**Geoid**" and "**Name**" and select ok.



Screenshot of the OpenRefine convert multiple columns to number

The final step in this dataset is to export the dataset. You can do that using the export button in the upper-right corner, chose to "**export to csv**". You will need to rename OpenRefine output as using the table number because QGIS will adopt the name of the table as a part of the variable name. For example rename:

- ACSST5Y2010_S1901 = S1901.csv
- ACSDP5Y2010_DP04 = DP04.csv
- ACSDT5Y2010_B02001 = B02001.csv

Make sure you clean and rename all the census tables.

CHAPTER 10: ADDING DATA

Adding Data in QGIS

You should already have downloaded QGIS. This chapter will focus on adding our data to QGIS, basic map formatting and getting to know the tool's basic functions.

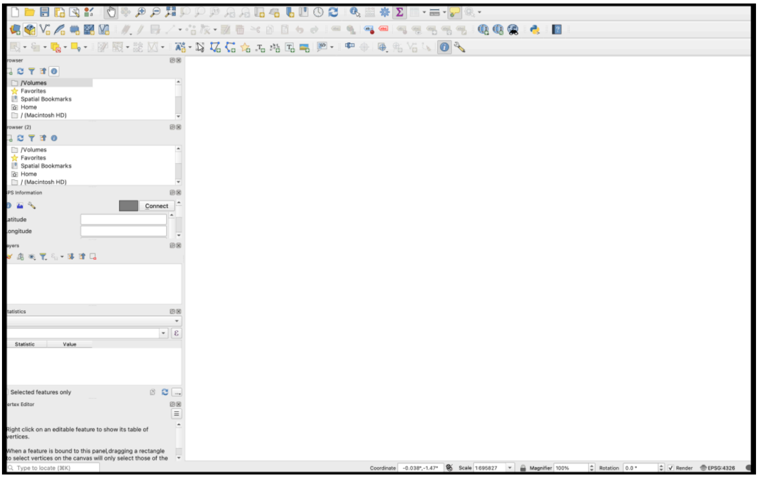
This chapter is completely focused on add geospatial data. Some of the data are in ***shapefile*** format, so those data are easy to add. Other data are in the form of ***services***, which requires adding via url.

We'll also bring in tables, which aren't quite geospatial, but have information in them to transform them to a geospatial layer.

In this chapter, we will also begin the practice of ensuring data are in the correct***projected coordinate reference system***.

Step 1. Launch QGIS.

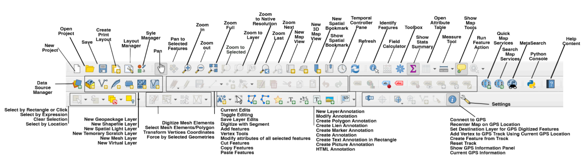
Your workspace should look something like this



The QGIS new project interface.

Tool Tour

Some of these tools we will use much more frequently that others. There are also other tools not in this tool bar that we will leverage, and plugins that we will install along the way.



The QGIS tools bar described
.

Commonly used tools include

- data storage manager
- zoom and pan

- identify features
- new map view (layout)
- refresh
- save

Datasets

If you are not finding and cleaning the data yourself, you can find the exercise data through an external site.

Step 2: Drag and Drop

Drag the .shp files from your external file manager and drop them into QGIS in this order:

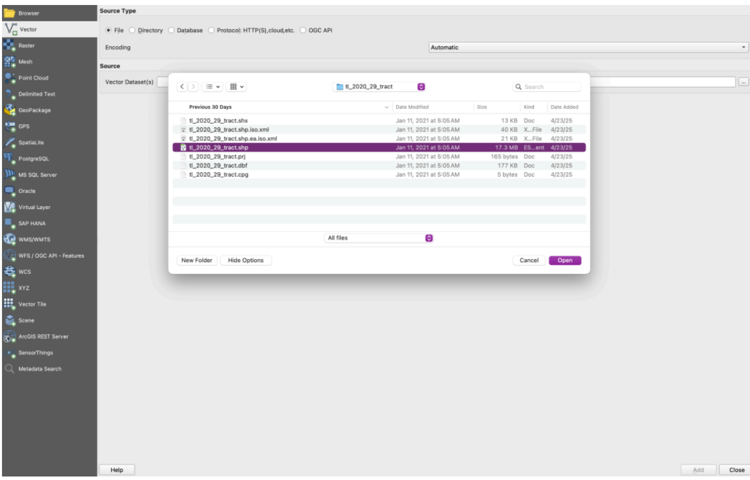
1. stlco_muni_bdy
2. stlco_fema_floodway
3. stlco_parcels
4. stlco_transport

Step 3: Browse to Layer Using the Data Manager

Use the data source manager within QGIS

- Click the file manager icon
- In the left-hand menu click the ***Vector*** tool

- Navigate to the .shp file, double click
- Accept the encoding defaults
- Click add

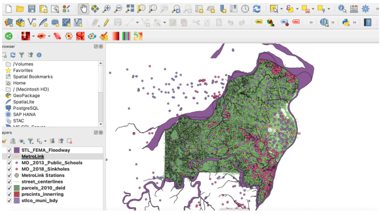


Navigate to files with the file manager.

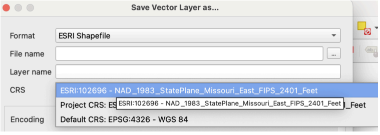
The shapefile layer will appear in the workspace.
Add remaining geometry datasets

- tl_2010_29189_tract10
- mo_2018_sinkholes
- mo_public_schools

Step 4: Check and change the Coordinate System



Screenshot of added data in QGIS

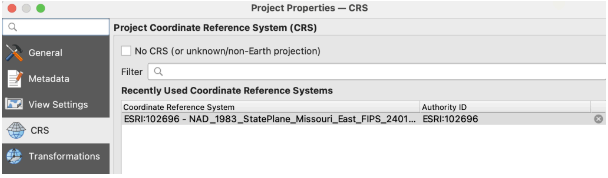


Screenshot of projecting data in QGIS

- Right click on the layer to view the CRS
- The CRS should be: ESRI:102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet
- EPSG is an index of coordinate systems. You can look up an EPSG number at https://epsg.io/
- Any layer not in the above projection should be exported with it.
    - Right click on the layer again and choose to Export, and then Save Features As
    - Name the same but add a "_p" and make sure it's in the correct folder.
    - Use the CRS picker to choose the right projection, which we found in EPGS.io.
    - Once we have put one layer in that projection, we can use that layer to assign projections.
- Ensure the project is also in the **ESRI: 102696** –

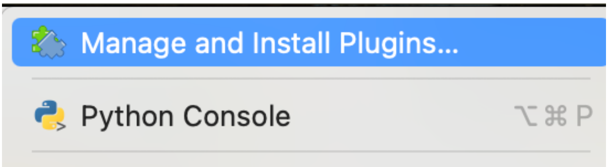
NAD_StatePlane_Missouri_East_FIPS_2401_Feet projection.

- Go to Properties
    - CRS
    - Recently Used Projections Systems
    - Select: **ESRI:102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet**



Screenshot of project CRS in QGIS

Step 5: Add a basemap (optional)



Use the manage plugins tool to add tools and features not in out-of-the-box QGIS

- Go to plugins
- Manage and install plugins
- Search for **quickmapservices,** install the plugin
- Under ***Web*** in the toolbar, an option for *QuickMapServices* will now appear.
- Go to QuickMapServices, OSM, OSM Standard

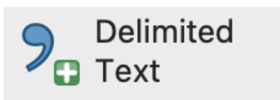
Step 5: Add Coordinate Data

Use the data source manager to add files.

We'll start with our dataset which contains locations of ice cream shops, coffee shops, and juice bars.

Use the **data source manager** within QGIS

- click the *Delimited Text* tool
- **Navigate to**: cafe_juice_ic.csv
- **File Format:** comma separated values
- **Geometry Definition**: Point Coordinates
  - o X field: Latitude
  - o Y field: Longitude
- Geometry CRS: WGS 1984
- Add
- Close



Use the delimited text option in the file manager.

The points are placed in the workspace, but we need to project them into our local coordinate system, **ESRI:102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet**

- Right click on the new layer. **Export**
- Save features as
- **File name:** cafe_juice_ic_p.shp
    - **CRS:** ESRI:102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet
    - OK
    - Right click and remove layer: cafe_juice_ic ***(don't remove cafe_juice_ic_p)***

Step 6: Install Geocoding Plugin:

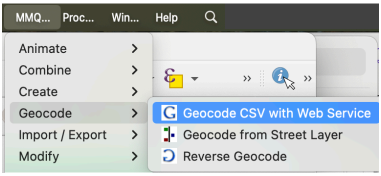
Before we can Geocode in QGIS, we must install a plugin.

- Go to the **Plugins menu** in the top tool menu
- Select Manage and Install Plugins
- Search for ***mmqgis***
- Install plugin***:*** mmqis will be added to the top tool menu

MMQGIS is a set of Python plugins for manipulating vector map layers in Quantum GIS: CSV input/output/join, geocoding, geometry conversion, buffering, hub analysis, simplification, column modification, and simple animation.

Step 7: Geocode Addresses

In the menu bar, go to **MMQGIS**



Screenshot of geocoding menu in QGIS

- Geocode CSV with Web Service
- **Input:** navigate to bakeries_clean
- Match fields in table to the web service geocoder
    - **Address:** Address
    - **City:** City
    - **State:** State
    - **Country:** Country
- **Web service:** Open Street Maps / Nominatim

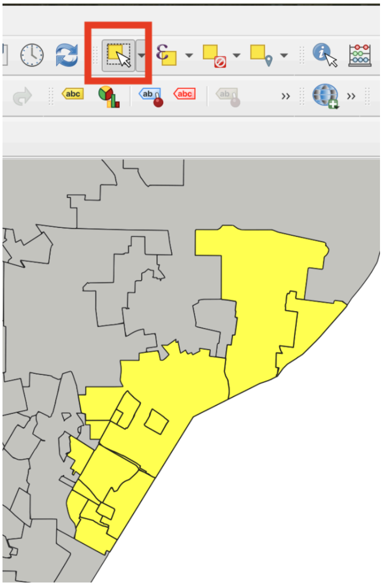
- **Duplicate Handling**: Use only first result MMQGIS
- Output: data folder in your directory
- **Not found Output:** data folder in your directory (using mmqgis, you will double check the addresses that didn't match outside of the GIS)
- **Export** the temporary layer using the same process as we did in the last layer we exported using the same CRS, named bakeries_clean_p

CHAPTER 11: CURATING YOUR DATASET WITH SELECTING FEATURES

Up to now we have added datasets which include information beyond our study area. As you refine your geospatial data an essential first step is to select the features that you want to use and remove the rest. In this chapter we will learn to manually ***select features***, ***select by expression*** querying the layer's attribute table, and by select by comparing the location of features in one layer to the location of features in another layer. It's important to note, that ***selecting by location*** is a time that projection matters. We want to ensure that our datasets are using the same CRS before we try to compare them.

Step 1: Manual Selection

In the case of manual selection, you are likely only selecting a few features or a rough area of features. There are several useful tools for creating selections.



Screenshot of manual selection in QGIS

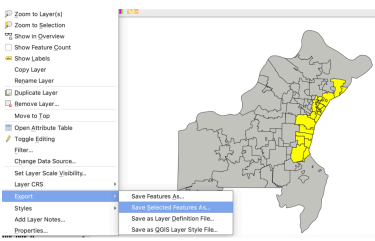
- Turn off all layers except the **stlco_muni_bdy layer**
- **Select features**: this is the most basic of these tools. It allows you to click and select a single feature. If you **click control** (Windows) or **command** (Mac) you can select several features drag and select a group of features
- **Deselect Features:** this allows you to clear the selection
- **Select Features by Polygon**: this tool allows you to draw a polygon by clicking to add vertices. When finished you can right click to create the selection.
- **Select Features by Freehand**: this tool lets you draw a shape on the layer and create the selection when you release the curser.

- **Select Features by Radius**: this tool allows you to click into the layer to create a starting point and drag a radius to create the selection when you let go of the curser.
- **Invert Selection:** this tool allows you to select all the features previously not selected.

Step 2: Select polygons

Select the polygons at the St. Louis City Boarder, except the top and bottom (unicorporated) polygons.

Export that selection to innerring_muni, using the same crs: ESRI:102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet

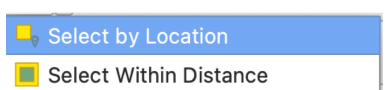


Screenshot of exporting selected features in QGIS

Step 3: Select by location

Now that we have our innerring_muni layer selected, we will use it to narrow our study area.

3a. Turn on the



Screenshot of select points within a distance in QGIS

tl_2010_29189_tract10_p layer

We're going to select from the innering_muni features to select inner ring tracts

- Use the select by location tool
- **Select layers from:** the tl_2010_29189_tract10_p
- Where the features: intersect
- By comparing the features from: innerring_muni
- **Modify by** creating a new selection

3b. Remove the tracts that fall too far outside the inner ring

Using select by location, intersect, sometimes features that intersect in one small area or share a boundary with, might be included. We must manually deselect those.

- Click and hold the command/control key
- Select the polygons that are not wanted

3c. Export the layer

- Save this to your file structure
- **File name:** innerring_tracts_2010
- **CRS:** ESRI:102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet
- **Encoding:** save only selected features.

- Turn off the tl_2010_29189_tract10_p layer

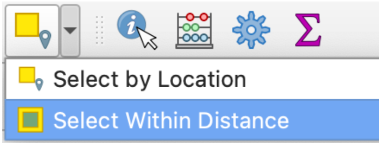
3d. Turn on the stlco_parcels layer

- Use the select by location tool
- Select layers from: the stlco_parcels
- Where the features: are within
- By comparing the features from: innerring_muni
- **Modify by:** creating a new selection

3e. Export the layer

- Save this to your file structure
- **File name:** innerring_parcels
- **CRS:** ESRI:102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet
- **Encoding:** save only selected features.
- Turn off the stlco_parcels layer

3f. Repeat 3e for other layers



Screenshot of select points within a distance inQGIS

- MO_2013_Public_Schools_p (within)

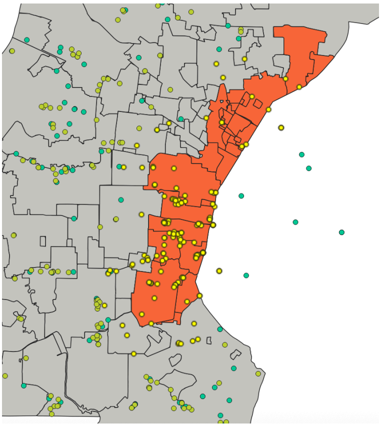
- MO_2018_Sinkholes_p (within a distatance, 2 miles)
- MetroLink_Sations (intersect)
- FEMA_Floodways (intersect, manually removing anything that falls so far outside the inner ring)

3g. Export each of these layers as we have done before

4. Select within a Distance

4a. Turn on the stlco_cafe_juice_ic_p layer

- From the selection menu, click select within a distance
- **Select layers from:** the stlco_cafe_juice_ic_p
- By comparing the features from: innerring_muni
- Where the features are within: 1.5 miles
- **Modify by:** creating a new selection



Screenshot of added points in QGIS

4b. Turn on the stlco_bakerier_p layer

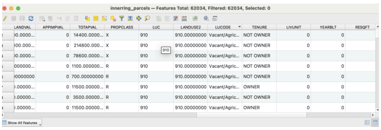
- From the selection menu, click select within a distance
- **Select layers from:** the stlco_bakeries_p
- By comparing the features from: innerring_muni
- Where the features are within: 1.5 miles
- **Modify by:** adding to the current selection

Step 5: Select by Attribute

We can use **select by attribute** to identify features based on a defined value. We do this by making a database query. Database queries are often executed using **standard query language (SQL) expressions**. We'll start by defining a question: ***Which parcels are zoned for residential use?***

5a. Find the column which contains the values

- Right click on innerring_parcels
- Look for a variable that describes parcel uses
- There is more than one possibility, but we will use LUCODE



Screenshot of identifying LUCODE field in attribute table in QGIS

5b. Create an expression

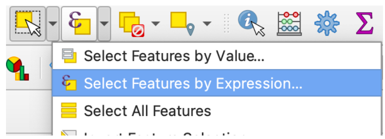
Field name: LUCODE

Operator: = (equals)

Value: Single Family

Syntax

Another very important thing is that the syntax must be correct. Syntax is necessary to define the query structure.



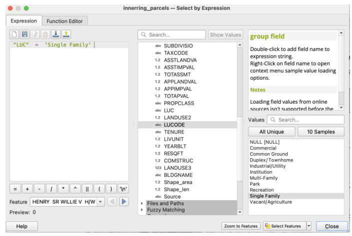
Screenshot of select features by expression in QGIS

- The field name requires "quotes"
- requires an **operator (e.g. =)** between the field and the value
- **values** should be in **single 'quotes'** unless it's numeric (e.g., integer or double)
- the case of the letter has to match the fields and values**.**

This the expression is: "LUCODE" = 'Single Family'

In QGIS, we can build this express by:

- Going to select features by expression
- An expression will start building in the left window
- In the feature geometry pane,
    - **Field:** Under the feature geometry pane, expand **fields and values**



Screenshot of LUCODE expression in QGIS

        - double click **LUCODE**, an expression begins building in the left pane
    - **Operator:** You can do one of these, they do the same thing
        - Type: =
        - expand the operators list in the feature geometry pane and select: =
        - oSelect: = button at the bottom of the tool
    - **Values**: In the lower right window, click ***Unique Values,*** which will list all the unique values for the LUCODE field; **choose Single Family.**
- The tool will check your syntax
- Once you click select features, the tool will provide a count of selected values

We'll export these features by right clicking on the layer

- Save it in your directory
- **File name:** innerring_sf_parcels
- **CRS:** ESRI 102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet
- Save only selected features.

CHAPTER 12. BASIC DATA MANAGEMENT IN QGIS

Data Management Tools

In this chapter, we're going to learn some basic tools to get our data in shape for analysis. We prepared our tabular Census data in OpenRefine so that we'll be able to take these steps in QGIS. Our first step is to join the prepared Census table to our tract data using a ***tabular join***. By doing that we represent Census values quantitatively by location. Next we will create a spatial join to give our coffee, juice, bakery, ice cream location attributes from our municipal boundaries layer. We are doing this so the location is tied to its municipality. The locations came in as separate layers, but it might be easier to work with them as one layer so we will merge those datasets. To make it easier to categorize the locations, we'll give them a new field were we can record their focus.

Step 1: Tabular Join

Once we have added our Census Tracts, we can join demographic data tables from the Census and visualize the values on the map.

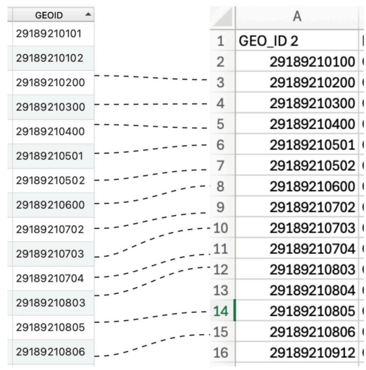


Illustration of a one-one join of GEOIDS

Review of Requirements for a successful join

- Join fields must contain matching values (a unique identifier is an ideal value for joining)
- Join fields must be the same data format (e.g., string, integer, etc.)
- Field headers must not contain spaces or special characters
- Field values are clean and standardized
- Field headers should not use reserved words (e.g., date," "day," "month," "table," "text," "user," "when," "where," "year," or "zone")

Join fields are often key fields in a table because they contain unique identifiers for each feature in the dataset. If our table has a field containing the unique identifiers that match the unique identifiers in the feature class, we can connect them.

1a. Add the census tables.

- Go to the data manager
- Add Delimited text
- Navigate DP04.csv
- Add to the project
- Repeat for:
  - B02001.csv
  - S1901.csv

1b. Join Income

Screenshot of add join button in QGIS

- Right click on one of the innering_tracts layers
- Select properties
- Select Joins
  - Click the plus sign and the join dialog appears
  - **Join layer**: S1901.csv
  - Join field: GEO_ID_2
  - Target field: GEOID10

If the join failed, go back through the requirements and make sure your data meets them.

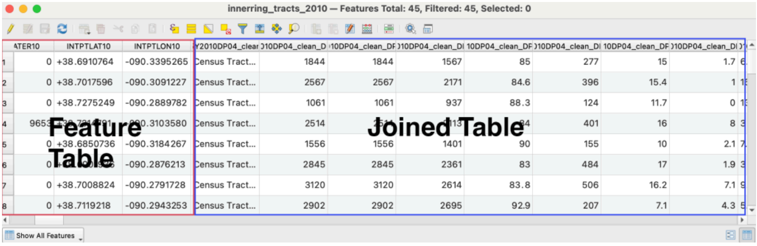
If the join succeeded but the values are null, go and check the requirements

Quickly symbolizing by graduated color can help you confirm the success of your join.

If you try to symbolize by graduated colors but don't see your fields, it's likely the software does not recognize your numbers as numeric, rather as numbers stored as text.

This often happens if row 2 has text in it. An easy solution is to add a dummy row two with a number in it.

This is necessary because we need our unique id column to read as numbers stored as text, but our value field to read as numbers stored as numbers. This dummy row will not affect the data because that field will not have a unique identifier and therefore will not join to the geospatial data. However, it is a good idea to note your dummy row in your documentation.



Screenshot of successful join in QGIS

1c. Export layer as innerring_tract_income

1d. Remove join

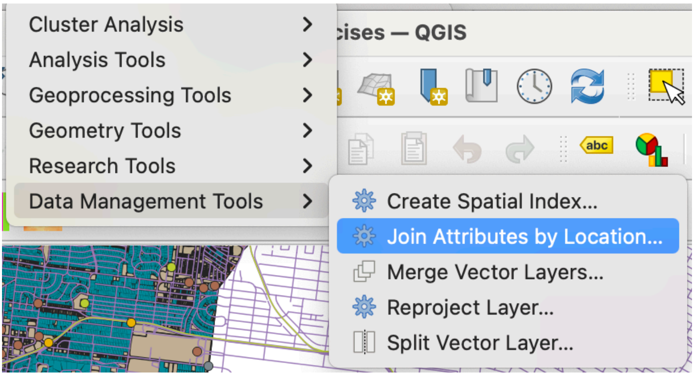
- Right click on innerring_tracts
- Go to properties
    - Joins
    - Click the minus icon
- The join will be removed.

1d. One by one, join the other layers, and export them to their own layer.

Step 2: Spatial Join

- In QGIS, a spatial join appends attributes of one feature to another feature. We are going to append the inner ring municipalities to include which school district is serves that area. To do this, we must go to the ***vector menu*** in the top tool bar.

- Select Data Management



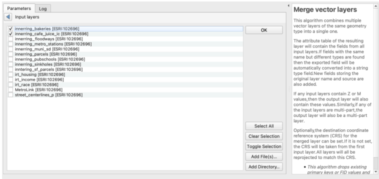
Screenshot of join by location in QGIS

- Join Attributes by Location
- Join to Features in: innerring_muni
- Features they (geometric predicate): intersect
- **By comparing to:** innering_pubschools
- Fields to add: CityDist
- **Joined layer:** create temporary layer
- Run
- A temporary layer will be created
- Open the temporary layer, there should be a new column added to it called CityDist
    - Some of the values will be null because they may not have a school district
    - Check to make sure not all values are null
- Export layer, name it innerring_muni_sd
- Destination CRS: ESRI 102696

–NAD_StatePlane_Missouri_East_FIPS_2401_Feet

Step 3: Merge datasets

I would like the business locations to be in the same layer. To achieve this, we can use the merge vector layers tool



Screenshot of merging data in QGIS

- Go to the ***Vector menu*** at the top
- Select Data management
- Select Merge Vector Layers
    - **Input layers:** innerring_cafe_juice_ic; innerring_bakeries
    - **Destination CRS:** ESRI 102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet
    - **Merged**: create temporary layer
    - Run
- Open the attribute table, there should be 153 features in the new table (63 innerring_cafe_juice_ic plus 90 innerring_bakeries)
- Export merged layer:
    - Save features as
    - Navigate to your directory structure

  - **File name:** innering_business
  - **CRS:** ESRI 102696 – NAD_StatePlane_Missouri_East_FIPS_2401_Feet
  - OK

Step 4: Add a new field

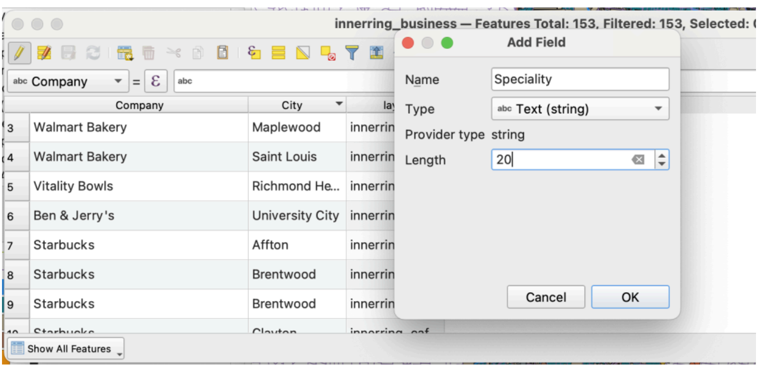
We might like to categorize the businesses by their specialty. Café, Bakery, Juice Bar, Ice Cream Shop. We can accomplish this by adding and populating a field.

Before adding the new field, I'm going to toggle off some the less helpful fields to ensure the table is easier to work with.

- Open the ***innerring_ business*** layer's attribute table
- Click the organize columns tool
- We will uncheck all the columns that are not serving our purpose; we can always toggle them on again if needed.
- Uncheck all fields except:
  - Company
  - City
  - Layer

So we can easily distinguish between what the specialty of each location is, we will add field to categorize; this will allow us to parse the data by that category.

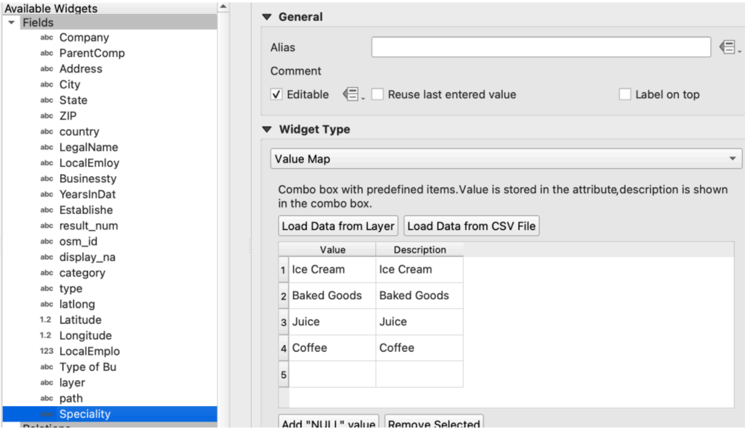
- Turn on (toggle) editing mode
- Use the ***add field*** tool
- **Name:** Specialty
- **Type:** Text (string)
- Length: 20
- OK
- Be sure to save the edits in the table



Screenshot of of adding field parameters in QGIS

Step 5: Create Value Mapping

Now we need to create standard values to keep the dataset tidy, also known as data validation.

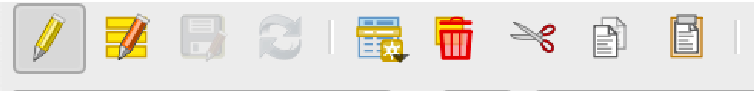


Screenshot of adding standard values to a field in QGIS

- Go to the layer properties
- Select ***Specialty*** from
- Choose the attribute form
- In the **widget type** box, change it to ***value map***
- Add the values and descriptions. In our case, they are the same:

Value	**Description**
Ice Cream	Ice Cream
Baked Goods	Baked Goods
Juice	Juice
Coffee	Coffee

Now we can populate the table with standard values using a dropdown menu. We can use the layer type column to help us with Baked Goods, but otherwise we'll need to look at the business name, in some cases we may need to google the name to be sure.



Screenshot of editing menu in QGIS

Step 6. Remove rows that don't fit into a specialty that fits into our four groups.

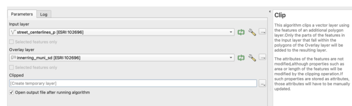
- Example: Sugarfire specializes in barbecue
- Select the row in the table by clicking the box all the way to the left of each row
- Use delete tool in the table to remove the row/feature
- Save the edits in the table

Step 7: Clip Street Centerlines

Now we will us a tool to clip out the St. Louis County parcels that are within the inner-ring municipalities.

- Turn of all the layers except the street_centerlines_p
- From the vector menu, select geoprocessing tools
- Choose the ***clip*** tool

- **Input layer**: parcels_2010
- **Overlay layer**: innerring_muni
- **Clipped**: Navigate to your folder, save as innerring_streets
- **Check:** open output file after running
- Run

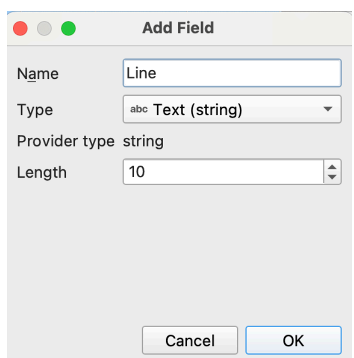


Screenshot of the clip tool in QGIS

Step 8: Edit Metrolink

Right click and open the attribute table on the Metrolink. Notice there's not field distinguishing the Red Line (North West) and the (South West) Blue Line.

- Click the edit button
- Click the add field button
    - **Name**: Line
    - **Type:** text
    - **Length:** 10

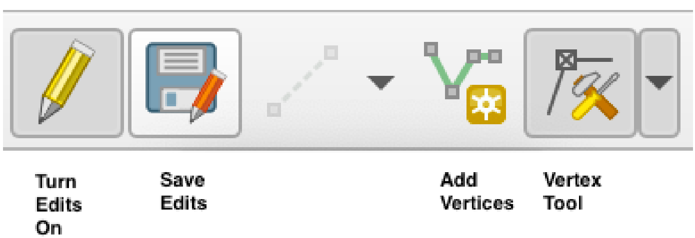


Screenshot of the add field dialog in QGIS

The Metrolink divides into Northwest and Southwest clearly in the map, however the geometry makes the Red Line division later than it should.

We can edit the lines so we can symbolize red and blue. With the editor still on

- Select the add add feature tool
- Draw the part of the line that should be the Red line, not both lines
- Save edits
- Select the row that represents both lines and stretches into the red line
- Use the Vertex tool to select each vertex reaching into the red line and delete them
- Save edits

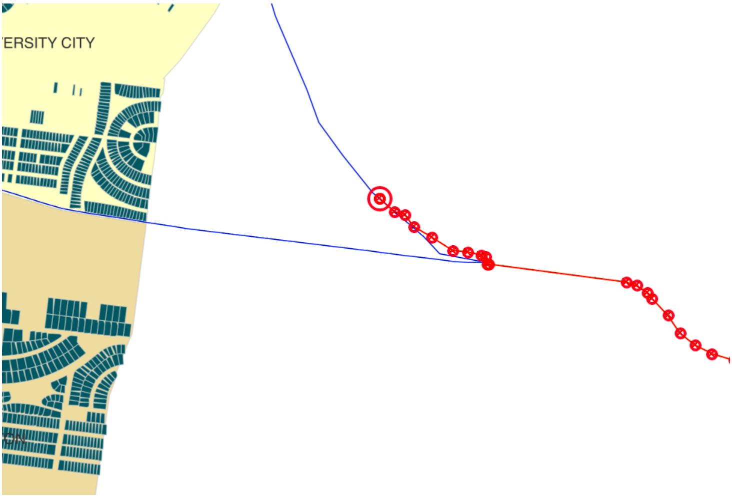


Screenshot of edit vertices tools in QGIS

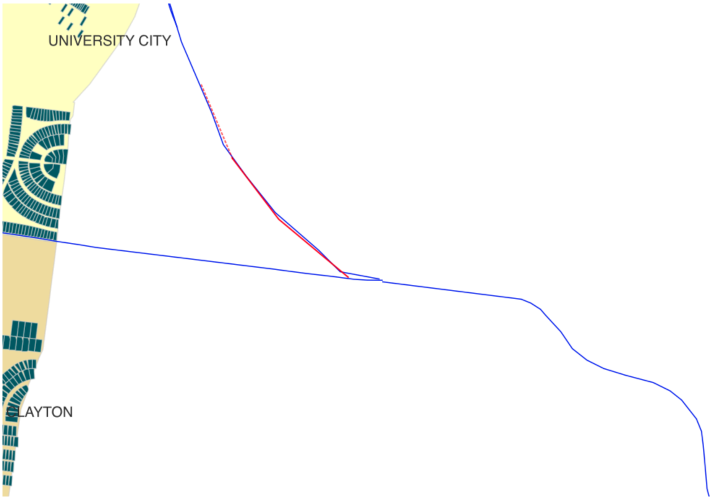
In the line column, populate the fields:

	OBJECTID	ID	Shape_Leng	Shape_len	Line
1	1	1	8608.77343...	8608.77345...	Blue
2	2	2	31726.27536...	31726.27554...	Blue
3	3	3	44728.3896...	44728.3900...	*NULL*
4	4	4	30680.3423...	30680.3419...	Red
5	5	5	10053.41844...	10053.41820...	Red

Screenshot of the Metroline attribute table in QGIS



- Red
- Blue
- Blue/Red



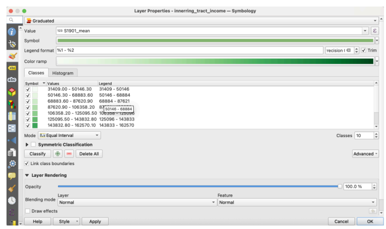
CHAPTER 13: THEMATIC / CHOROPLETH MAPS

Chapters 13- 16 will focus more on analyses than we have thus far by exploring some different way of quantifying and visualizing how much of something there is in a location. First, we will simply visualize quantities by separating them into bins, next we'll use a tool called kernel density estimator to find areas of concentrated features, or hot spots. Fiwe will count the number of one feature that falls into another location.

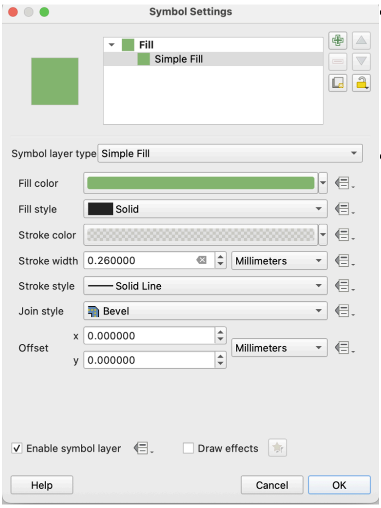
- Graduated colors
- Kernel Density (chapter 14)
- Create a buffer (chapter 15)
- Points in polygons (chapter 15)

Step 1. Creating thematic maps by graduated symbology (choropleth): numbers represented by colors or symbols

We are going to symbolize by a value in our attribute table. As discussed in Chapter 7, we can tell a different story with our data based on how we break up the values. Anytime we want to express the magnitude of a phenomenon using classification, we make tradeoffs. It's important to be transparent about those tradeoffs.

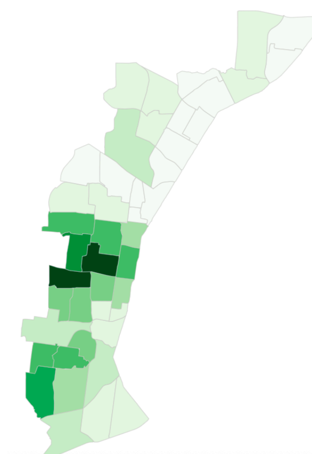


Screenshot of graduated colors parameters for mean income in QGIS



Screenshot of formatting symbols in QGIS

- Turn off the other layers except for innerring_tract_income
- Right click on the innerring_tract_income
  - Go to Properties
  - Click on symbology
  - From the dropdown menu, change from single symbol to **graduated colors**
  - We want to symbolize the **mean income** for families for **innerring_tract_income**.
  - **Value:** S1901_mean
  - I used a green ramp to reflect income, but a neutral ramp would be fine too.



Screenshot of gradated mean income in QGIS

    - **Mode:** equal interval
    - Classes: 10
- To change the boarder color, drop down **Symbol**
    - You don't need to change the fill
    - Change the stroke to **light gray** and make it **50% transparent**

1b. Symbolize the other census layers

Black Population

For the race tract we will use an expression to normalize the values so they are more comparable. In this case we will divide the variable estimation in a tract by the total population estimation in a tract. Innerring_tract_race

Screenshot of expression for normalized race in QGIS "B02001_eba"/"B02001_tot".

- **Value:** "B02001_eba" / "B02001_tot" (estimated black population/total population). To get this value click the expression button
- **Mode:** Natural Breaks
- Classes: 10

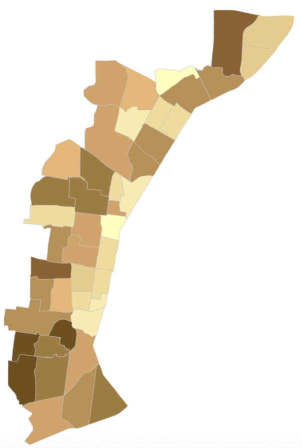
Mean Housing Value

Innerring_tract_housing

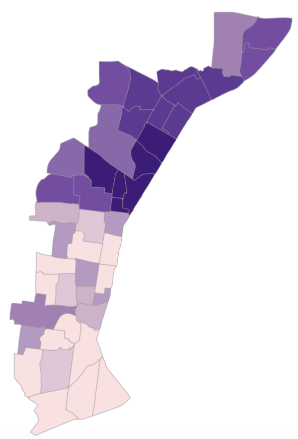
- **Value:** DP04_estme (mean housing value)
- **Mode:** Natural breaks
- **Classes:** 10

Households which are Owner Occupied

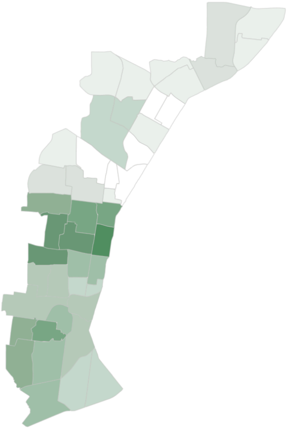
- Duplicate the innerring_tract_housing layer
- Rename the layer innerring_tract_housing_oo
- **Value:** DP04_prcto
- **Mode:** natural breaks
- **Classes:** 10



Screenshot of owner occupiers shaded shaded by graduated colors in QGIS



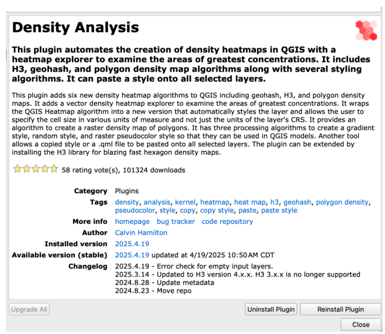
Screenshot of black population shaded shaded by graduated colors in QGIS



Screenshot of median home values by graduated colors in QGIS

CHAPTER 14: KERNEL DENSITY ESTIMATION

Step 2: Density Analysis Plugin



Screenshot of density plugin QGIS

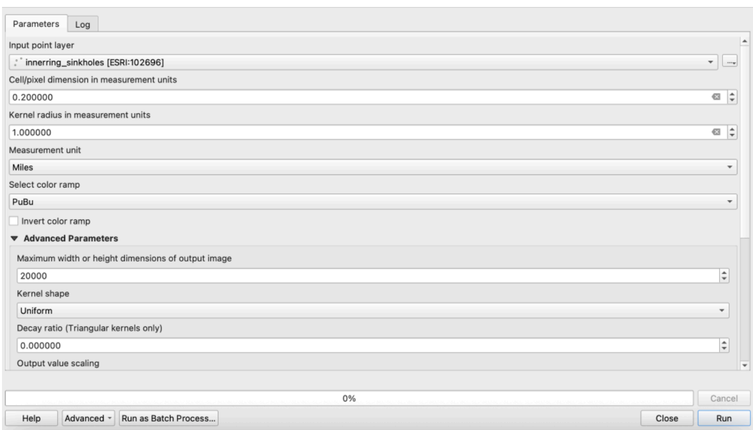
A very useful tool for demonstrating density of a phenomenon is to run Kernel Density Estimation (KDE). KDE measures density of features in relation to their neighborhood using weights. KDE can be used for vector data and creates a smoothed, raster output. As of publishing, QGIS does not have a KDE tool out-of-the-box but you can add the Density Analysis plugin.

- In the Top menu bar, go to Plugins
- Search for Density Analysis
- Install plugin

Screenshot of density toolbar in QGIS

When the plugin is successfully installed you will see a new toolbar in your workspace.

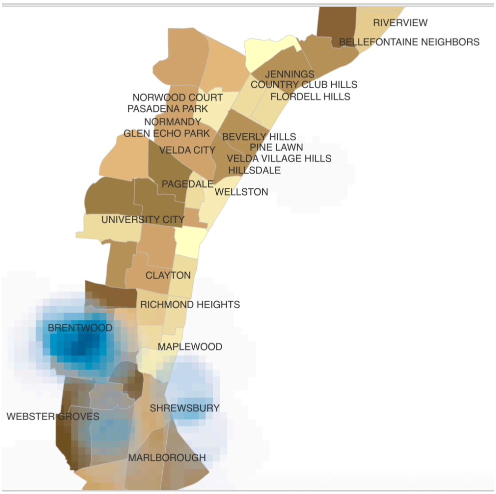
Step 3: Kernel Density Estimation



Screenshot of kernel density estimation parameters in QGIS

- from the Density toolbar, click the ***Raster Density tool***
    - **Input point layer**: innerring_sinkholes_2018
    - **Cell/pixel dimension:** .1 (the higher the value, the lower the resolution)
    - **Kernel radius**: 1.0 (this is the search distance)
    - **Units**: miles

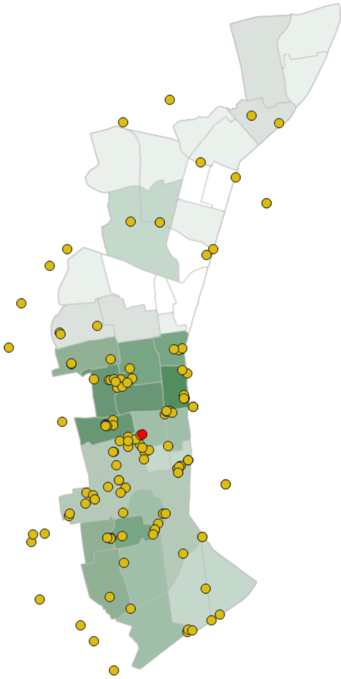
- **Color ramp**: you choose
- **Kernel shape**: uniform (feel free to try other shapes)
- **Decay ratio**: default
- **Output value scaling**: default
- **Interpolation**: linear
- **Mode:** continuous
- **Number of gradients**: default



CHAPTER 15: BUFFER ANALYSIS AND POINTS IN POLYGONS

Create a central point from which to buffer

Before we create our buffer, we will find the geographical center of our businesses. To do this we will calculate the mean coordinates, which computes a point layer with the center of mass of geometries in an input layer. Doing this computation will provide us a a center point for our buffer.



Screenshot of the central point feature result in QGIS

Step 1: From the

vector menu, go to analysis

- Select mean coordinates
- **Input layer**: innerring_businesses
- **Weight**: none
- **Unique ID:** none
- **Mean coordinates:** create temporary layer
- Open output file after running algorithm
- Run

A feature (red in the image) will appear at the center of the innerring_businesses

Step 2: Buffer the central point.

We might like to know how many businesses fall within the center. To do that we can make a radius around the central point.

- From the vector menu go to the geoprocessing tools
- Select buffer
- **Input features:** mean coordinates
- **Distance:** 2 miles
- **Segments**: 25
- **End cap style:** round
- **Miter:** default
- Run

Step 3: Count points in polygons

- Go to the vector menu, analysis, ***Count Points in Polygons***
- **Polygons:** buffer
- **Points:** innerring_business
- **Weights**: none
- **Class field**: none
- **Count fields name**: NUMPOINTS

Step 4; Buffer floodways

- We might also like to know what parcels are falling near t flood zone. We'll also buffer the flood way.
- Go to the vector menu, geoprocessing tools, select buffer
- Input features: innerrring_fldwy
- **Distance:** 200 feet
- **Segments:** 25
- **End cap style:** round
- **Miter:** default
- Run

Step 5: Now we'll find parcels that fall into the 200 foot buffer

- Select by location
- **Select features from:** Innerring_sf_parcels

- Where the features are ***within***
- **Comparing to**: fldwy_buffer
- Run

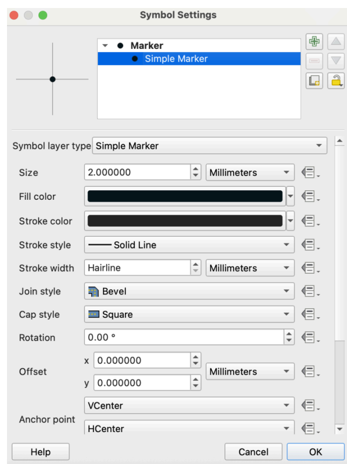
If you open the attribute table, you can see 3764 parcels fall in that buffer.

Step 6: Export the fldway_parcels

We have exported data many times up to this point, repeat those steps.

CHAPTER 16: COMMUNICATING THE MESSAGE

- Now that we have done some basic analysis it's time to share some of the insights we've made. Through doing the work, we may have had some insights, and we'll probably have more as we put things together. Some reminders:
- Everything doesn't have to go in one map; be sure to curate what information is going into the map
- Color choices are important
- Some features only use certain colors, e.g., water is blue
- Using non-neutral colors can relay meaning, be aware when using colors that have meaning to provide the context for the use

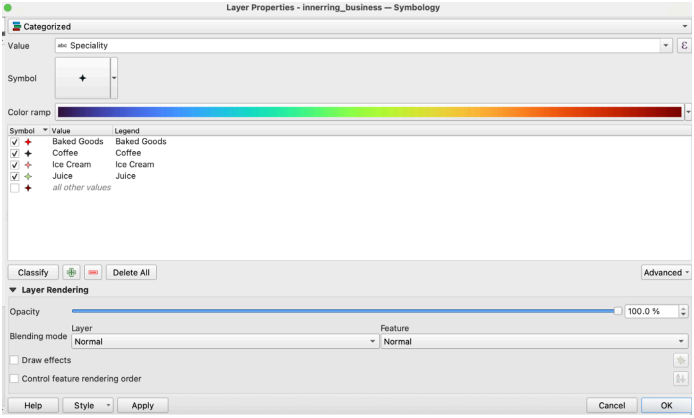


Screenshot of formatting all symbols in QGIS

- Be sure to include information that helps your reader understand the meaning

Communication Checklist

- Meaningful **title** which helps readers contextualize the map.
- **Legend** which describes the features on the map.
- **Scale bar** that helps readers understand distance relationship between features on the map
- **North Arrow** which helps orient the view to its relationship / position in the world
- **Credits** which include who put this collection of data, analysis, inference, and communication together (you), the date or relative date it was created, the sources and dates of each dataset used, (Projected)
- **Labels** of features on the map that help the view understand the study area
- Coordinate Reference System

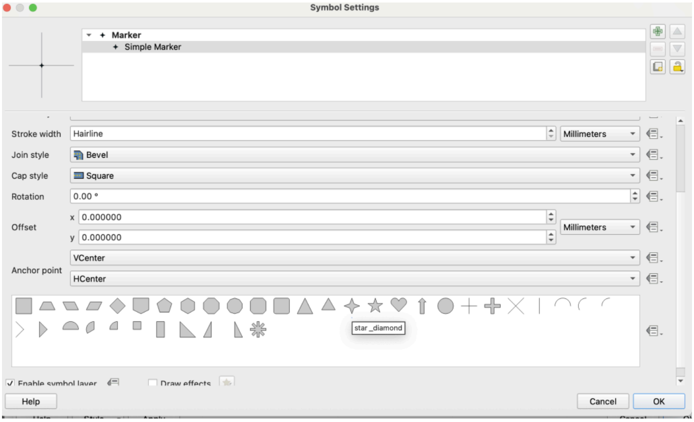


Screenshot of category symbology in QGIS

How to symbolize by Unique Values

We have already explored symbology by single symbol and by graduated colors, another common symbology is unique values. This is a method of visualizing a category defined it the attributes.

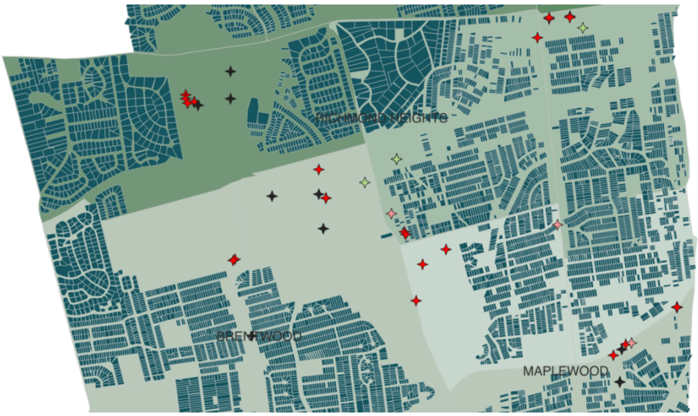
- Right click on innerring_businesses
- Go to symbology
- Select Categorized from the first dropdown menu



Screenshot of simple SVG marker in QGIS

- Value: Specialty
- You can double click on symbol and change the symbol shape
- You can also change the stroke, icon size, and angle in this menu
- Each shape will get a different color, but have the same shape.

SVG collections



Screenshot of a map with simple SVG markers in QGIS

It's possible to get a library of icons in QGIS, got to the plugins menu and install the **resource sharing plugin.**

- In the symbology dialog, click symbol
- Change the method to svg
- Scroll down to collection to the svg browser and choose an icon

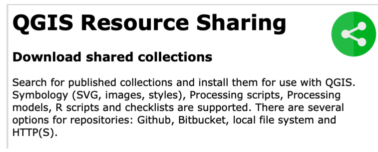
Custom svg icons

It possible to upload custom icons.

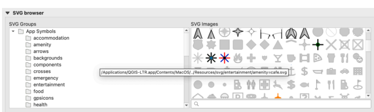
Note, the image must be in .svg format and saved to your local directory

You can use the sample svg that are in the data folder for these exercises

In the QGIS project preferences

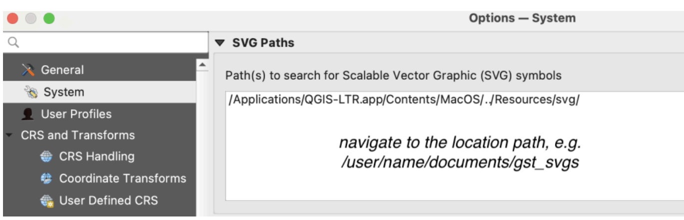


Screenshot of resource sharing plugin in QGIS



Screenshot of svg icons in QGIS

- Go to systems



Screenshot of setting path to svg custom icons in QGIS

- Navigate to the location where your .svg is stored
- Double click the symbol for baked goods
- Change the method from simple maker to svg marker
- Scroll down in the svg browser until you see the icons you added

- Double click the croissant press ok
- Repeat for:
  - ice cream
  - juice
  - coffee

Screenshot of browsing to custom svg icons in QGIS

Custom markers adapted from Noun project

Bakery symbol

Ice Cream Parlor Symbol

Juice Bar Symbol

Cafe Symbol

How to label

It's often not possible to label everything on your map, but strategically naming features on your map, can help the user. In the case of the municipalities, we may not have the feature symbology showing prominently, but we can label them to give views a relative idea of where they are.

- Double click innerring_muni_ds
- Select labels

In the first drop-down menu we must select how features will be labeled, the options are:

- None – no labels for a feature layer are the default
- Single – if we label feature, this is the most common choice, you might label by the feature name for example
- Rule based – this option allows us to create an expression for how the features should be labeled.
- Blocking – this allows us to determine the feature to label priority.

Select rule-based

- Use the plus sign at the bottom to add a rule
- Use the create expression button to create the rule
- The expression should be “MUNICIPALI” IS NOT ‘UNINCORPORATED’
- OK

Create a Layout

I’m going to create a few layouts based on what I’ve learned.

Map 1: Locations of Businesses, Schools by Student/Teacher Ratio, and Mean Income

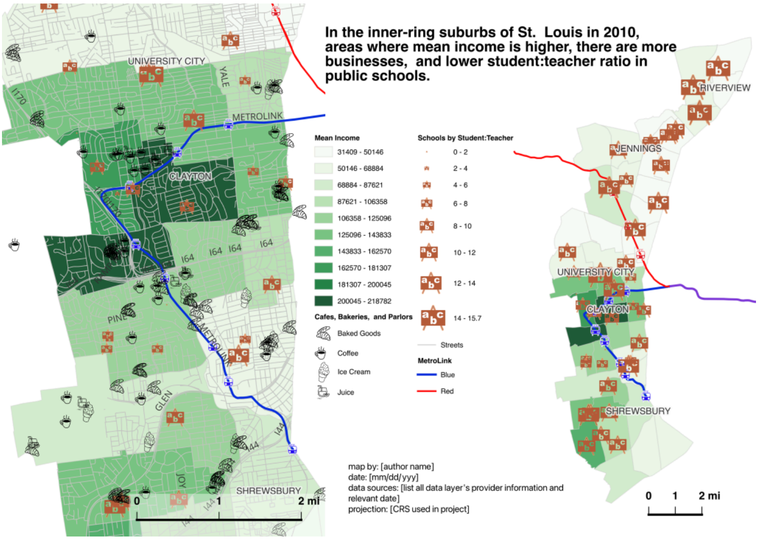


Image of map 1, focused on income, schools, and businesses

Map 2: Comparison between mean income, race, and housing values in census tracts in 2010.

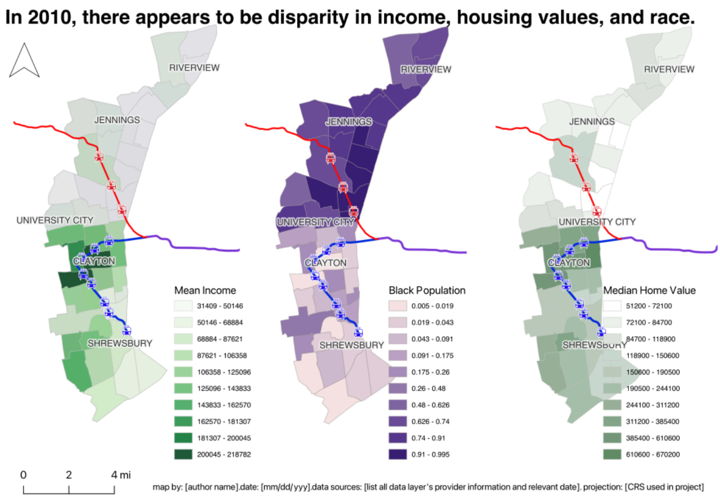


Image of map 2, focused on housing values, income, and race

Map 3: Median home values in 2010 census tract and the relationship to floodways, and sinkholes.

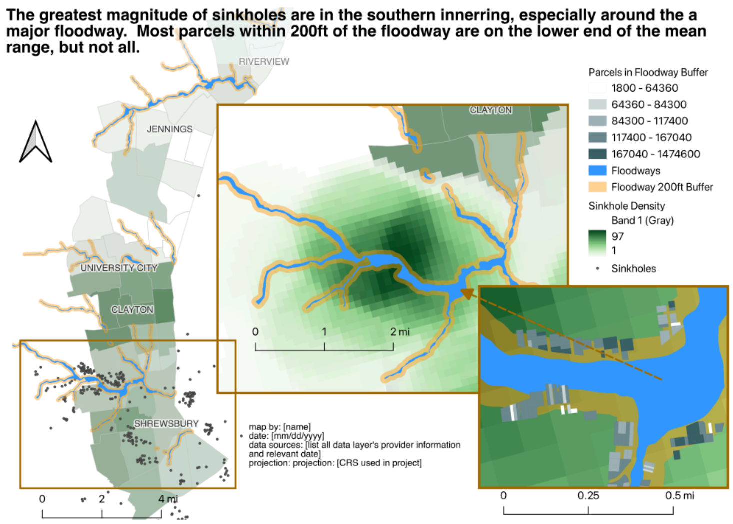


Image of map 3, focused on income, floodways and sinkhole density, and effected parcels

Map 4: Median home values in 2010 census tract and the relationship to owner occupancy and individual single-home values

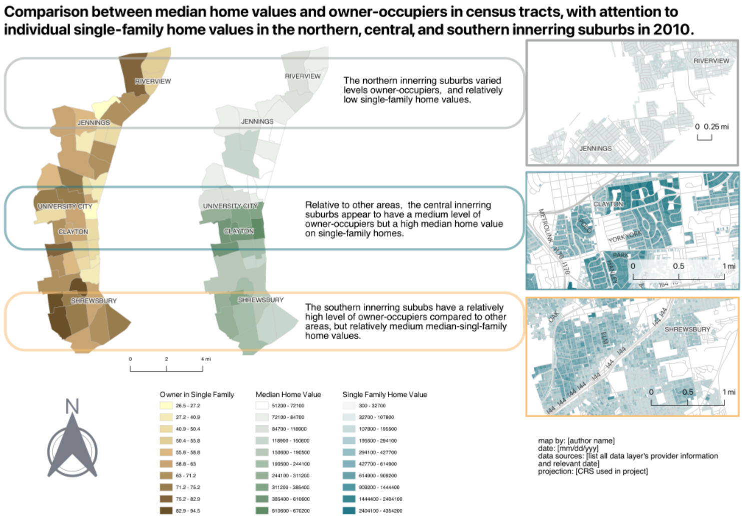


Image of map 4, focused housing values, owner occupancy, and single-family homes

IMAGE CREDITS

Sources of images are listed by chapter and in filename order Links current as of October 8, 2025.

Chapter: 1

Image 001: John Snow's 1854 Cholera map. Source: Wikimedia, https://commons.wikimedia.org/wiki/File:Snow-cholera-map-1.jpg

License: Public Domain, https://creativecommons.org/public-domain/pdm/.

Image 002: Example Mental Map created. Source: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Chapter: 2

Image 003: Illustration of GIS components. Source: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 004: QGIS logo, Source: QGIS, https://qgis.org. **License:** GNU General Public License, https://qgis.org/license/

Chapter: 3

Image 005: Illustration of the shape of the earth. Source: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 006: Illustration of latitude lines. Source: Wikipedia, https://commons.wikimedia.org/wiki File:Latitude_and_Longitude_of_the_Earth.svg
License: CC0 1.0, https://creativecommons.org/publicdomain/zero/1.0/

Image 007: Illustration of longitude lines. Source: Wikimedia, https://commons.wikimedia.org/wiki File:Latitude_and_Longitude_of_the_Earth.svg
License: CC0 1.0, https://creativecommons.org/publicdomain/zero/1.0/

Image 008: Illustration of lines meeting to form a graticule. Source: Wikimedia, https://commons.wikimedia.org/wiki/File:Spherical_coordinates_on_a_globe_dA.svg.
License: CC-BY-SA-4.0, https://creativecommons.org/licenses/by-sa/4.0/.

Image 009, Illustration of datum the division between NAD27 and NAD83 . Source: Wikimedia,

https://commons.wikimedia.org/wiki/File:Datum_Shift_Between_NAD27_and_NAD83.png.
License: CC-By-3.0, https://creativecommons.org/licenses/by/3.0

Image 010, Illustration of an orange representing a 3D earth unpeeled to represent a 2D map. Source: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 011: Illustration of a light bulb projecting from the center of the globe. Source: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 012: Illustration of projection categories. Source: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 013: Illustration of Azimuthal Equaistant project. Source: Wikimedia, https://commons.wikimedia.org/wiki/File:Taipei_centered_azimuthal_equidistant_projection.gif
License: CC-SA, https://creativecommons.org/licenses/by-sa/4.0/

Image 014: Illustration of UTM zones. Source: Wikimedia, https://commons.wikimedia.org/wiki/File:Utm-zones-USA.svg
License: CC-By-3.0, https://creativecommons.org/licenses/by/3.0/

Image 015: Illustration of State Plane zones divided by

whether they are Lambert or Mercator projections, adapted by Jennifer Moore based on a similar image used by Fang, and dataset:
"USA State Plane Zones NAD83." n.d. Accessed October 6, 2025. https://hub.arcgis.com/datasets/esri::usa-state-plane-zones-nad83/about.

Chapter 4

Image 016, Illustration of points, lines, and polygon layers. **Source**: Jennifer Moore**.**

Image 017, Illustration of a topological model. **Source**: Jennifer Moore**.**

Image 018, Clipping of a Seurat painting demonstrating pointillism. Source: Wikimedia.

Image 019, Illustration of cell-by-cell raster encoding. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 020, Illustration of run-length raster encoding. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 021, Illustration of quadtree raster encoding. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 022, Illustration of a raster satellite image at different resolutions. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 023, Illustration of a hierarchical database schematic. **Source**: Jennifer Moore**.**

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 024, Illustration of a network database schematic. **Source**: Jennifer Moore**.**

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 025, Illustration of a relational database schematic. **Source**: Jennifer Moore**.**

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 026, Illustration of a directory structure schematic.

Source: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Chapter 6

Image 027, Illustration of a project lifecycle for data. **Source**: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 028, Illustration of a project diagram. **Source**: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Chapter 7

Image 029, Illustration of a GIS workflow. **Source**: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 030, Illustration of single symbolization. **Source**: Jennifer Moore.
License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 031, Illustration of unique symbolization. **Source**:

Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 032, Illustration of graduated symbolization. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 033, Screenshot from Colorbrewer . **Source**: ColorBrewer,
https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3. AZXq`

License: Apache License, Version 2.0, https://www.apache.org/licenses/LICENSE-2.0

Chapter 8

Image 034, Illustration of the project diagram for part 2. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Chapter 9

Image 035, Screenshot from Census Tigerline . **Source**: United States Census Bureau, TIGER/LINE SHAPESFILES, https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

License: Public Domain, https://creativecommons.org/public-domain/pdm/.

Image 036, Screenshot of filter from Census Data Portal. **Source**: United States Census Bureau, Census Data Portal, https://data.census.gov

License: Public Domain, https://creativecommons.org/public-domain/pdm/.

Image 037, OpenRefine logo. **Source**: OpenRefine, https://openrefine.org/.

License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 038, Screenshot from QGIS application. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 039, Screenshot of Census table geoid in Excel app. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 040, Screenshot of the OpenRefine upload screen. **Source:** OpenRefine, https://openrefine.org.

License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 041, Screenshot of the OpenRefine upload screen. **Source**: OpenRefine, https://openrefine.org/.

License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 042, Screenshot of the OpenRefine parsing screen. **Source**: OpenRefine, https://openrefine.org/.
License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 043, Screenshot of the OpenRefine create facet by star action. **Source**: OpenRefine, https://openrefine.org/.
License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 044, Screenshot of the OpenRefine matching rows. **Source**: OpenRefine, https://openrefine.org/.
License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 045, Screenshot of the OpenRefine split columns dialog. **Source**: OpenRefine, https://openrefine.org/.
License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 046, Screenshot of the OpenRefine remove columns dialog. **Source**: OpenRefine, https://openrefine.org/.
License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 047, Screenshot of the OpenRefine convert to number. **Source**: OpenRefine, https://openrefine.org/.
License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Image 048, Screenshot of the OpenRefine convert multiple

columns to number. **Source**: OpenRefine, https://openrefine.org/.
License CC-By-4.0, https://creativecommons.org/licenses/by/4.0/.

Chapter 10

Image 049, Screenshot of the QGIS empty project. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 050, Screenshot of the QGIS tool bar with labeling. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 051, Screenshot of dragging and dropping in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 052, Screenshot of added data in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 053, Screenshot of projected data in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 054, Screenshot of project CRS in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 055, Screenshot of data manager icon in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 056, Screenshot of delimited text icon in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 057, Screenshot of plugin management in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 058, Screenshot of geocoding menu in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 059, Screenshot of manual selection in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 060, Screenshot of exporting selected features in QGIS. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 061, Screenshot of select points within a distance in QGIS. QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 062, Screenshot of attribute table in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 063, Screenshot of added points in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 064, Screenshot identifying LUCODE field in attribute table in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 065, Screenshot of selected features by expression in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 066, Screenshot of LUCODE expression in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Chapter 12

Image 067, Illustration of a 1-1 join.. **Source**: Jennifer Moore. **License:** CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 068, Screenshot of add join button in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 069, Screenshot of successful join in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 070, Screenshot of join by location in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 071, Screenshot of merging data in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 072, Screenshot of adding field parameters in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Image 073, Screenshot of adding standard values to a field in QGIS. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 074, Screenshot of editing menu in QGIS. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 075, Screenshot of select row boxes in table in QGIS. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 076, Screenshot of the clip tool in QGIS. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 077, Screenshot of the add field dialog in QGIS. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 078, Screenshot of the edit vertices tool in QGIS. **Source**: QGIS, https://qgis.org.

License:GNU General Public License, https://qgis.org/license/

Image 079, Screenshot of the Metroline attribute table in QGIS. **Source**: QGIS, https://qgis.org.

License: GNU General Public License, https://qgis.org/license/

Image 080, Screenshot of Metrolines to edit in QGIS.

Source: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 081, Screenshot of the Metroline vertices in edit mode in QGIS. **Source**: QGIS, https://qgis.org.
License:GNU General Public License, https://qgis.org/license/

Chapter 13

Image 082, Screenshot of formatting symbols in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 083, Screenshot of formatting symbols in QGIS.**Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 084, Screenshot of gradated mean income in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 085, Screenshot of expression for normalized race in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 086, Screenshot of median home values by gradated

colors in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 087, Screenshot of owner occuipiers shaded by gradated colors in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 088, SCreenshot of black population shaded by gradated colors in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 089, Screenshot of density plugin in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 090, Screenshot of density toolbar in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 091, Screenshot of kernel density estimation parameters in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 092, Screenshot of kernel density estimation result in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Chapter 15

Image 093, Screenshot of the central point feature result in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Chapter 16

Image 094, Screenshot of formatting all symbols in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 095, Screenshot of catagory symbology in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 096, Screenshot of simple svg marker in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 097, Screenshot of a map with simple svg markers in QGIS. **Source**: QGIS, https://qgis.org.
License: GNU General Public License, https://qgis.org/license/

Image 098, Screenshot of resource sharing plugin in QGIS. **Source**: QGIS, https://qgis.org.

License: GNU General Public License, https://qgis.org/license/

Image 099, Screenshot of svg browser in QGIS. **Source**: QGIS, https://qgis.org.

License: GNU General Public License, https://qgis.org/license/

Image 100, Screenshot of setting path to svg custom icons in QGIS. **Source**: QGIS, https://qgis.org.

License: GNU General Public License, https://qgis.org/license/

Image 101, Screenshot of browsing to custom svg icons in QGIS. **Source**: QGIS, https://qgis.org.

License: GNU General Public License, https://qgis.org/license/

Image 102, Croissants. **Source**: bakery by Gung Yoga, https://thenounproject.com/icon/bakery-7767333/, Noun Project.

License: CC-By-3.0, https://creativecommons.org/licenses/by/3.0

Image 103, Ice Cream. **Source**: ice cream cone with sprinkles by MARADONNA, https://thenounproject.com/icon/ice-cream-cone-with-sprinkles-6874819/, Noun Project.

License: CC-By-3.0, https://creativecommons.org/licenses/by/3.0

Image 104, Juice. **Source**: juice by Rikas Dzihah, https://thenounproject.com/icon/juice-7957614/, Noun Project.

License: CC-By-3.0, https://creativecommons.org/licenses/by/3.0

Image 105, Coffee. **Source**: Coffee by CT, https://thenounproject.com/icon/coffee-59078/, Noun Project.

License: CC-By-3.0, https://creativecommons.org/licenses/by/3.0

Image 106, Image of map 1, focused on income, schools and businesses. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 107, Imge of map 2, focused on housing values, owner occupancy, and race. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 108, Image of map 3, focused on income, floodways, and sinkhole density. **Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

Image 109, Image of map 4, focused on housing values, owner occupancy, and single-family homes.**Source**: Jennifer Moore.

License: CC-By-4.0, https://creativecommons.org/licenses/by/4.0/

SOURCES

Links current as of October 8, 2025.

"2010 TIGER/Line® Shapefiles." n.d. Accessed September 21, 2025. https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2010&layergroup=Census+Tracts.

"B02001: RACE – Census Bureau Table." n.d. Accessed September 21, 2025. https://data.census.gov/table/ACSDT5Y2010.B02001?q=b02001&g=050XX00US29189$1400000&y=2010.

Campbell, Jonathan. n.d. *Essentials of Geographic Information Systems*. 2011. Saylor Foundation. Accessed September 1, 2020. https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/.

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Strong Ideas Series. The MIT Press.

Dorrell, David, and Joseph P. Henderson, eds. 2018. *Introduction to Human Geography*. Blue Ridge: University of North Georgia. https://open.umn.edu/opentextbooks/textbooks/870

"DP04: SELECTED HOUSING CHARACTERISTICS – Census Bureau Table." n.d. Accessed September 21, 2025. https://data.census.gov/table/ACSDP5Y2010.DP04?q=dp04&g=050XX00US29189$1400000&y=2010.

Fang, Yiping. n.d. "Spatial Thinking in Planning Practice: An Introduction to GIS." Open Textbook Library. Accessed August 17, 2020. https://open.umn.edu/opentextbooks/textbooks/234.

Monmonier, Mark S. 2018. *How to Lie with Maps*. Third edition. The University of Chicago Press.

National Research Council, Division on Earth and Life Studies, Board on Earth Sciences and Resources, Geographical Sciences Committee, and Committee on Support for Thinking Spatially The Incorporation of Geographic Information Science Across the K-12 Curriculum. 2006. *Lear ning to Think Spatially: GIS as a Support System in the K-12 Curriculum*. National Academies Press. https://doi.org/10.17226/11019.

Rosenfeld, Christine, and Nathan Burtch. n.d. *Human Geography*. Accessed September 30, 2025. https://viva.pressbooks.pub/humangeog/.

"S1901 – Census Bureau Tables." n.d. Accessed September 21, 2025. https://data.census.gov/table?q=s1901&g=050XX00US29189$1400000&y=2010.

Scott, Dorris. 2024. *Digital Cartography*. WashU Libraries. https://doi.org/10.7936/354r-qa33.

Sinton, Diana Stuart, and Jennifer J. Lund. 2007. *Understanding Place: GIS and Mapping across the Curriculum*. 1st ed. ESRI Press.

Stevens, Joshua, and Raechel Bianchetti. 2012. "4.2 Attribute Measurement Scales." In *Mapping Our Changing World*. The

Pennsylvania State University. https://www.e-education.psu.edu/geog160/c3_p8.html.