Introduction to GIS with QGIS

Frank Donnelly, Head of GIS & Data Services, Brown University Library [1]

April 2024

Contents

Introduction

Frank Donnelly, Head of GIS & Data Services, Brown University Library
frank_donnelly@brown.edu

Last Updated: April 26, 2024 (14th ed.)

Introduction

This tutorial was created to accompany day-long introductory GIS workshops at the Brown University Library, and to serve as a guide for self-directed learning. *Introduction to GIS with QGIS* introduces geographic information systems (GIS) as a concept for envisioning information and as a tool for conducting geographic analyses and creating maps. You will learn how to navigate a GIS interface, how to prepare layers and conduct a basic geographic analysis, and how to create thematic maps.

This tutorial was written using QGIS version 3.34 "Prizren", a free and open source (FOSS) cross-platform (Windows, Mac, Linux) desktop GIS software package. You can download the software and user manual from the QGIS website at `https://www.qgis.org/`. Version 3.34 is the *long term* release; make sure to download this version and NOT the *latest* release. Given differences between versions it's best to use this tutorial with 3.34, although much of the material will apply to all 3.x versions of QGIS. Download the data for the tutorial at `https://libguides.brown.edu/gis_data_tutorials/intro_qgis`. Once you download and unzip the data file, you'll see that the data is separated into different folders for each part of the tutorial.

Objectives

Participants will be able to bring both the tools and the knowledge they gain from this workshop to enhance their projects and the organizations they work for. Specifically, this workshop will enable participants to:

- Add data to GIS software and navigate a GIS interface.
- Perform basic geoprocessing operations for preparing vector GIS data.

- Convert text-based data to a GIS data format.
- Conduct geographic analyses using standard GIS tools and vector data.
- Create thematic maps using the principles of map projections, data classification, symbolization, and cartographic design.
- Locate GIS data on the web and consider the merits of different data sources.
- Demonstrate competency with a specific GIS package (open source QGIS).

Outline

- Chapter 1: General introduction and overview of GIS.
- Chapter 2: Introduction to GIS Interface (learn how to navigate the interface: adding data, layering data, symbolization, changing zoom, viewing attributes, viewing attribute table, making basic selections, difference between data formats, organizing projects and data).
- Chapter 3: GIS Analysis (using a policy example in Rhode Island, attribute table joins, plotting coordinate data, selecting and filtering attributes and features, basic geoprocessing tasks, querying data to make decisions).
- Chapter 4: Thematic mapping (using US states as an example, map projections, coordinate systems, data classification, symbolization, labeling, map layouts).
- Chapter 5: Going Further with GIS (exploring and evaluating online sources for free data, exploring open source and ArcGIS software resources for learning more) .

Organization of this Tutorial

This document is divided into five chapters and subdivided into sections for specific tasks. Each section begins with steps for learning a specific application or process (the what and when), followed by commentary that explains various facets of the process (the how and why). The process and the commentary were separated in order to keep the steps as concise and easy to follow as possible with few digressions; you follow the steps first, and then go back and understand the details of why you followed the steps. The names of certain tools and menus may vary slightly between operating systems, but functionality should be the same.

The following conventions are used throughout:

- Each section begins with steps for learning a specific application or process, followed by commentary that explains various facets of the process.
- Steps are enumerated and begin with *italicized text*.
- A STOP stop sign appears at the conclusion of a series of steps, to clearly delineate where the steps end and the commentary for that section begins.
- Menus, tabs, and items are capitalized if they appear capitalized in the interface.
- Images of toolbar buttons appear in the text whenever they are referenced.
- The names of files and layers appear in sans serif font.
- Urls for websites appear in `typewriter text`.

Changes From Previous Manual

This manual (14th edition) is a revision of the previous edition, and accounts for software updates between QGIS 3.28 Firenze and QGIS 3.34 Prizren. QGIS 3.34 is the latest long term release (LTR), which will be supported for over a year even as newer versions of QGIS are rolled out. The 11th through 14th editions were derived from *Introduction to GIS Using Open Source Software* (10th edition) which was written for a series of workshops at Baruch College CUNY. Beginning with the 11th edition, the manual was retitled "Introduction to GIS with QGIS", examples for New York City were replaced with examples for Rhode Island, and the business-focused site selection analysis in Chapter 3 was replaced with a policy-oriented example.

Specific changes between the 14th and 13th editions:

- Chapter 1 - Minor updates.
- Chapter 2 - Modified commentary in 2.2 to discuss GeoJSON files, changed 2.2.2 heading accordingly.
- Chapter 3 - Adjustments to plotting coordinates in 3.3 based on user interface changes to the CRS library window, added notes about model designer to commentary in 3.6.2.
- Chapter 4 - Modified 3.1 to account for changes in the CRS window, swapped US National Atlas example with Web Mercator as the latter is more widely encountered and better illustrates differences in systems, dropped step about opening .prj files as the shapefile format is becoming less common and reorganized steps in this section. Updated data on UN internet access by country and Natural Earth boundary files for the supplemental exercise at the end of 4.6.
- Chapter 5 - Fixed some broken links, revised section on university and non-profit data sources, added Ocean State Spatial Database as a resource for Rhode Island.
- Appendix - A: dropped Stamen from the web mapping services (they are no longer available), added WMS service for General Bathymetric Chart of the Oceans (GEBCO). C: added the National Land Cover Database classification legend following the ID Codes, referenced it in Chapter 2 raster exercise.

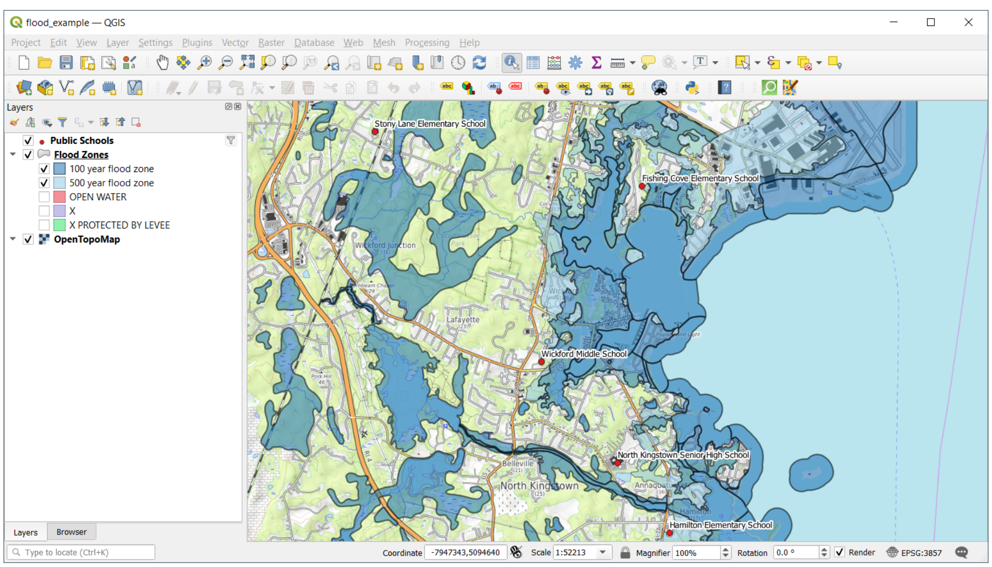
Chapter 1

An Overview of GIS

The goal of this chapter is to provide you with a basic foundation in GIS concepts and software in preparation for the rest of the tutorial.

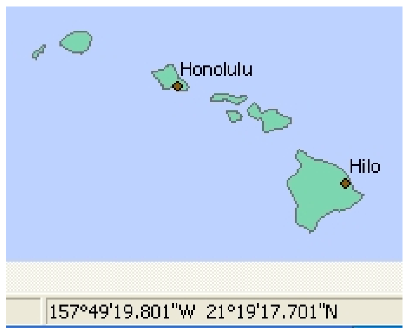
1.1 Basic GIS Concepts

Geographic information systems (GIS) are an integrated collection of software and data used to visualize and organize geographic data, conduct spatial analysis, and create maps and other geospatial information. Narrow definitions of GIS focus on the software and data, while broader definitions include hardware (where the data and software is stored), metadata (data that describes the data), and the people who are part of the system and interact with it as creators, curators, and users. Another definition: GIS is a visual system that organizes information around the concepts of place and location that can be used for geographic analysis, map making, database management, and geospatial statistics. GIS can be applied to virtually any discipline or endeavor.



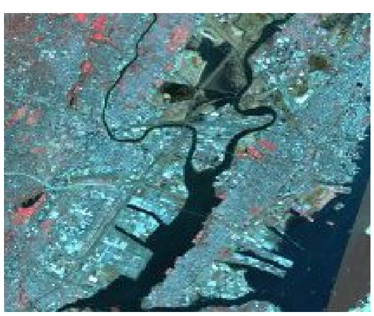
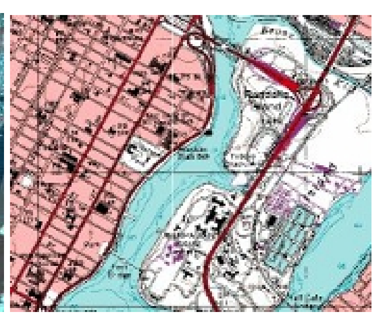
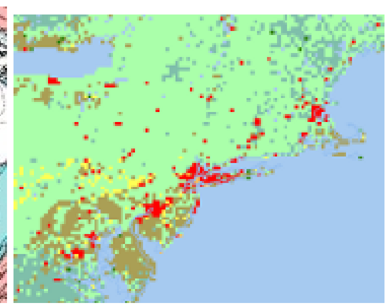
In a GIS, geographic features are represented as individual files or layers that can be added to a project. These features are not maps in and of themselves, but are the raw materials used for map making and analysis. For much of the 20th century cartographers drew geographic features on individual mylar or acetate sheets and layered those sheets over a paper base map to create maps. GIS uses the same principles of layering, with individual files consisting of features that can be layered on top of each other. GIS software acts as an interface, or window, for viewing and manipulating GIS data. The ability to add different layers is quite powerful, as combining the layers allows for analysis that would be impossible if you were viewing single layers by themselves (see example above of topographic map, flood zones, and public schools, illustrating schools vulnerable to flooding).

Each GIS file is georeferenced, meaning that the file is actually tied and related to real locations on the earth. If we mouse over Hawaii in GIS software like in the image below, we see its longitude and latitude coordinates. Just as paper maps were drawn based on map projections and coordinate systems, each GIS file has also been created based on a particular spatial or coordinate reference system, which means that files that share the same reference systems can be overlaid. Since coordinate reference systems are standardized, GIS data can easily be shared. If two files do not share the same system, most GIS packages can convert files from one system to another so they'll match. This distinguishes map making in GIS versus a graphic design package. Maps created in a graphic design package are just simple lines and shapes with no connection to the earth, and the components of the map can't be easily replicated to make other maps. GIS files used to create maps in a GIS package can readily be shared and used to create any map, because they are tied to the earth using standardized systems.



There are different structures or formats for representing geographic data, and these structures are stored in several different file types. Most file types are cross-platform and compatible with different GIS software packages. Major formats and files include:

- Raster - a continuous surface that is divided into grid cells of equal size. Each cell is assigned a particular value and appears as a specific color based on that value. Files in the raster format are similar to digital photos. Common raster objects include air photos, satellite imagery, and paper maps that have been scanned. Raster files can also consist of photos or imagery that have been generalized or have had value added to them to create a new layer, like a land use and land cover layer or a grid showing temperature. There are many different file formats, some common ones include Tiffs (.tif) and JPEGs (.jpg). Unlike regular .tif or .jpg files, GIS raster files are georeferenced.

- Vector - discrete coordinates and surfaces represented as individual points, lines, or polygons (areas). Vector files are built from strings of coordinates and appear more "map-like", and are always abstractions rather than actual images (i.e. shapes to represent boundaries, points to represent cities). The ESRI shapefile (.shp) is a common format. A newer format called a Geopackage (.gpkg) is a simpler alternative for storing a single vector layer. Vector geometry can also be stored in text formats like GeoJSON (.json).

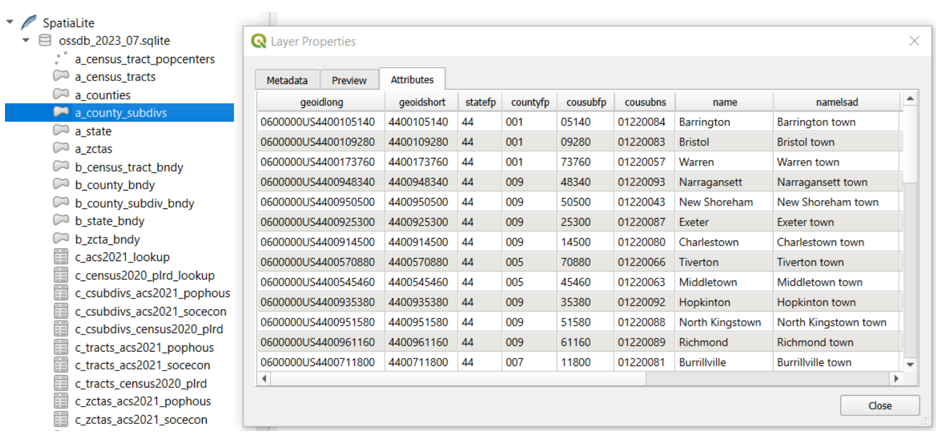
- Tables - data tables that contain records for places can be converted to GIS files and mapped in several ways. If the data contains coordinates like longitude and latitude, it can be plotted and converted to a vector point file. If each data record contains unique ID codes for each place, those records can be joined to their corresponding features in a GIS file and mapped. Tables are commonly stored in text files like .txt or .csv, spreadsheets like Excel or Calc, or as database tables.

	A	B	C	D	E	F
1	ISO	COUNTRY	iu95	iu00	iu05	iu10
2	AFG	Afghanistan			1.22	4.00
3	ALB	Albania	0.01	0.11	6.04	45.00
4	DZA	Algeria	0.00	0.49	5.84	12.50
5	AND	Andorra		10.54	37.61	81.00
6	AGO	Angola		0.11	1.14	10.00
7	ATG	Antigua and Barbuda	2.20	6.48	27.00	47.00
8	ARG	Argentina	0.09	7.04	17.72	45.00
9	ARM	Armenia	0.05	1.30	5.25	25.00
10	ABW	Aruba		15.44	25.40	62.00
11	AUS	Australia	2.76	46.76	63.00	76.00
12	AUT	Austria	1.89	33.73	58.00	75.17

un meta

READY

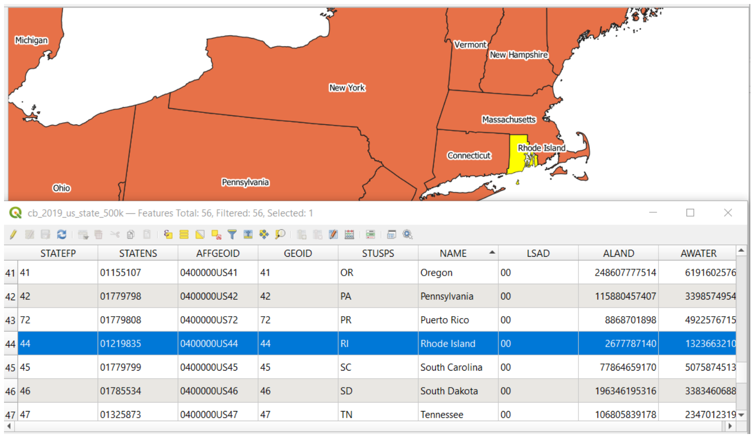
- Geodatabases - containers that can hold related raster, vector, and tabular data. They are good for consolidating and organizing data, and many can be used for spatial queries and analysis. Geodatabases / spatial databases can be desktop (ESRI file geodatabases .gdb, Spatialite files .sqlite) or server based (PostGIS, Oracle).



geoidlong	geoidshort	statefp	countyfp	cousubfp	cousubns	name	namelsad
0600000US4400105140	4400105140	44	001	05140	01220084	Barrington	Barrington town
0600000US4400109280	4400109280	44	001	09280	01220083	Bristol	Bristol town
0600000US4400173760	4400173760	44	001	73760	01220057	Warren	Warren town
0600000US4400948340	4400948340	44	009	48340	01220093	Narragansett	Narragansett town
0600000US4400950500	4400950500	44	009	50500	01220043	New Shoreham	New Shoreham town
0600000US4400925300	4400925300	44	009	25300	01220087	Exeter	Exeter town
0600000US4400914500	4400914500	44	009	14500	01220080	Charlestown	Charlestown town
0600000US4400570880	4400570880	44	005	70880	01220066	Tiverton	Tiverton town
0600000US4400545460	4400545460	44	005	45460	01220063	Middletown	Middletown town
0600000US4400935380	4400935380	44	009	35380	01220092	Hopkinton	Hopkinton town
0600000US4400951580	4400951580	44	009	51580	01220088	North Kingstown	North Kingstown town
0600000US4400961160	4400961160	44	009	61160	01220089	Richmond	Richmond town
0600000US4400711800	4400711800	44	007	11800	01220081	Burrillville	Burrillville town

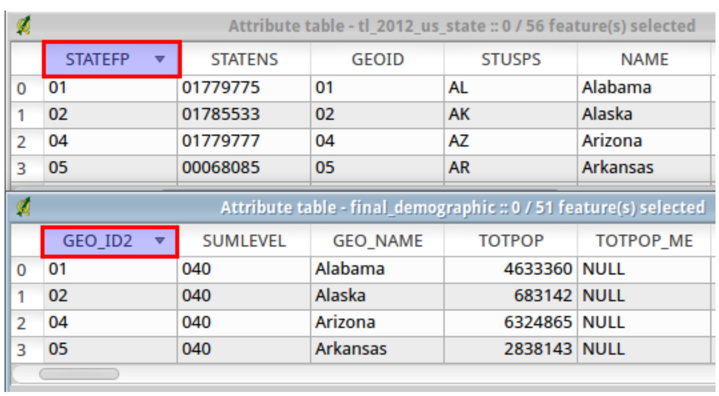
- Web Services - GIS can tap into files stored on the web that are published using a variety of services. Instead of downloading, users can connect to web-based layers and render them directly in GIS. Layers are rendered using Web Mapping Services (WMS) which render layers as rasters, and as Web Feature Services (WFS) which render layers as vectors and allow for the display and manipulation of attributes. Web Map Tile Services (WMTS) and XYZ Tiles are popular open-standards for delivering raster maps as a series of individual tiles that appear seamless. This method renders rasters quickly and more efficiently than WMS.

Raster and vector GIS files exist spatially, in that you can see the grid or shapes and their corresponding location on the earth. Vectors also have a tabular component that is particularly valuable. For example, every feature in a vector file of state boundaries has an associated record that's attached to it and stored in a table for those states. This attribute table contains columns or fields that store values for each state, such as the state's name, values like population or area that describe it, and ID codes that uniquely identify each one. The names can be used by the GIS to label each state, and the values like population can be thematically mapped.



	STATEFP	STATENS	AFFGEOID	GEOID	STUSPS	NAME	LSAD	ALAND	AWATER
41	41	01155107	0400000US41	41	OR	Oregon	00	248607777514	6191602576
42	42	01779798	0400000US42	42	PA	Pennsylvania	00	115880457407	3398574954
43	72	01779808	0400000US72	72	PR	Puerto Rico	00	8868701898	4922576715
44	44	01219835	0400000US44	44	RI	Rhode Island	00	2677787140	1323663210
45	45	01779799	0400000US45	45	SC	South Carolina	00	77864659170	5075874513
46	46	01785534	0400000US46	46	SD	South Dakota	00	196346195316	3383460688
47	47	01325873	0400000US47	47	TN	Tennessee	00	106805839178	2347012319

The ID codes for each state can be used to join the attribute table for the GIS file to a tabular file that contains state-level data. For example, a GIS file of state boundaries with a state code can be joined within GIS using relational database techniques to a text or spreadsheet file that has state-level data and that uses the same codes to identify each state. The data in the table, which was just a regular table with no geospatial geometry, can now be visualized and mapped in GIS. There are a number of standard ID codes that can be used for joining data; two common systems are ANSI / FIPS (created by the US government to identify geographic features in the US) and ISO (created by the International Standards Organization to identify countries and their subdivisions).

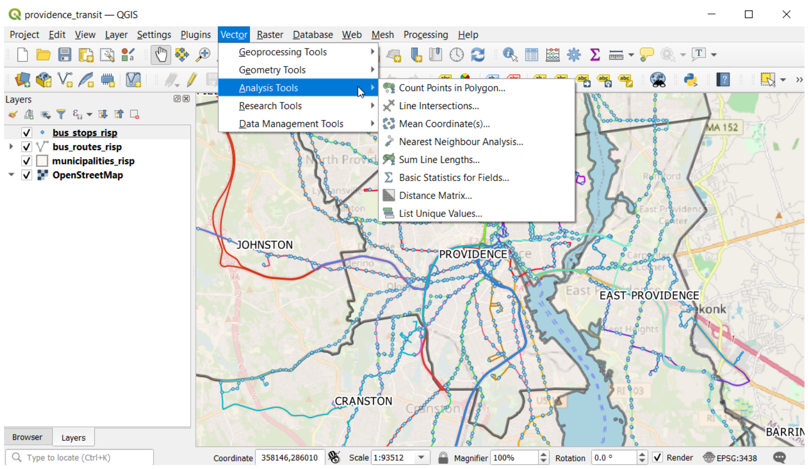


	STATEFP	STATENS	GEOID	STUSPS	NAME
0	01	01779775	01	AL	Alabama
1	02	01785533	02	AK	Alaska
2	04	01779777	04	AZ	Arizona
3	05	00068085	05	AR	Arkansas

	GEO_ID2	SUMLEVEL	GEO_NAME	TOTPOP	TOTPOP_ME
0	01	040	Alabama	4633360	NULL
1	02	040	Alaska	683142	NULL
2	04	040	Arizona	6324865	NULL
3	05	040	Arkansas	2838143	NULL

1.2 GIS Software

A standard interface for GIS software has evolved over time. Typically, GIS software has a data view that consists of a layers menu that lists files that have been added to a project, a data window that displays the GIS files, and a set of toolbars and menus for accessing various tools and launching various processes. Re-ordering the layers in the layers menu changes their drawing order, and right or left clicking on a layer in the menu will reveal individual properties for that particular layer. You can also access the attribute table of the layer and a symbol tab for changing how the features are depicted or classified. There are several tools for zooming in and out to examine different layers and to change the extent of the view.



Coordinate reference systems (CRS) are managed differently in various GIS software packages. In general, the options are: define the CRS for the project before adding the files, or the project automatically takes the system of the first file added. If you try to add GIS files that have different systems, most packages will try to re-project the data on the fly, while others will simply fail to draw the new layers. Even if the software can correctly draw a layer without the user defining it, or even if it can re-project layers on the fly, users will run into problems later on when trying to manipulate the GIS files. You should always be sure to specify the reference system properly and make sure that all files share the same one - GIS gives you the ability to re-project data.

GIS software provides users with a variety of ways for querying geographic data, either by selecting records in the attribute table or shapes in the view, or by conducting searches where you build queries to high-light features that contain specific attributes, or that have some relationship with another geographic layer. GIS software comes with a variety of editing tools that allow you to modify the geometry of GIS files. For example, you can merge features together, break them apart, or clip out or select certain areas to create new files. Collectively these processes are known as geoprocessing. You geoprocess layers in order to prepare raw data for analysis, to create new layers or data, to simplify layers for cartographic or aesthetic purposes, and to perform certain kinds of analyses. GIS also provides the ability to edit files on a feature by feature basis, and most packages allow users to write scripts or build models to automate processes.

Most GIS programs have a separate map layout or print layout, where the user can create finished maps with standard map elements like titles, legends, scale bars, north arrows, and accompanying text. Finished maps can be exported out of the GIS as static files, such as pdfs or jpgs.

Users can always save their GIS projects in a GIS project file. The scale and extent of the data view, symbolization and classification assigned to layers, map layouts, and links to GIS files used in the project are stored in the file. It's important to understand that GIS data files are NOT stored inside the project file - the GIS data and the GIS project file exist independently. When adding data to a GIS, you are establishing a link from the GIS project to the GIS data. Furthermore, changing the colors of the features or classifying them in a certain way has no effect on the actual GIS data files themselves. When you change symbols, you are only changing how the GIS program views the data - you're not modifying the actual data.

Essentially, the GIS software acts as a window for viewing and working with GIS data, which is stored outside the window. The GIS project file stores the window dressing of scale and symbolization. You never actually change the GIS data unless you go into an edit mode or conduct an operation that creates a new GIS file. This relationship is of crucial importance when it comes time to move or share files - if you move your project file or your data, the links between them will become broken, and you'll need to re-establish the location between the project and the data in order to repair your project file. Unlike data files, project files are only compatible with the GIS software package for which they were designed.

1.3 Open Source

In this tutorial we will be using QGIS, which is free open source software (FOSS). Open source software is an alternative to proprietary software:

- Open source software is free; you don't have to purchase it and you can freely distribute it to anyone else, as opposed to proprietary software which you must purchase and typically can not share with anyone (since it's copyrighted).
- The source code, or actual computer programming, that was used to create the software is transparent, as opposed to proprietary software where the code is hidden and encrypted.
- Under the open source model the programming code is transparent and you are free to change and make improvements to it; this is strictly prohibited with proprietary software.

Open source software can be created in several ways. A programmer or developer creates software from scratch, because they have some need that isn't being met by current software. Over time, as other programmers discover the project they may choose to contribute to building or improving this software, and they rally around the creator and begin to form a group that becomes devoted to the project. The Linux operating system and the Python programming languages essentially began this way. Alternatively, a group of people who receive support from a business or entrepreneurs take software that was formerly proprietary but is no longer commercially viable, and they build on this product and re-release it as open source. The Mozilla Firefox browser (formerly the proprietary Netscape) and LibreOffice (formerly the proprietary Star Office and a branch of OpenOffice) are examples of the latter.

Why do people create FOSS software?

- It gives programmers a chance to practice their skills.
- It gives programmers a way to build thier professional reputation within different programming circles.
- Open source is an ethos for some, who believe that software and information should be free.
- Some see it as a superior model - since the code is open, there is a better chance that improvements can be made more quickly and that bugs can be discovered more easily than in proprietary software, as it's scrutinized by an entire community.
- Businesses may prefer it because it does not tie them to costly, proprietary software that may go out of date or out of business - with open source there is always someone who can take over a project and keep it going, since the code is free and transparent.
- If proprietary software for a certain application is inefficient, insufficient, expensive, or non-existent, FOSS software can be created to meet the need.

The Open Source Geospatial Foundation (OSGEO) was created to support the collaborative development of FOSS GIS and promote its use (`http://www.osgeo.org/`). Dozens of packages exist; in this tutorial we will be using QGIS, which was initially developed by a group of volunteers in 2002 as a simple GIS viewer but has evolved into one of the premier FOSS GIS packages.

The advantages of using QGIS: it's free, you can download it, it runs on any operating system, it is mature enough that it supports all essential GIS tasks plus many advanced ones, and it's relatively easy to use. Many of it's early disadvantages, such as lack of documentation and name recognition, and the shortage of more advanced features, have dissipated as the software has become increasingly stable and sophisticated over time.

Open software tends to be modular rather than monolithic; you often have several, independent software applications to perform different functions, rather than one, large piece of software that does it all. A typical FOSS GIS workstation may include several applications like QGIS (for viewing data, basic analyses, map making, working with vector and raster data), GRASS (another FOSS package that's good for advanced raster modeling), Python (for data processing and creating customized tools), and a geodatabase application (PostGIS for server-based databases and Spatialite / SQLite for desktop use). Individual FOSS software will often contain a large group of core components as well as a number of plug-ins that were subsequently designed to add new functionality. Plug-ins may be created by the developers or by third parties, and over time can be incorporated as core functions in later versions of the software.

ArcGIS, created by a company called ESRI, has been on the market for several decades and is the dominant, proprietary (non-FOSS) GIS software on the market. It's used by many government agencies and universities. Since it is rather expensive to purchase for individual use, you tend to see it more often in institutional settings. If you are affiliated with a college or university, chances are you'll be able to access it somewhere on your campus. ESRI does distribute trial versions of the software for education and home use. Making the transition from a proprietary package to FOSS is relatively straightforward as most GIS software operate under the same properties and principles and share similar user interfaces. Best of all, many of the common raster and vector data formats are cross-platform and can be used in any GIS software package.

Chapter 2

Exploring the Interface

The goal of this chapter is to familiarize you with the interface and basic features of GIS in general and QGIS in particular. You'll explore the various features of vector and raster layers, and will add and configure some layers that you'll use later in Chapter 3.

2.1 The QGIS Interface

This section will introduce you to the QGIS interface; you will configure the interface in preparation for the rest of this tutorial. The GIS layers we'll use come from the US Census Bureau and Ocean State Spatial Database Project, and have been prepped for this tutorial so that they share a common coordinate reference system. Download and install QGIS 3.34 Firenze, and download the sample data and unzip the archive file before you begin.

2.1.1 Steps

1. *Launch QGIS.* (If you're using MS Windows, look under the Start > Programs > QGIS3 or QGIS Desktop 3.34).

2. *Configure plugins*. Go to Plugins > Manage and Install Plugins. Click on the Installed Plugins button in the menu on the left. Keep the plugins for DB Manager and Processing checked, and un-check the others. Note that this menu is where you can find core and third-party plugins for QGIS. Hit the Close button when you're finished.

3. *Configure the panels and toolbars*. Likewise, we're going to turn off toolbars that we won't need. Right click on a blank area of the toolbar to get the panels and toolbar view menu. In the Panels section, make sure Browser and Layers is checked. In the toolbars section, these eight toolbars should be checked: Attributes, Data Source Manager, Digitizing, Help, Label, Map Navigation, Project, and Selection. If other panels or toolbars are checked, uncheck them. Every time you check or uncheck a feature the menu will disappear, so you will need to right click on a blank area of the toolbar to get it back.

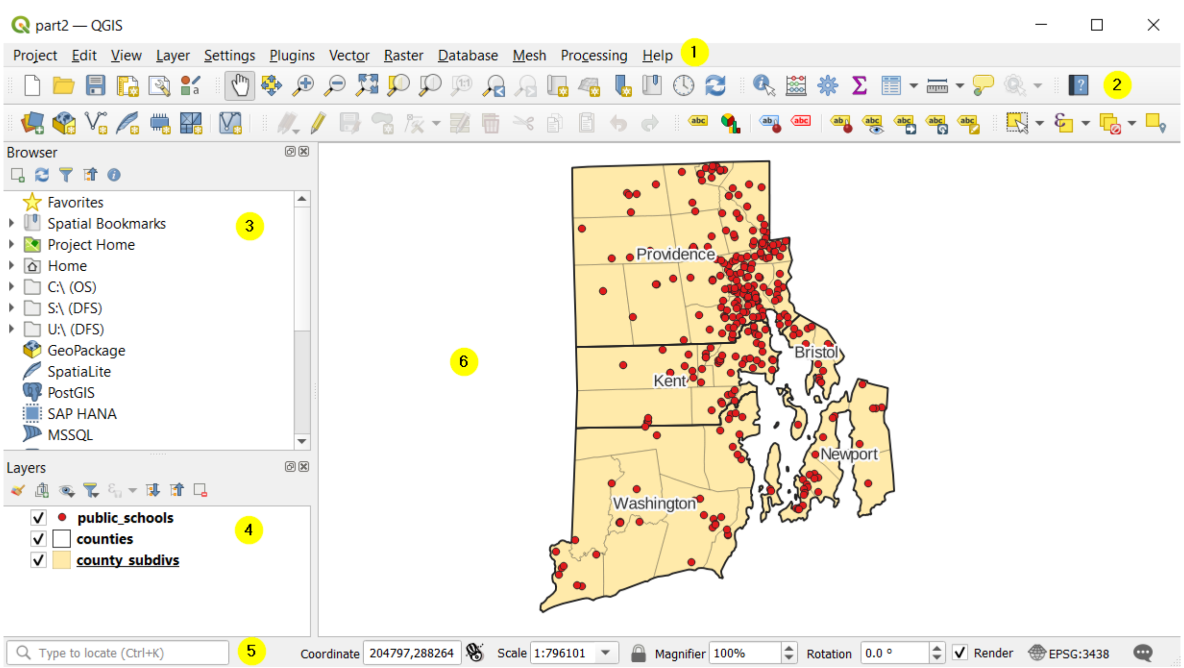
4. *Move toolbars*. Move the toolbars around by hovering over the left edge of a toolbar until you see a crosshairs, left click and hold, then drag and drop. Configure the toolbars to your liking (suggestion: try aligning them so you have only two rows of them at the top of the screen and all buttons are visible).

5. *Un-stack the panels*. To keep our interface less cluttered, we're going to modify the configuration of our Browser and Layers panel from a stacked view (where they appear together, one on top of the other) to a tabbed view (where only one is visible at a time). At the top of the Browser, left-click on the 'Browser' title and drag the box down over the Layers panel, and release. The Browser will now occupy the entire space. At the bottom of the screen you'll now see two tabs, where you can switch back and forth between the Browser and the Layers.

2.1.2 Commentary

Interface Components

1. *Menu Bar*: provides access to various features and functions of the software using standard dropdown menus. The location of the menus and items is largely fixed, although if you activate certain plugins they may add additional items to a dropdown.
2. *Toolbar*: replicates many of the features and functions in the Menu Bar, providing access to common features in a single click. The location of the toolbars is not fixed; if you hover over the edge of the toolbar and hold down the left mouse button you can drag and dock the toolbar wherever you like (this means that the location of tools on your screen may not match those of other screens, or this tutorial).
3. *Browser Panel*: the browser allows you to see your file system, database connections, and web services, and lets you drag files from your file system into your project. By default the Browser initially occupies this space in the interface, but there are a number of other features you can enable that can share or occupy this area; you can also switch the location of the Browser with the Layers panel.
4. *Layers Panel*: a list of the map layers that are part of your current project. You can check or uncheck layers to turn them on and off, drag them to change the drawing order, select one in order to perform specific tasks on that layer, and right click on a layer to access menus and tools for working with that specific layer. Like the Browser, the Layers panel can be stacked or you can enable tabbed viewing to see just one at a time.
5. *Status Bar*: shows the current scale of the map view, the coordinates of the current position of the cursor, and the coordinate reference system (CRS) used by the project.
6. *Map Canvas*: geographic display that shows all of your active layers. Also known as the Map View. When you launch QGIS, this area is populated by other panels: News, Recent Projects, and Templates. As soon as you add data or open a project, these panels will disappear.



- *Want to turn a panel or toolbar off? Wondering where a panel or toolbar went?* If you right click on a blank area of either the Menu Bar or the Toolbar, you'll get a list that shows all of the panels and toolbars. You can check and uncheck items to turn them on and off.
- *Can't figure out what a button means or does?* If you hover over a button, a small window appears that displays the name of the button.

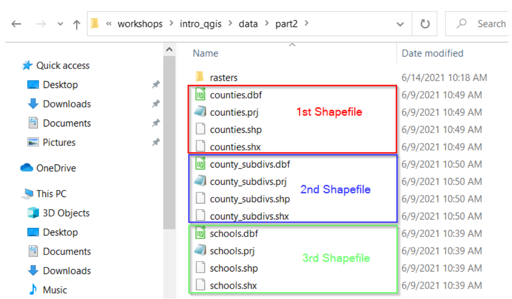
- *Are there hotkeys?* Most menu items and tools can also be accessed by using hotkeys or keyboard shortcuts (for example, CTRL S will save the current project). For a full list of hotkeys, view the QGIS manual. Many of the common Windows shortcuts (like CTRL C for copy and CTRL V for paste) will work in QGIS.
- *Where is the QGIS manual?* It's available on the QGIS website at `https://www.qgis.org/en/docs/` You can click the help contents button to go there directly.

2.2 Adding Vector Data

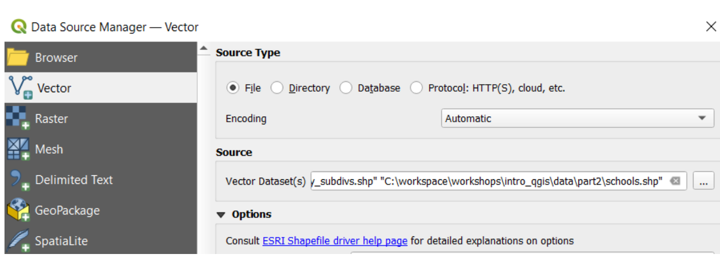
In this section you'll learn how to add vector GIS files to QGIS and to symbolize them. Shapefiles are a common GIS data format that you'll routinely encounter in your future work.

2.2.1 Steps

1. *Examine your data.* Minimize QGIS for a moment and take a look at the data files under the data folder for part 2 in your operating system's file browser or window. These are shapefiles that we will add to QGIS and work with for this project. There are three shapefiles; each one is composed of multiple files that have the same names but different extensions.

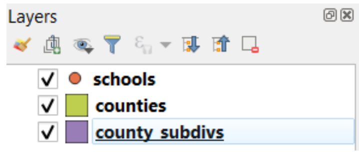


2. *Add the three shapefiles.* Maximize QGIS to return to the program. On the Tool Bar, hit the Data Source Manager button. When the Manager appears, select the Vector button in the list on the left. Under Source, hit the ellipsis button and browse through the folder list to the data folder for part 2. In the Files of Type dropdown in the lower right-hand corner of the window, make sure the ESRI shapefiles option is selected. While holding down the control key (command key if you're using a Mac) select the three shapefiles ending with the extension .shp. Hit Open to add them. This returns you to the Manager. Click Add, and then Close the Manager (if you're prompted to transform files, just hit OK and the operation will proceed; we'll return to coordinate systems later in the tutorial). Your layers should appear in the Layers Panel and Map View (make sure that you tab from the Browser to the Layers view so you can see them).

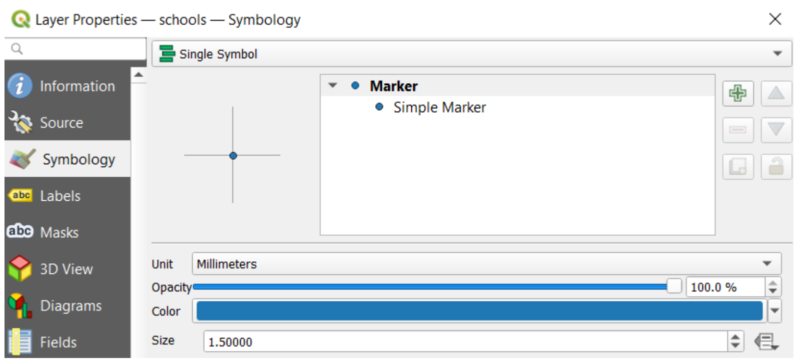


3. *Experiment with changing the drawing order.* Click on the first layer that's listed in the Layers Panel (LP), hold down the left mouse button, and drag it to the bottom of the list. This moves that layer from the top of the drawing order to the bottom; layers in the Layers Panel (LP) are stacked on top of each other, and their order in the list determines which are visible relative to others. Experiment with changing the order of the layers and observe what happens.

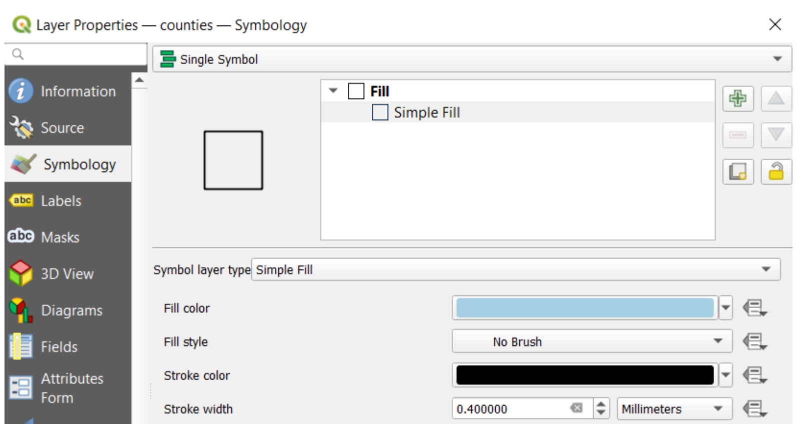
4. *Order the layers.* Drag the layers in the Layers Panel (LP) so they appear in this order, from top to bottom: schools, counties, and county_subdivs (cities and towns).



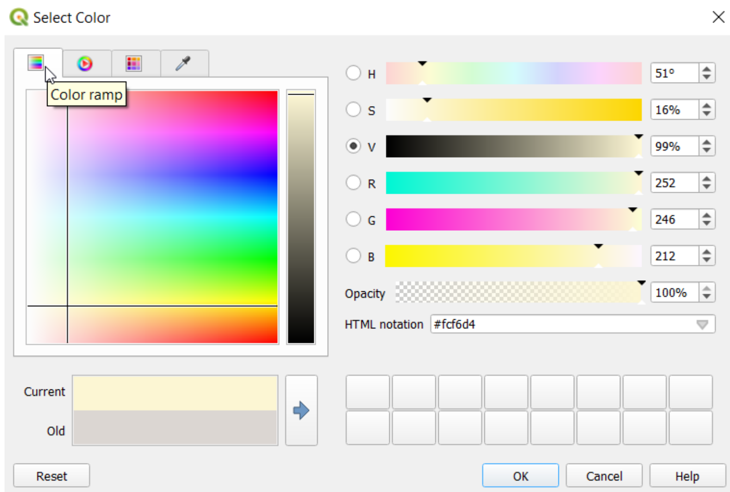
5. *Change the color and size for the schools.* Double-click on the schools layer in the LP to open the Layer Properties menu for that layer (alternatively, you can right click on the layer and select Properties). Click on the Symbology tab. Click the drop down menu beside the Color box. Change the color to blue by choosing from the palette of Standard Colors. In the size box change the value from 2.0 to 1.5. Click OK on the Symbology menu to make the change.



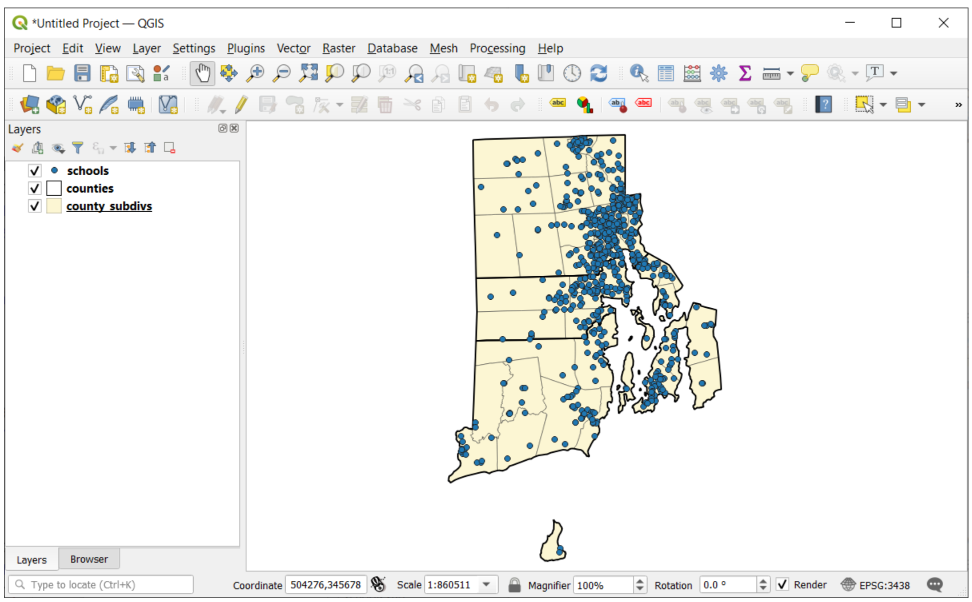
6. *Give the counties no fill.* (i.e. make them hollow with no color). We can't see the county subdivisions and counties simultaneously because they both are filled in and one covers the other. Double-click on the counties layer in the LP to open the Layer Properties menu for that layer. Click on the Symbology tab. Select the Simple fill box at the top. This modifies the options you'll see on the bottom. Change the Fill style dropdown from Solid to No Brush. In the Stroke Width box, change the value from .26 to .40. Hit OK.



7. *Change the colors for the county subdivisions.* Double-click on the county_subdivs layer in the LP to open the Layer Properties menu for that layer. This time we'll choose from a broader color palette. Click directly on the color box itself to open the palette options. The Color Ramp tab is selected by default. Click around in the palette, adjust the hue slider on the right, and select a light shade of either brown, yellow, or grey (or if you prefer, you can click the Color Wheel tab and select there). Click OK in the color selector. Click the Simple Fill option at top, and change the stroke width from 0.26 to .10. Click OK in the symbology tab to accept the changes and close the window.



8. *Verify symbolization.* After completing these steps, your QGIS window should resemble the image below.



2.2.2 Commentary

Shapefiles, Geopackages, and GeoJSON

A shapefile is a common legacy file format used for storing vector GIS data. It was created by ESRI, the company that produces ArcGIS (the predominant software in the proprietary GIS market). Shapefiles are an open format that can be used in just about any GIS software package. A shapefile can consist of point, line, or polygon features for a given geographic area, and can never consist of multiple types of geometry (i.e. you can't have a shapefile with points and lines). Polygon features can be single-part (where every individual polygon is an individual feature) or multi-part (where multiple polygons can be grouped together as single features).

Despite it's singular sounding name, a shapefile consists of several individual files. The following three pieces are mandatory:

- .shp file - shape file, contains the geometry
- .shx file - shape index file, an index of the geometry
- .dbf file - attribute file, contains attributes for the features

The following pieces are typically (ideally) included:

- .cpg file - a plain text file that contains the encoding for the features
- .prj file - a plain text file that contains the projection and coordinate system
- .sbn and .sbx files - spatial index of the features
- .shp.xml file - XML metadata

It is important that all of the pieces of the shapefile are kept together in the same folder, otherwise the file will not work - so be careful when moving files around! Renaming files is problematic - if you rename one you must rename all of them with the same name, otherwise they won't function together. Instead of using your computer's file explorer, use the Browser panel, right click on a file, and choose Manage - Rename.

Geopackages (.gkpg) are a newer format. Built on a SQLite database, they were designed to provide a concise, flexible, and efficient way for storing either a vector or raster layer in a single file while avoiding the legacy issues associated with shapefiles. Geopackage is the default format when saving or exporting vectors in QGIS, but you can change the default if you prefer. Many data providers continue to provide data in the shapefile format, and thus in the first few chapters we'll continue to use it as you're likely to encounter it.

GeoJSON files (.geojson or .json) are increasingly common, with data saved in a plain text format that uses nesting and key value pairs. While QGIS can render these files, they usually can't be used in tools or operations because they are not structured in a tabular / gridded format. To use them, convert them to shapefiles or geopackages.

Adding Data and Drawing Order

When you add map layers or data to a map view, you are technically not adding data to the window, i.e. copying the file and inserting it into the project. Rather, you are establishing a link between the GIS interface and the files, which exist independently from the software. When you use GIS software to change the symbolization of the layers (colors, outline, labels, etc) you are not modifying the data file itself; you are simply telling the software to display the layers in a certain way. The software is essentially a window for viewing the data files. The only way to change the data files themselves (their geometry or attributes) is within an editing mode which you must specifically launch.

For much of the 20th century maps were created by taking individual layers on translucent mylar sheets and laying them over top of a paper base map. For example, an outline of the United States with boundaries of each state could serve as a paper base map, with individual mylar sheets layered on top that had rivers and cities. The order of the sheets determined which features appeared on top, covering up other features. GIS functions the same way; the order of the layers determines which appear on top. If you move a polygon layer with a solid fill (i.e. counties) over top of a point layer (i.e. schools), you will not see the schools as the county layer is covering it up. In order to show both layers, you would have to move the schools layer on top of the counties. Alternatively, you could make the counties layer hollow by removing the fill, which would allow the schools layer to be visible if it was on the bottom.

You would typically use a hollow fill for a polygon if you wanted to display its boundaries on top of another polygon layer that has a fill, as we did with the counties and county subdivisions in this example; we placed the counties on top and made them hollow to see county subdivisions underneath, and made the county boundaries thicker to make them stand out from the subdivisions. Where the boundaries coincide, the county boundaries cover up the thinner subdivision boundaries. Alternatively, if we wanted to see both sets of boundaries, we could have made the subdivisions hollow and placed them above the counties, leaving the counties solidly filled, and changed the color and thickness of each set of boundaries to make them visible.

2.3 Exploring the Map View

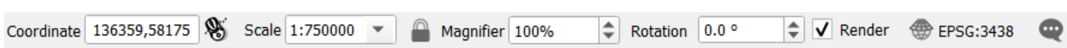
In this section you'll learn how to navigate the map view.

2.3.1 Steps

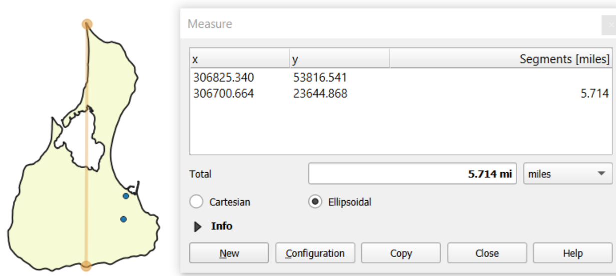
1. *Experiment with the Zoom tools.* Try each of the zoom tools in the Menu Bar (if you ever get lost, hit Zoom Full).

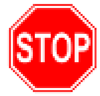
   - Pan - move around the map by holding the left mouse button down and drag (does not change the zoom)
   - Pan Map to Selection - move map to selected features without changing the zoom (skip this one for now)
   - Zoom In - click to zoom in once, draw a box to zoom in to an area, or use the mouse wheel
   - Zoom Out - works the same as the Zoom In tool
   - Zoom Full - will zoom the window to the maximum extent of all visible layers
   - Zoom to Selection - zooms to selected features (skip this one for now)
   - Zoom to Layer - zooms to the maximum extent of the feature currently selected in the LP
   - Zoom to Native Pixel Resolution - will zoom to the optimal scale for rasters (skip this one for now)
   - Zoom last - returns to your previous zoom
   - Zoom next - moves you forward to your next zoom (if you've already used zoom last)
   - Refresh - redraws the screen (useful if your layers didn't draw completely or properly)

2. *Notice change in coordinates.* Move the cursor around the map. In the Status Bar (below the Map View) notice how the coordinates change based on the position of the cursor. If you hover over the box that says EPSG: 3438, a pop-up window tells us the coordinate reference system (CRS) of the project based on our layers is NAD83 / Rhode Island (ftUS). This is the local state plane system that is appropriate for our area. In the state plane system coordinates are measured in feet. We'll cover coordinate reference systems later in the tutorial. The scale box can also be used to change the zoom (a higher number to zoom out and a lower number to zoom in).



3. *Measure some distances.* Use the zoom tools to center the map view on Block Island, which is the island at the bottom of the window that's most distant from the rest of the state. Select the measuring distance tool in the toolbar, which opens the Measure window. Change the unit of measurement to miles or kilometers. Hover over the map and you'll notice that crosshairs will appear. Click on the northern tip of the island. Drag the crosshairs to the southern tip of the island in a straight line. As you do this, you'll see a line is drawn from the original point you clicked and the measurement window will update with distances. If you left click on the southern tip of the island, it will lock the line segment and allow you to draw a second segment from that point; or if you right click it will stop measuring at that point. Hit the Close button when you've finished experimenting (Note - to measure distances between many features, use the Vector > Analysis > Distance Matrix tool).





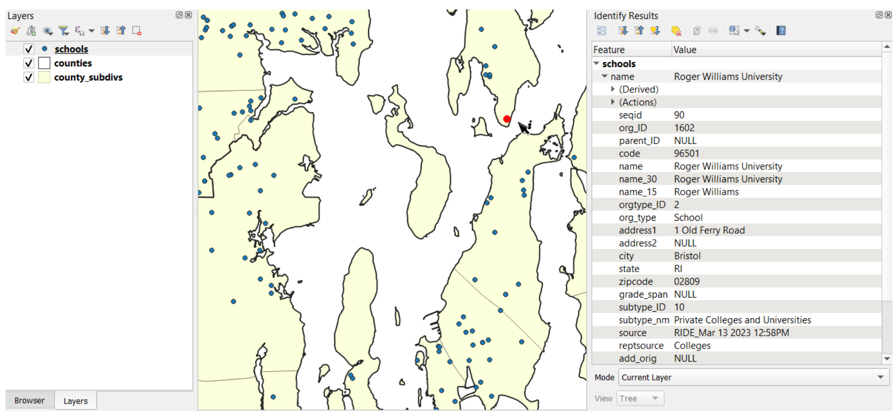
2.4 Exploring Features

In this section you'll learn how to explore and interact with features in the Map View and Attribute table.

2.4.1 Steps

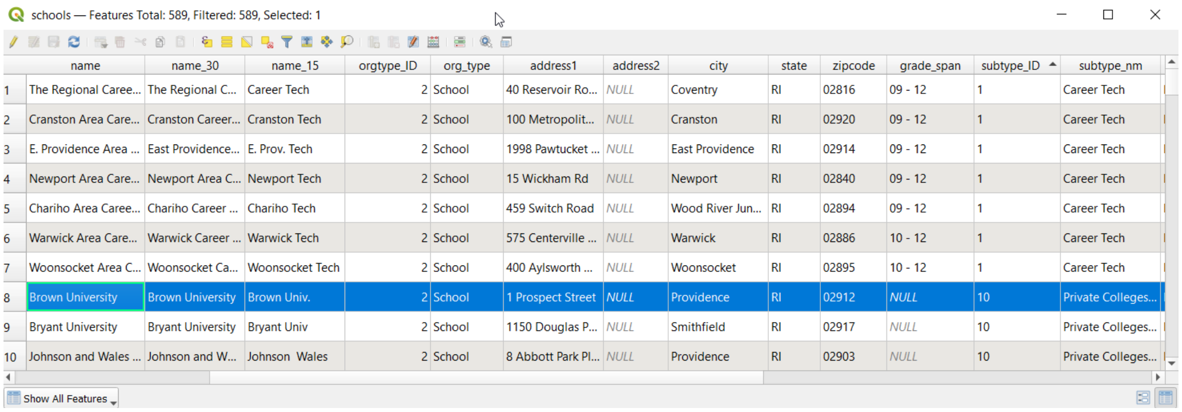
1. *Identify features.* Hit the Zoom to full button to return to the full extent of our features. Then, hit the identify features button in the toolbar. Select the county_subdivs layer in the LP to make it active. Click on a town, and that town becomes highlighted and information about that feature is displayed in a new Identify panel. You can adjust the width of the Feature and Value columns in the panel to see the labels more clearly, and can drag the panel to un-dock it if you wish. Click on another town to change the selection.

2. *Identify features from a different layer.* Make the schools layer the active layer by selecting it in the LP. Click on any school in the map view to get information about that school. If you end up selecting multiple schools, the total number selected will appear beside the layer name at the top of the Identify panel, and you can cycle through each one. To precisely select just one school, zoom into the map and try selecting again. Where is this information coming from?



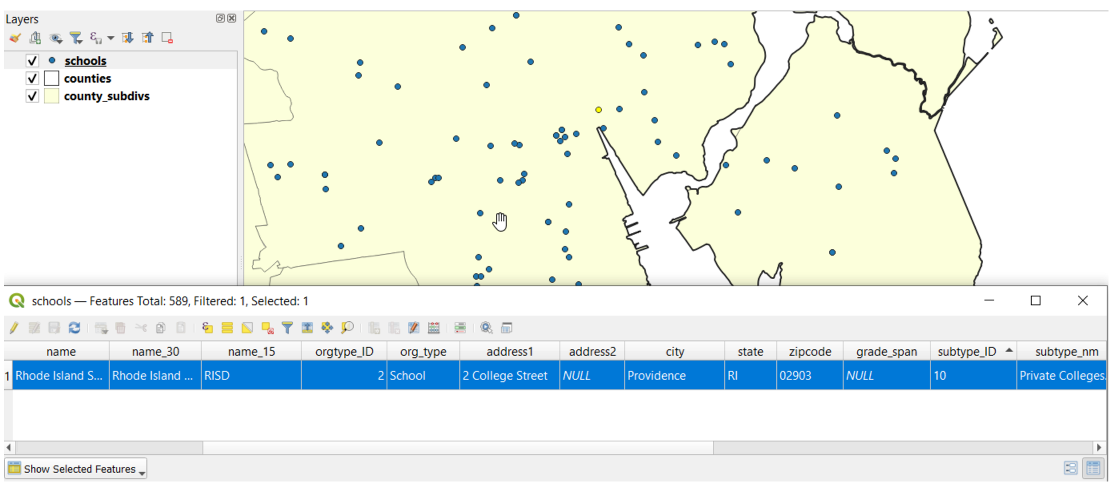
3. *Open the attribute table.* Close the Identify panel, and with the schools layer still selected in the LP, right click on the layer and select Open attribute table (alternatively, you could click the open attribute table button on the toolbar). For every school (feature) in the layer, there is a record for the school in the attribute table of that layer. Explore the table by scrolling across it and down.

4. *Select a feature from the table.* Sort the table by clicking on the field (column) heading that indicates the type of school (subtype_ID). Click on the record for Brown University in the table. Close the attribute table. Hit the Zoom to selected button and you'll see it is selected in yellow; zoom out a little to gain some geographic context, to see the boundary for Providence (Note - you can select multiple records from the table by holding down the CTRL key and selecting records one by one, or select a range by selecting a record, hold the SHIFT key, and select the last record).



	name	name_30	name_15	orgtype_ID	org_type	address1	address2	city	state	zipcode	grade_span	subtype_ID	subtype_nm
1	The Regional Caree...	The Regional C...	Career Tech	2	School	40 Reservoir Ro...	NULL	Coventry	RI	02816	09 - 12	1	Career Tech
2	Cranston Area Care...	Cranston Career...	Cranston Tech	2	School	100 Metropolit...	NULL	Cranston	RI	02920	09 - 12	1	Career Tech
3	E. Providence Area ...	East Providence...	E. Prov. Tech	2	School	1998 Pawtucket ...	NULL	East Providence	RI	02914	09 - 12	1	Career Tech
4	Newport Area Care...	Newport Area C...	Newport Tech	2	School	15 Wickham Rd	NULL	Newport	RI	02840	09 - 12	1	Career Tech
5	Chariho Area Caree...	Chariho Career ...	Chariho Tech	2	School	459 Switch Road	NULL	Wood River Jun...	RI	02894	09 - 12	1	Career Tech
6	Warwick Area Care...	Warwick Career ...	Warwick Tech	2	School	575 Centerville ...	NULL	Warwick	RI	02886	10 - 12	1	Career Tech
7	Woonsocket Area C...	Woonsocket Ca...	Woonsocket Tech	2	School	400 Aylsworth ...	NULL	Woonsocket	RI	02895	10 - 12	1	Career Tech
8	Brown University	Brown University	Brown Univ.	2	School	1 Prospect Street	NULL	Providence	RI	02912	NULL	10	Private Colleges...
9	Bryant University	Bryant University	Bryant Univ	2	School	1150 Douglas P...	NULL	Smithfield	RI	02917	NULL	10	Private Colleges...
10	Johnson and Wales ...	Johnson and W...	Johnson Wales	2	School	8 Abbott Park Pl...	NULL	Providence	RI	02903	NULL	10	Private Colleges...

5. *Select a feature from the map.* With the schools layer still selected in the LP, hit the select feature button in the toolbar. Select the school that is just to the west (left) of Brown University. Hit the open attribute table button. At the bottom of the table, hit the filter drop down menu and choose the option to Show Selected Features. This reveals the record for the RI School of Design; this is the school that you've selected in the Map View. These two steps demonstrate that the table and map are linked, and you can select features in one and display them in the other. (Note - you can select multiple features in the map by holding down the CTRL key and clicking on features one by one, or by hitting the dropdown beside the select feature button and choosing one of several options).

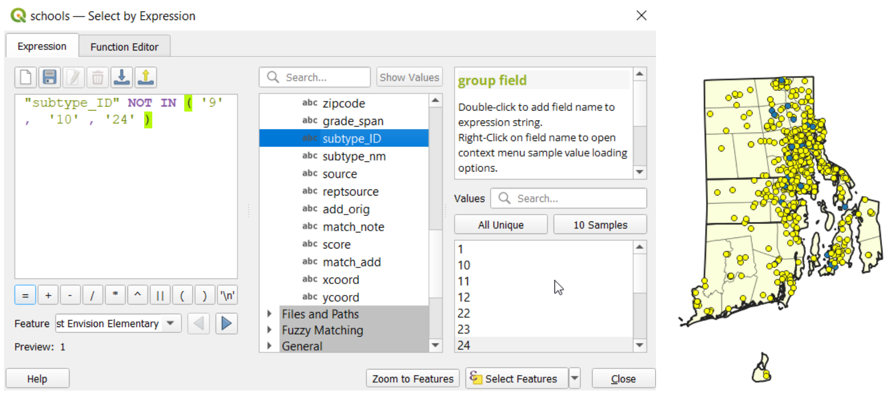


	name	name_30	name_15	orgtype_ID	org_type	address1	address2	city	state	zipcode	grade_span	subtype_ID	subtype_nm
1	Rhode Island S...	Rhode Island ...	RISD	2	School	2 College Street	NULL	Providence	RI	02903	NULL	10	Private Colleges

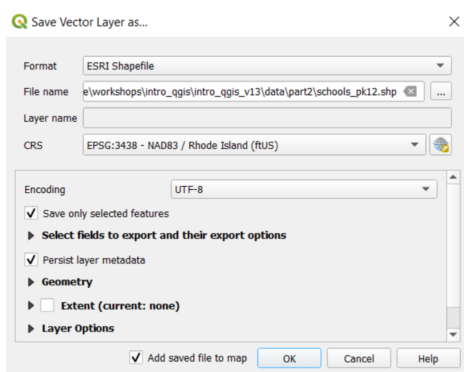
6. *Select features by attribute.* Note the types of schools indicated in the subtype_ID and subtype_nm columns, then close the attribute table. With the schools layer still selected in the LP, hit the dropdown arrow beside the Select features by value button and select the Select features using an expression button on the tool bar. This opens the Select by expression window, which allows you to select features based on shared attributes. In the Function list (center menu), scroll down to Fields and Values and hit the arrow symbol to expand the options. Double-click the subtype_ID field, which adds it to the Expression box at the left. Hit the All Unique button under the Field values box (on the right) to display all possible values for the subtype field. Type the following into the expression box:

```
"subtype_ID" NOT IN (  '9' , '10'  '24' )
```

We have to quote the ID numbers, as they are saved as text values. Click the Select Features button. You've just selected all of the schools that are not colleges and universities (i.e. grades pre-K to 12). Close the expression editor window and take a look. (Note - the expression window is also available in the attribute table window, via the same button .)



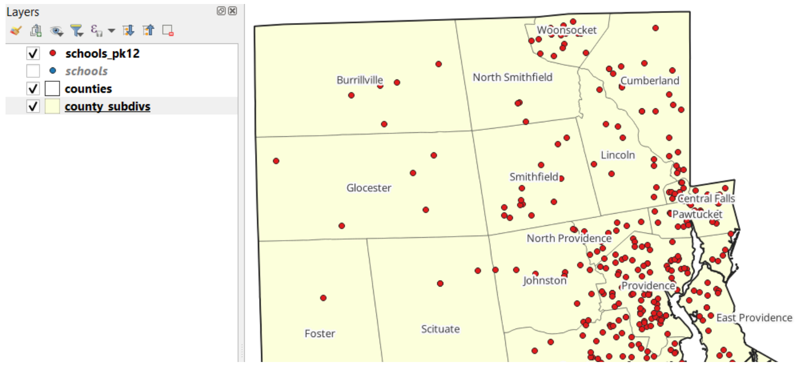
7. *Save selected features.* To save the selected features as a new layer, select the schools in the LP, right click, and choose Export > Save Selected Features As. Change the format from Geopackage to ESRI Shapefile. Hit the ... ellipses button beside the File Name box, browse to the data > part2 folder, and name the file schools_pk12 (ALWAYS make sure to browse and specify where the file should be saved). Hit Save. Back on the Save Vector Layer menu, make sure the Save only selected features box is checked. Hit OK. Voila, you've created a new vector layer!



8. *Clear selected features.* Uncheck the new schools_pk12 layer in the LP. You'll see the original schools layer with the yellow selected features. Click the Deselect features button on the toolbar to remove selected features from all layers. Alternatively, with the Select features button active, you could click on an area of the map that has no schools to clear the features, or you could clear the current selection from the attribute table.

9. *Symbolize pre-K to 12 schools.* Uncheck the original schools layer to turn it off. Check the schools_pk12 layer to turn it on, and drag it to the top of the LP. Double click on schools_pk12 and enter the Symbology tab. Change the color of the schools to red, and change the size from 2.0 to 1.5.

10. *Labeling features.* Attributes stored in the table can also be used to label features. Select the county_subdivs layer in the LP to activate it. Double-click on the layer in the LP and access the Labels tab via the layer properties menu. In the top dropdown menu change the option from No labels to Single labels, and in the Label with dropdown menu choose the NAME column. Change the size of the font to 8. Then click the Buffer option on the left, and check the Draw text buffer box. Click OK. Explore the map a little, and notice how the labels shift as you zoom. We'll experiment more with labeling later on. Go back to the labels tab and change the top dropdown to No labels, then hit OK.



2.4.2 Commentary

Attribute Tables

Every vector feature has a record in the attribute table; you can't have a feature without an attribute or vice versa. In a shapefile, the geometry is stored in the .shp file, an index of the geometry is in the .shx file, and the attributes are stored in a .dbf file. As we'll explore throughout this tutorial, attributes can be used for selecting, symbolizing, and labeling features in layers.

In GIS software attribute tables are managed and handled in the same manner as tables in a relational database. Each column has a data type associated with it which determines the kind of data that can be stored in that column and the types of operations that can be performed on it. Data types include strings (aka text) and various types of numeric fields (integers for whole numbers, reals for numbers with decimal places, etc). When you use the Expression Builder to select features, like "subtype_nm" = 'State Colleges and Universities' , you are actually creating SQL code, which is a standard language for manipulating data in a database. Text values must be surrounded by quotes; this includes numeric ID codes that are saved as text or 'string' fields. You do not use quotes when querying actual numeric values saved as integers or decimals.

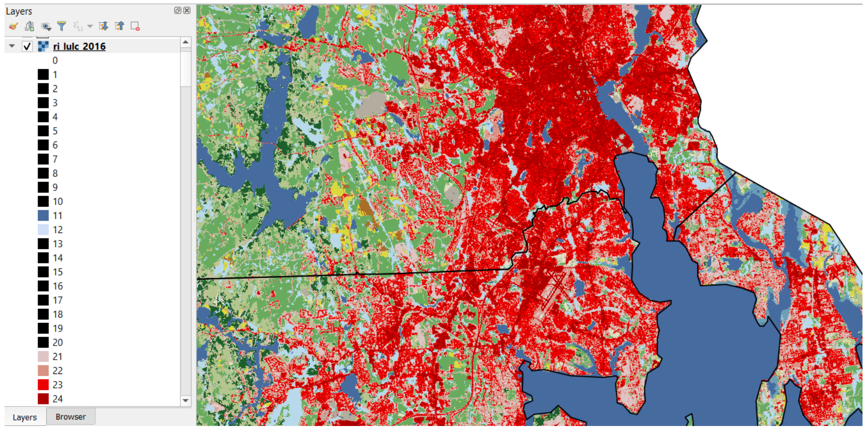
2.5 Adding Raster Data

In this section you'll get a very brief introduction to raster data, to demonstrate how this model differs from the vector model.

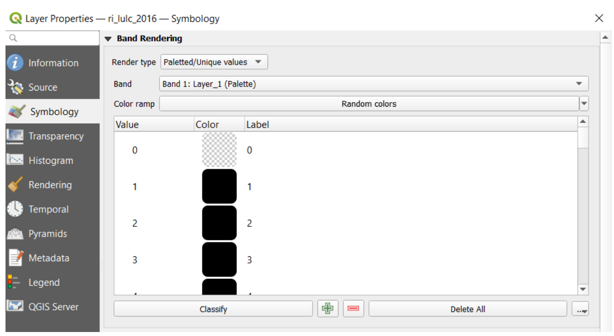
2.5.1 Steps

1. *Add raster data.* Hit the Data Source Manager button. Then select the Raster button in the list on the left. Under Source, hit the ellipsis button. Browse to the data folder for part 2, look in the rasters subfolder and select the ri_lulc_2016.tif file and add hit open. Add the layer and close the manager. Once the layer is added, drag it in the LP so it appears below the county layer, and turn the schools_pk12 layer off.

2. *Explore raster map.* If you zoom in you can see the individual pixels that make up the raster, where each grid cell represents one value denoting its land use and land cover, such as water, urban, forest, or farmland. In this particular raster, a specific color palette has been pre-assigned to the values based on the Anderson LULC classification scheme (see appendix). In the LP, extend the arrow below the ri_lulc_2016 layer to see the categories. Values assigned to black represent no data, while the other colors represent a specific land use. Double click on the raster in the LP to open the Properties menu, and click on the Information tab (icons on the left). In this menu you'll see the file type, it's components, the size of the raster (2128 by 3184 pixels), the coordinate system, the pixel size (approx 105 feet tall and wide) and descriptive stats about the values.



3. *Symbolize raster map.* If this raster did not have a pre-defined palette, we could classify it ourselves. Click on the Symbology tab. Change the Render type dropdown from Paletted / Unique Values to Singleband gray and hit OK. Inspect the map. Return to the Symbology tab, and this time choose Singleband pseudocolor. Under the color ramp drop down choose a color scheme, then underneath the empty Value box hit the Classify button, then hit OK. Examine the output; these two schemes would make sense if the pixels represented quantities like elevation or pollution content, which would be assigned a shade or color based on value ranges. In this example, the values are categorical, so Paletted / Unique Values makes more sense. Go back to the Symbology tab and change the option back to Paletted / Unique Values. Hit the Classify button, then hit OK. Note that by reclassifying the data, the empty value classes have been removed so it's easier to see the range of actual LULC classes.



4. *Add a different raster.* The LULC raster was derived by classifying and interpolating values from satellite imagery. Satellite imagery and aerial photographs are also stored as rasters, as are historic paper maps that have been scanned and georeferenced. Uncheck the LULC raster in the LP to turn it off. Go back to the Data Source Manager and select the add raster layer button. Browse to the data folder for part 2 and in the rasters subfolder select the RI_Providence_353442_1939_risp file. Add it to the map and close the manager.

5. *Explore the topo map.* This is a georeferenced topographic map for the City of Providence from 1939. Use the zoom tools to center this layer in the Map View, then select the layer in the LP, right click, and choose the option to Zoom to Native Resolution. This adjusts the zoom of the map so that it appears at the optimal level for the raster. For reference, check the schools_pk12 layer to turn it back on. Slide the raster layer down in the LP so that it appears below the county layer. Explore the area around Providence and note how the raster layer lines up with the other layers.

6. *Extract raster features?* Unlike vector features, rasters have no attributes in a table as the layer exists as a grid of pixels, and not as discrete entities. You can extract raster features based on their pixel values; see the commentary for details.

7. *Clean up*. Uncheck the rasters to turn them off, and check all of your vector layers to turn them back on. This example has demonstrated how rasters differ from vectors, and how you can get basic information about them.

STOP

2.5.2 Commentary

Raster Data

Raster layers differ from vector layers in composition (continuous surface of pixels versus discrete geometric areas), file formats (many raster formats versus relatively few vector formats), resolution (optimal scale for raster layers matters more than vector layers), size (raster files tend to be much larger), and attribute tables (raster layers do not have attribute tables; the color of individual pixels denotes feature values). Given the differences in format, the tools for working with vector and raster layers are distinct.

Many geographic objects are represented in raster formats including satellite imagery, aerial photography, paper maps that have been scanned and digitized, and imagery that has been interpreted to represent value-added data that does not conform to political boundaries, such as land use and land cover, temperature, and population density.

The land use and land cover raster in this exercise came from the National Land Cover Database, which is created in concert by many US federal agencies. It was originally stored as a national file in an ERDAS image format, but for this exercise it was prepped in QGIS by clipping the raster to the extent of Rhode Island, using a vector boundary file of RI from the Census Bureau with the Raster > Extraction > Clip by Mask Extent tool. It was also reprojected from a continental map projection to the local state plate system, so it could be readily used in this exercise, and saved in a tif format.

You can select and extract raster cells by value. In the LULC raster, the pixel values pertain to a specific land use category (see appendix C for the classification system legend). You would select the Raster menu in the menu bar, and then the Raster Calculator. In the Raster Bands, you would double-click on the band to add it to the expression. In the expression you can type the operators (equal to, greater or less than) and the pixel value(s) you want. You would need to save the selection as a new raster (under Output Layer).

The historic topographical map was created by the US Geological Survey (USGS) and downloaded from their National Map archive. The old paper topo maps were scanned and georeferenced so they were assigned a coordinate reference system and can be used in GIS packages. The USGS provides the maps in .tif format called a GeoTIFF; a lossless image file that has georeferencing information (coordinates and map projection) embedded in it. The raster was prepped for this exercise by converting the old coordinate system used in the original map to the modern one our layers are in. We'll tackle coordinate systems later on in this tutorial.

The Georeferencer tool is located under the Layers menu. This tool allows you to georeference images and maps that lack a coordinate reference system, so that you can integrate them with other GIS layers and files. Tools like this one were used to georeference the historical USGS topo map.

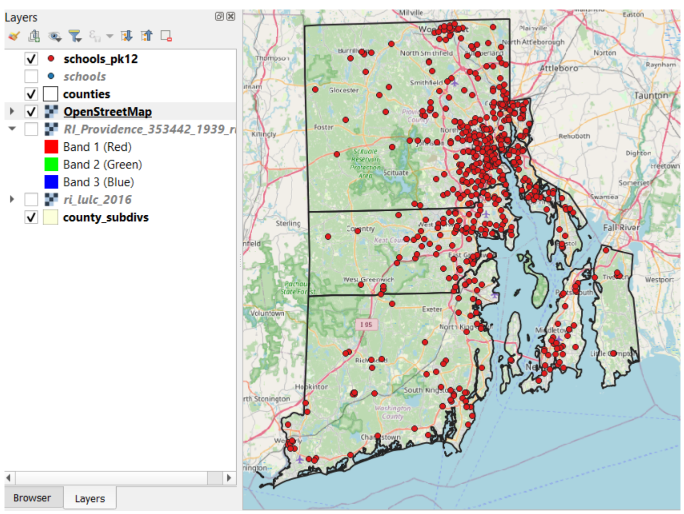
2.6 Adding a Web Map as a Base Map

In this section we'll add the web-based OpenStreetMap to our project, to serve as a reference base map for our other layers. You will need to be connected to wifi / internet in order to perform these steps.

2.6.1 Steps

1. *Add the OpenStreetMap.* Uncheck both raster layers from our last exercise to turn them off. Hit the  Zoom to full extent button to return to our default zoom. Flip from the Layers panel to the Browser panel. Hit the arrow beside the XYZ Tiles layers to see the OpenStreetMap (OSM) underneath. Click on OSM, right click, and Add Layer to Project. If prompted with a coordinate transformation screen, just click OK. This renders the OSM directly from the web into your project.

2. *Reorder Layers.* Flip back to the Layers panel. Drag the OSM in the layers list so that it appears below the counties but above the county subdivisions. Zoom in and pan around the map, and note how the schools and counties overlay with the OSM. As you pan and zoom, new tiles for the OSM are fetched from the web and redrawn.



3. *Add Transparency.* Drag the OSM layer to the bottom of the LP, so the county subdivisions covers it. Select the county subdivisions in the LP and go into the Properties and Symbology tab. Drag the Opacity slider down to 60% and hit OK. Now you can see through the subdivisions to see the OSM map below it.

4. *Having drawing issues?* If your internet connection is slow and there is a significant lag when moving around the map, turn the OSM layer off, zoom or pan to the area you want to explore, then turn OSM on again.

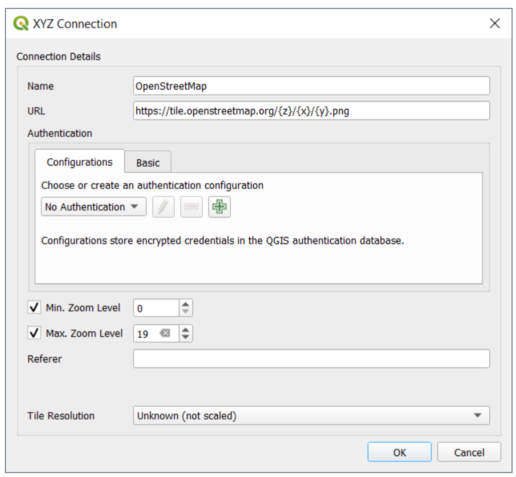
2.6.2 Commentary

Web Mapping Service

A web map service (WMS) is an open standard for serving georeferenced maps via the web. WMS layers are saved in a geodatabase system on a webserver, and are typically rendered as raster-based layers via a website or a desktop-based GIS program when a client connects to a host and requests the layer. As an end user zooms closer or further, the actual map that's rendered may switch from showing a generalized, small scale map for a large area to a detailed, large scale map for a small area. WMTS and XYZ Tiles are specialized versions of WMS, in that they render rasters as smaller, individual tiles that cover a small area within the view extent defined by a GIS or browser window. This method is usually faster and more efficient than rendering an entire raster layer.

Web layers are particularly valuable as a source for base maps. In our example, the OSM provides a stylized map that depicted streets and major features without us having to go through the trouble of downloading and stylizing a number of vector layers (which is time consuming) or downloading and stitching together several rasters (which consumes a lot of time and disk space).

QGIS includes a connection to the OpenStreetMap by default. If you know the specific URLs for additional web mapping services, you can add them in the Browser Panel as either as WMS / WMTS or XYZ layer. Select the format, right click, and choose the option to add a new connection. The screen for the OSM connection (accessed by selecting the connection in the Browser and choosing the edit option) is displayed below; at minimum you must provide a name for the connection (make one up) and a link. See the appendix for a list of free web services you can add to QGIS.



If a web map layer looks out of focus, you can adjust the scale by right clicking on a blank area of the tool bar and selecting the Tile Scale Panel. By adjusting the slider bar in the panel you change the zoom based on the resolution of the available tiles. If the slider is greyed-out, select the web layer in the LP, right click, and choose Set CRS - Set Project CRS from Layer. This will change the coordinate reference system (CRS) of the map view to match that layer (usually a web mercator projection), and you'll be able to adjust the slider. When you're finished, you can reset the map window back to the CRS of your other layers.

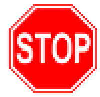
2.7 Saving Your Project

You'll learn how to save your project.

2.7.1 Steps

1. *Save your project.* Hit the save project button. Navigate to the data folder for part 2, and save your project there as part2.qgz. The project file saves the symbolization, labeling, and current zoom for your data, and links

to your data files (vectors and rasters); the data files themselves are NOT stored inside your project file but exist independently. In order to use your project in the future, the project file and the data files you used must be kept together.



2.7.2 Commentary

Project Files

When you add vector and raster data to a project file you are not saving the data inside the project; you are saving links to those files. Elements like symbolization, data classification, the extent of your last zoom, and any finished maps you create are stored in the project file. When you click on the project file to open it, the software looks at the paths to your data, re-establishes the links, and then applies the settings (symbols, zoom, etc) that you have saved in your project file. This relationship is of crucial importance when it comes time to move or share files - if you move your project file or your data the links between them will break, and you'll need to re-establish the location between the project and the data in order to repair your project file. If you open a project and QGIS can't find the data, you will have the opportunity to restore the links by browsing through your folders and selecting each file that corresponds to a layer you have in the LP of your project. Once you restore the links, you can save the project and it will save the new links.

Paths to files can be stored as absolute links or as relative links. Relative paths are the default in QGIS, and are the safer choice. A relative link contains the directory and file information for the folder the project file is in (i.e. path would be .\counties.shp) and all folders below it (i.e. path would be .\data\counties.shp). Since anything above the project's directory is omitted, relative paths are a good choice if you know that you'll be sharing your project data or moving it around. Relative paths are a bad choice if your data is not going to be stored underneath your project folders (i.e. it's stored above the project directory, in a parallel directory, or another drive or server all together).

An absolute link contains the complete path of a file, such as F:\My_Stuff\GIS_Practicum\part2\data\counties.shp. Use absolute paths when you're working in an established environment where you know that you won't need to move data and projects around, or in situations where your project files won't be stored directly above or in the same folder as your data. Absolute paths are a bad choice if you know you'll be moving data around; they're particularly bad if you're working on a usb drive in a MS Windows environment, as the paths can change as you move from machine to machine (i.e. F:\My_Stuff... on one machine becomes E:\My_Stuff... on another machine; QGIS won't be able to locate the files stored on F:\My_Stuff because it doesn't exist that way on the 2nd machine). If you ever need to change the setting to use absolute paths, you can modify this option under Project > Properties > General tab in the General Settings section labeled Paths.

Think carefully about where to save project files in relation to your data, and once you've created your project file keep project files and data in a consistent place. Also remember that you must keep all of the individual components of a shapefile together (.shp, .shx, .dbf, .prj, etc); otherwise the shapefile will not function. If you want to share your project file with someone, you will also have to send them your data; the project file cannot exist independently from the data. You can share views or maps you've created in a static format (image file or PDF) that is separate from your project and data files; we'll explore that later in this tutorial.

Chapter 3

Geographic Analysis

The goal of this chapter is to introduce analysis and geoprocessing tools and techniques using an example that incorporates basic tasks that all GIS users would need to perform, regardless of application or field of study. In our hypothetical example, the State of Rhode Island has been given funding from the Institute of Library and Museum Services for a specific initiative, and they need to identify public libraries that would be eligible to receive funding. The initiative provides after school programs to promote literacy and research practices for public school students in low to moderate income urban areas. To qualify for this funding, a public library must be:

1. Sufficiently large enough to host the program (no less than 2000 square feet)
2. Within 1/3rd of a mile of a K-12 public school
3. Within or near (just across the boundary from) an area:
    (a) That is urban
    (b) Where the percentage of children is higher than the state average
    (c) Where median household income is below the 80th percentile for the state

To do this analysis, we'll take the following steps which will illustrate many key GIS processes:

3.1 Create a new project from an existing one

3.2 Take census data on age and income from a data table and join it to a vector layer of census tracts, which we'll use for measuring population relative to libraries

3.3 Plot a CSV file of public libraries in the US using longitude and latitude coordinates, to create a vector layer of libraries

3.4 Create subsets of layers for libraries (select by location to create a Rhode Island layer, and delete bookmobiles), K-12 public schools (use filter to hide private schools), and census tracts (select by attributes using quantitative criteria) to prepare for our analysis.

3.5 Geoprocess layers to clip US urban areas to Rhode Island, calculate the intersection between urban areas and tracts that met the criteria, draw buffers of 1/3rd of a mile around libraries and count public schools that fall within each buffer.

3.6 Perform multiple selections to select library buffers that include a school and a library of sufficient size where the buffer intersects urban areas of tracts that met the criteria. We'll add a column to the attribute table to flag eligible libraries.

3.1 Creating New Project From Existing One

This section will show you how to create a new project from an existing one and will set the working environment for the rest of part 3. We will remove layers we no longer need while turning some others off.

3.1.1 Steps

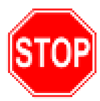
1. *Open project. Launch QGIS.* Hit the open project button (or go to Project > Open). If you closed the project from Part 2, browse through your folders to the QGIS project file you created for part 2, and select it to open it. Or, go to Project > Open Recent and select it there.

2. *Save Project As.* Once your project has loaded, go to Project > Save As. Browse to the data folder for part 3. Save the project in that folder as part3.qgz. Hit save. You've now saved a new copy of your old project, and are currently working in this new copy (you can tell by looking at the title at the top of the window, where the project name is listed). We will work with this new project, part3.qgz in this chapter.

3. *Remove layers.* We don't need the two raster layers for this exercise, nor do we need the original schools layer. Select each one in the Layer Panel (LP), right click on the layer, and select Remove.

4. *Turn off layers.* Uncheck the county_subdivs and the OpenStreetMap layers for now (keeping them in the project, but turned off). You should now have four layers in the project, arranged in this order: schools_pk12 and counties (checked on) and county subdivisions and OpenStreetMap (checked off).

5. *Zoom out and save.* Hit the zoom to full extent button to zoom out to the full extent of your layers. Then hit the save button.



3.2 Joining and Mapping Attribute Data

In this section you'll learn how to join an attribute table to a vector file and map the attributes in that table. You will add a layer for census tracts, which are statistical areas created by the US Census Bureau to represent census data for small areas; they're designed to have an ideal size of 4,000 residents. We will eventually use the tracts to measure two of the criteria for the initiative: selecting areas where the percentage of children is higher than the state average, and where median household income is below the 80th percentile for the state. In order to do that, we need to associate a census data table with the tracts.

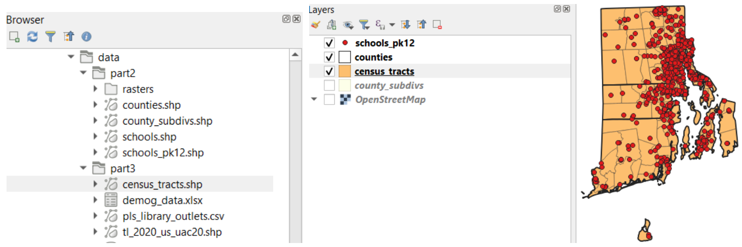
The boundary file was sourced from Ocean State Spatial Database Project, and is based on the the US Census TIGER files. The data table was created by downloading and aggregating age and income data from data.census.gov.

3.2.1 Steps

1. *Add the tracts shapefile using the Browser.* The Browser allows us to see our file system within QGIS and to add data directly to our project, instead of using the Data Manager. The Browser has folders that represent hard and external drives as well as icons for connecting to various spatial databases and web services. Tab over from the Layers to the Browser panel, use the arrow buttons to expand the folder tree where your project files are stored, and drill down to the part3 folder. Select the census_tracts.shp file. You can either right click and choose the option to add it to your project, or you can hold down the left mouse button and drag it into the Map Window. Then, flip back to the Layers Panel, and drag the tracts layer in the LP so it appears below the counties. Modify the symbology of the census_tracts, under the Simple Fill option change the strole width to 0.10.

2. *Open the census data file.* Minimize (don't exit) QGIS for the moment. Using your file manager, browse to the data folder for part 3. Look for an Excel spreadsheet file called demog_data.xlsx. Depending on what operating system you're using, open this file with a spreadsheet package like Microsoft Excel or LibreOffice Calc (double-click and it should open in the appropriate program).

3. *Examine the data file.* The data file contains one row for each tract in RI and several columns of attributes. The first column contains the unique ANSI / FIPS identifier used by the Census Bureau; we'll use it to join this table to the shapefile. The data comes from the American Community Survey (ACS) and includes: the total



population, the population under 18 and 65 and over (total estimates and percent totals), and median household income. Each estimate is paired with a column that contains a margin of error (moe). You would interpret the estimates thusly: For Census Tract 301 in Bristol County we're 90% confident (that's the confidence interval for the ACS) that there were 1,289 people under the age of 18 between 2017-2021, plus or minus 223 people.

	A	B	C	D	E	F	G	H	I
1	gid	geoname	vintage	totpop	totmoe	und18pop	und18moe	und18pct	und18pmo
2	1400000US44001030100	Census Tract 301, Bristol County, Rhode Island	2017-2021	4813	501	1289	223	26.8	3.3
3	1400000US44001030200	Census Tract 302, Bristol County, Rhode Island	2017-2021	3350	428	902	184	26.9	3.4
4	1400000US44001030300	Census Tract 303, Bristol County, Rhode Island	2017-2021	4693	403	1302	195	27.7	3.2
5	1400000US44001030400	Census Tract 304, Bristol County, Rhode Island	2017-2021	4345	481	1305	261	30	3.9
6	1400000US44001030500	Census Tract 305, Bristol County, Rhode Island	2017-2021	3314	375	378	103	11.4	3.3
7	1400000US44001030601	Census Tract 306.01, Bristol County, Rhode Island	2017-2021	3649	391	604	146	16.6	3.6
8	1400000US44001030602	Census Tract 306.02, Bristol County, Rhode Island	2017-2021	4203	439	736	189	17.5	4.4

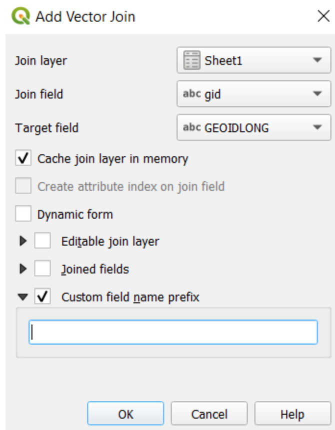
4. *Examine the attribute table of the tracts.* Close the Excel file, exit your spreadsheet software and maximize QGIS. Select the census_tracts layer in the LP, right click and open the attribute table. In the table, note the column labeled GEOIDLONG. It contains the same Summary Level and ANSI / FIPS code (the state-county-tract number) that was stored in the gid column in the data table. Since these columns are the same, we can use them to join the two files. Close the table.

census_tracts — Features Total: 247, Filtered: 247, Selected: 0

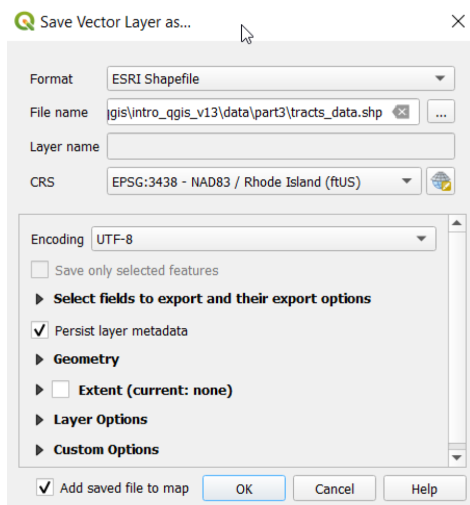
	STATEFP	COUNTYFP	TRACTCE	GEOIDSHORT	NAME	NAMELSAD	MTFCC	FUNCSTAT	INTPTLAT	INTPTLON	GEOIDLONG
1	44	001	030100	44001030100	301	Census Tract 301	G5020	S	+41.7442752	-071.3457473	1400000US44001030100
2	44	001	030200	44001030200	302	Census Tract 302	G5020	S	+41.7500065	-071.3238799	1400000US44001030200
3	44	001	030300	44001030300	303	Census Tract 303	G5020	S	+41.7551629	-071.3012118	1400000US44001030300
4	44	001	030400	44001030400	304	Census Tract 304	G5020	S	+41.7143862	-071.3015021	1400000US44001030400
5	44	001	030500	44001030500	305	Census Tract 305	G5020	S	+41.7313781	-071.2812695	1400000US44001030500
6	44	001	030601	44001030601	306.01	Census Tract 306.01	G5020	S	+41.7228958	-071.2536137	1400000US44001030601
7	44	001	030602	44001030602	306.02	Census Tract 306.02	G5020	S	+41.7177074	-071.2738822	1400000US44001030602
8	44	001	030700	44001030700	307	Census Tract 307	G5020	S	+41.6713548	-071.2753140	1400000US44001030700

5. *Add Excel file to the project.* You can either drag the spreadsheet from the Browser into your project, or use the Data Manager. In the Data Manager you would add the spreadsheet as a vector layer. Browse to your part 3 folder, change the file type drop down from ESRI Shapefiles to All Files. Click on demog_data.xlsx to select it, then hit Open, Add the file, and close the manager. It should appear in the LP as Sheet1, which is the first and only sheet in our workbook that has data. You can select it in the LP and hit the open table button to verify that the table displays correctly.

6. *Join data table to shapefile.* Close the table, and double click on tracts layer to open its properties menu. Hit the Joins tab. Hit the green plus button to add a join. The join layer will be the data table Sheet1. The Join field in that table is gid. The Target field in the tract layer is GEOIDLONG. At the bottom of the menu, check the box that says Custom field name prefix, and delete the text so that the option is blank. This will preserve the original column names from the spreadsheet. Keep the other defaults and hit OK, then hit OK again in the properties menu to make the join stick. Then right click on tracts in the LP and open the attribute table. Scroll over to the right, and you'll see all of the layers attributes and the data that is stored in the Excel file. Close the attribute table.

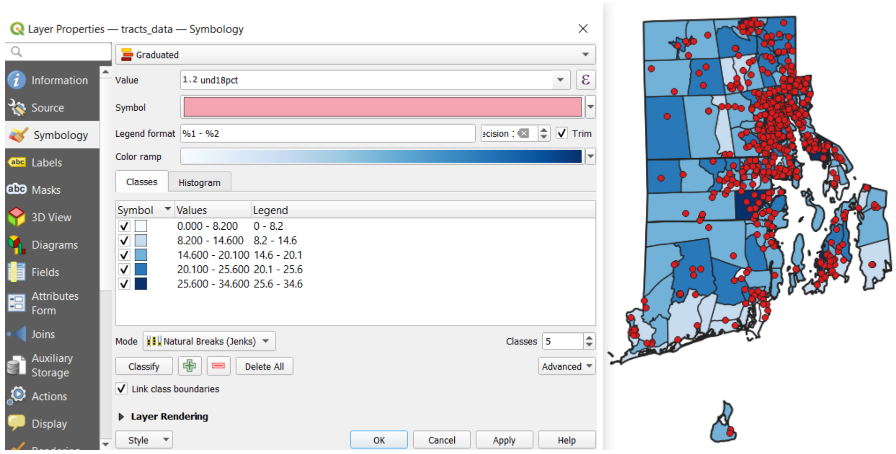


7. *Make the join permanent*. Our data table is joined to our shapefile, and if we save the project the join will be saved within the project. At this point we could map the joined data, but we won't be able to perform a lot of other operations unless we make the join permanent, so that the data is fused to the shapefile as new attributes. To do this, we simply have to save this shapefile as a new one. So, select census_tracts in the LP, right click and choose Export - Save Features As. Browse to the part 3 data folder and Save the layers as tracts_data. Leave the encoding and the CRS alone, and make sure the Add saved file to map box at the bottom is checked. Hit OK.



8. *Reorder the layers*. Select census_tracts layer in the LP, right click and remove it. Also, remove the data table Sheet1. Drag the new tracts_data layer so that is falls below the counties layer. If you open the attribute data for the new tracts_data you'll see the data from the table has been fused to the shapefile. Save your project.

9. *Map the data*. Now that the data is joined to the boundaries, we can map it. Double click the tracts_data layer in the LP and go to the Symbology tab. Change the top dropdown from Single symbol to Graduated. Change the Value to und18pct (percentage of the population that is under 18). Change the mode from Equal Counts to Natural Breaks. In the Color ramp drop down select a scheme that has a range of single-color values that go from light to dark. Hit the Classify button, and then hit OK. You should now have a choropleth (shaded area) map that shows the percentage of the total population that's under 18 for each tract, classified by natural breaks (divides data into categories based on gaps in values). We'll discuss color and classification schemes in more detail later on. Save your project.



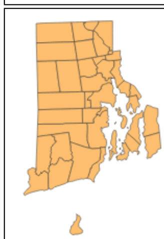
3.2.2 Commentary

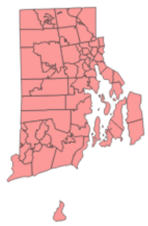
Geographic Units

For this exercise we're working with census tracts, which are statistical areas created by the US Census Bureau for publishing census data for small areas. Census tracts are designed to have an ideal population size of 4,000 residents; since they are roughly similar in population size, equivalent comparisons can be made between them. Their geographic size varies tremendously between urban and rural areas. Tracts are built by combining smaller census statistical areas (block groups and blocks) and their boundaries typically coincide with major topographical features (roads, rivers) or legal boundaries (state, county, and municipal).

Tract boundaries are revised every ten years in conjunction with the decenial census, and are used for publishing data over the course of the following decade. The tracts in our example are based on 2020 boundaries, as we were using 2021 ACS data.

Here is a summary of some of the most common geographic areas for thematic mapping in the US (most countries will have some corollaries):

	**Counties** - Legal subdivisions of states, counties are commonly used for mapping national or regional distributions given the large amount of data that's available for them. In Rhode Island, Connecticut, and parts of Massachusetts counties no longer have a legal function, but are still used for publishing statistical data. Dataset availability: decennial census, 1 and 5-year ACS, population estimates
	**County Subdivisions** - Represent legal subdivisions in states that have them (Municipal Civil Divisions) and statistical areas created by the Census Bureau for states that don't (Census County Divisions). In Rhode Island and throughout New England they represent MCDs; cities and towns that exercise local governance. As legal areas MCDs vary in both geographic and population size. Dataset availability: decennial census, 5-year ACS

	**PUMAs** (Public Use Microdata Areas) - Named statistical areas created by the Census Bureau to have approximately 100,000 residents; they're created by aggregating census blocks. In urban areas they can represent subdivisions of cities, in suburban areas they represent subdivisions of counties, and in rural areas they are often aggregates of several counties. Dataset availability: 1 and 5-year ACS
	**ZCTAs** (ZIP Code Tabulation Areas) - Statistical areas created by the Census Bureau to approximate areal USPS ZIP Codes. ZCTAs are created by aggregating small statistical areas called census blocks based on the predominant ZIP Code used by addresses within the blocks. While not ideal for representing communities, ZIP Codes are often used for this purpose since most people are familiar with them. ZCTAs do not correspond with other census geographies. Dataset availability: decennial census and 5-year ACS
	**Census Tracts** - Statistical areas created by the Census Bureau to have approximately 4,000 residents (with a range of 1,200 to 8,000). Tracts can be used for analyzing patterns within counties, cities, and neighborhoods and can be aggregated to create neighborhood-like areas; many cities create official neighborhood or sub-municipal areas based on tracts. Dataset availability: decennial census and 5-year ACS

The choice of a geographic unit is an important decision; it's often a balance between the availability of data for an area, the suitability of the unit for the analysis, the amount of work that has to be invested in processing and analyzing the data, and the final outputs that will be created (tables, charts, maps) to explain the data.

We used tracts for this exercise because the demographic data availability for them is good, there is a work-able number of them in the state (approximately 240), and they are commonly used for studying distributions below the county level. Other geographic areas present different trade-offs. County Subdivisions and PUMAs would provide us with more precise census estimates as their population sizes are larger, but for our case study they are too large geographically and mask a lot of variability within each area. ZCTAs are readily recognizable to most people and are frequently used for approximating communities, but were never designed for this purpose - they were designed for delivering mail and they vary tremendously in size, shape, and population. They also don't mesh well with other types of geography.

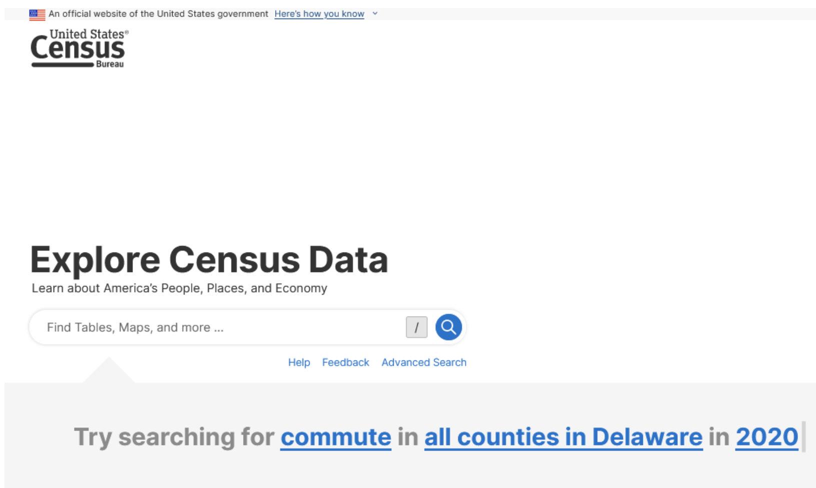
We could opt for areas smaller than census tracts, like census block groups, that would allow us to make finer geographic selections. The problem at this scale is data reliability; margins of error for ACS estimates (see Census Data in the following section) are unacceptably high at the block group level for most applications. We could use decennial census data for the under 18 population to circumvent this problem, but data on income is only available in the ACS. Again, these choices represent compromises.

Census Data

The demographic data used in this exercise comes from the American Community Survey (ACS). Most people are familiar with the decennial, ten-year census, which is a 100% count of the population that's used to reapportion seats in Congress. The ACS is a rolling sample-survey of population characteristics. Each year the census publishes annual ACS estimates for all geographic areas in the US that have at least 65,000 people. Estimates are at a 90% confidence interval and are published with a margin of error. Since the survey's sample size is too small to publish reliable data for areas with smaller populations, the Bureau averages data over a 5-year period for smaller areas. This 5-year series includes all geographic areas down to census tracts and block groups (although data for the latter is highly unreliable). Each year the Bureau releases a new annual data set and updates the 5-year averaged series by adding the latest year of data and dropping the oldest one. For our exercise we are using 5-year average data, as that's the only ACS series that is published at the tract level.

The American Community Survey was designed to provide data on a frequent basis and to replace the form on the decennial census that collected detailed socio-economic characteristics of the population. Beginning with the 2010 Census, the decennial census only provides basic demographic indicators of the population such as age, sex, race, and the total number of households and housing units. The decennial census is a count (not a survey) of the population and continues to be useful for making historical comparisons, providing a baseline for creating estimates, for doing analysis below the census tract level, and for providing exact counts when estimates aren't suitable. The 2020 census is similar in composition to 2010.

All recent data from the US Census Bureau is available for download from data.census.gov `https://data.census.gov`. All of the data is free and in the public domain. When you download the data you may have to process it to aggregate certain variables before you can use it. The data that we are using in this exercise has been preprocessed to aggregate certain columns and delete unnecessary ones. Boundary files to associate with the data must be obtained separately (we discuss this in Chapter 4).



Census data from other countries may be more difficult to obtain, as is may not be free or in the public domain, and may not be available in a digital format. You can check the website of the statistical agency for an individual country to see what is available, or you can visit the websites of international organizations like the United Nations or World Bank to obtain basic population data for all countries.

Identifiers

The ability to join data tables in a database or a data table to a vector file is made possible by the use of identifiers, which are codes used to uniquely identify features. If features in two separate data tables share the same identifier, those data tables can be matched or joined together based on that common identifier, allowing you to create new data or to map data in a table.

There are several standard codes for identifying features. In the United States, ANSI / FIPS codes are a classification system for identifying all legal, administrative, and statistical areas in the country. For example, ANSI / FIPS 44007003602 is the code for Census Tract 36.02 in Providence County in Rhode Island. The first two digits (44) are the code for Rhode Island, the next three digits (007) are the unique code within RI for Providence County, and the last six (003602) are for the census tract number (the last two digits of the tract number are reserved for numbers to the right of a decimal point). In an attribute table these codes may appear in separate columns (state, county, tract) or in a single column as one string. In our exercise the demographic data used the longer version of the code which incorporates the census summary level (digits prior to 'US') and the ANSI FIPS. For example, in 1400000US44007003602, the number

140 is the summary level code for census tracts.

A list of US ANSI / FIPS codes for states and territories is available in the appendix of this tutorial, and the US Census Bureau maintains lists of codes on its website. The US government has also created two-letter alpha FIPS codes for each of the world's countries and uses them for international data published by various agencies. However, international data is more commonly coded with ISO codes (ISO 3166) which are available in a two-letter alpha format, a three letter alpha format, and a three-digit numeric format.

Sample Country Codes

Country	FIPS 10	ISO 3166		
Denmark	DA	DK	DNK	208
Djibouti	DJ	DJ	DJI	262
Dominica	DO	DM	DMA	212
Dominican Republic	DR	DO	DOM	214

It is a best practice to store ID codes as text and not as numbers since they don't represent quantities. Storing ID codes as numbers can result in data loss and misidentification. If codes begin with a value of zero and the ID is stored as a number, the zero will be dropped and the code will be incorrect. Examples of codes with leading zeros include Census ANSI / FIPS codes, USPS ZIP Codes, and 3-digit ISO country codes.

In order to join two tables together based on an identifier, you need to be sure that each field is stored in the same data format; if one is stored as text and the other is numeric, the join will fail. Furthermore, you need to ensure that each record is unique because one to many joins are not allowed; if you have a data table that has multiple records for one country, only one of those records will be joined to the GIS data file and the others will be dropped. Finally, you should avoid use place names as identifiers or join fields because there are often many inconsistencies (imagine the number of different ways for spelling or abbreviating country names like the United States or South Korea).

3.2.3 Tabular Data Files for QGIS

There are several different formats that you can use to get non-spatial tables into QGIS for the purpose of joining the tables to spatial files.

- *Spreadsheet files: .xlsx, .xls, .ods*. Includes newer and older Microsoft Excel formats and the Open Document Spreadsheet format used by OpenOffice and LibreOffice Calc. Add them to QGIS using the Add vector menu or the Browser. See the following subsection for details.
- *Delimited text files: .csv or .txt*. Plain text files with fields separated by delimiters (commas, tabs, or pipes) can be created in just about any program, and text files with coordinates can be converted into a spatial layer. CSVs can't preserve data types (i.e. text versus numbers) as easily as spreadsheets can. Add text or csv files to QGIS using the Add delimited text layer menu. Beginning with version 3.28, QGIS allows you to specifically designate data types for each column when importing, for instances where the software doesn't make a 'good guess' (designating an ID field as a number instead of text, which results in leading zeros being dropped).
- *dBase files: .dbf*. An older data table format that's still used in GIS, but that's been largely deprecated elsewhere. Can still be created and modified with Libre / OpenOffice Calc or various database programs.
- *Database tables*. QGIS is able to connect to a number of databases like PostGIS and Spatialite, and can access both spatial and tabular data via database connections.

Spreadsheet Files

With QGIS 3.x contemporary Excel spreadsheet files (.xlsx) and open document spreadsheets (.ods) are supported. There are a number of caveats and quirks in working with spreadsheets - follow these rules to insure that your data will load properly in QGIS:

- Your spreadsheet should consist strictly of rows of data with columns of attributes that describe them - you cannot have titles, footnotes, sum totals, or any stray text or information. It must be a strict grid of data.
- The first row will be your header row with the names of the columns; you cannot have multiple header rows. Names for columns should be kept short, should contain no spaces or punctuation (except underscores), and should not begin with numbers.
- In order to preserve the formatting of the data, you should specify formats for each column - text, numbers, date, etc. Remember that the data type of the unique ID column in your spreadsheet must match the data type of the unique ID in your spatial file. If one is saved as text and the other is a number, the join will fail.
- You should never mix text and numeric data in the same column. Columns should be either text or numbers. If you mix text (like footnotes) into your numeric columns then the entire column will be saved as text, and you won't be able to treat the numbers as numbers (i.e. for classifying data, performing calculations, etc.) in QGIS.
- Do not embed spreadsheet formulas in your data - QGIS won't know how to interpret them. If you have data that was created from a formula, you need to do a copy and paste special and replace the formulas with the actual values that result from the formulas.
- Avoid using any stylistic formatting on the data (colors, underlining or italicizing) or the cells (borders, merged cells).
- If you try to add a workbook to QGIS that has multiple sheets of data, it will prompt you to choose a sheet. You may want to rename the sheets to clearly identify them, rather than using the default Sheet1, Sheet2, etc.
- The names of attribute columns in shapefiles are limited to 10 characters, so if you intend to join a data table to a shapefile keep the column names in the data table short, otherwise the names will be truncated and will be difficult to interpret. Also, when doing a join to a shapefile it's a good idea to remove the custom field name prefix in the joins menu; otherwise the column names will all begin with the name of the worksheet and the actual names of the columns will be truncated and difficult to understand. Geopackages can handle column names of up to 128 characters (but it's still better to keep names short).

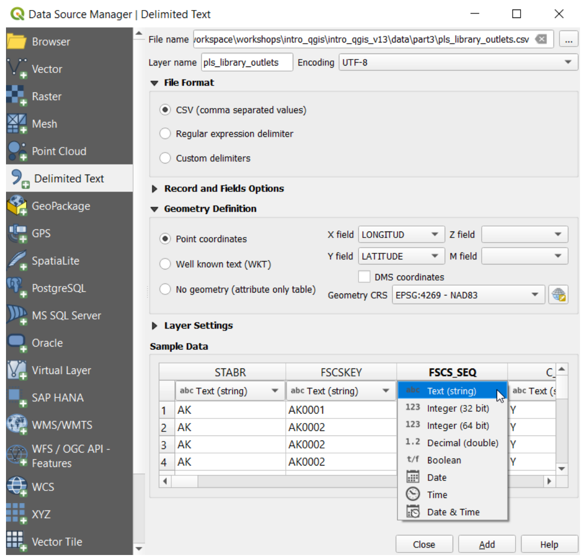
3.3 Plotting Coordinate Data

It's fairly common that you will encounter point-based data (representing events, buildings, or topographic features) with coordinates stored in a text fomat. In this section you'll learn how to take a text file with coordinate data, plot the data in GIS, and convert it to a shapefile. In this exercise you'll create a layer of public libraries from a CSV file that lists each library with its longitude and latitude coordinates. Since these coordinates don't match the state plane system coordinates of our existing layers, we'll have to plot them first and transform them to our system.

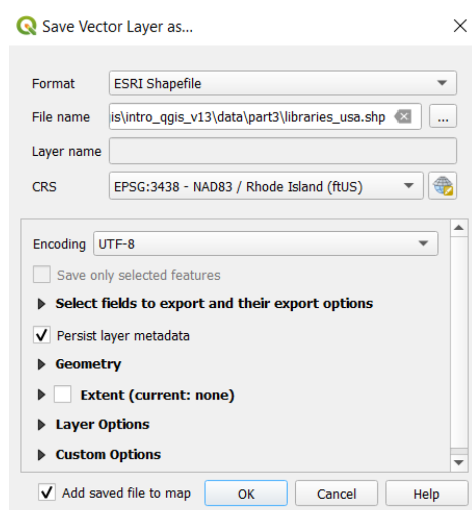
3.3.1 Steps

1. *Inspect the text file.* Go to your data folder for part 3, open the file pls_library_outlets.csv in a text editor (like Notepad on MS Windows - select file, right click - open with - Notepad) and examine it. This is a comma-delimited text file with data for public libraries across the US; each record represents one library outlet and each attribute column is separated by a comma. Close the file when you're finished.
2. *Launch a blank project.* We don't want to plot our longitude and latitude-based points over top of our RI state plane layers, as they don't share the same coordinate system. We'll use a blank workspace to plot our layers and then we'll add them back to our project. First, save your project. Then hit the New project button to get a blank workspace.
3. *Launch the delimited text layer menu.* Go to the Data Manager and select the delimited text layer option. Under File Name, Browse to the Part 3 folder and select pls_library_outlets.csv and hit Open. This will populate the menu screen. Under File format keep CSV as the format. Under Record and Field Options verify that the First

record has field names box is checked. Under Geometry definition, the Point coordinates radio button should be selected. Under the X field drop down select LONGITUD (the X coordinate is always longitude). Under the Y field drop down select LATITUDE (the Y coordinate is always latitude). For the Geometry CRS, hit the little globe icon, and in the filter type NAD83. Hit the dropdown menu under Geographic (2D) and select NAD83 (EPSG:4269) from the list and hit OK. Back in the Delimited Text Menu, scroll to the right in the sample data at the bottom, and change the column types for FSCS_SEQ and ZIP from Integer to Text. When finished, hit Add. If prompted with a coordinate system transformation window, hit OK, then Close the delimited text menu.

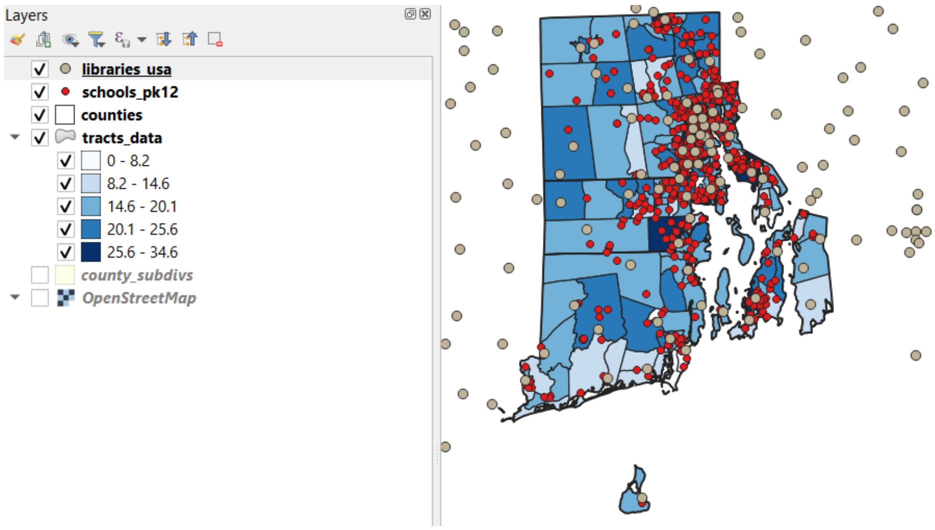


4. *Convert the plot to a shapefile.* Even though the points have been plotted, it isn't a vector file yet - we have to convert it. When we convert it, we can also transform it to match our other layers. Select and right click on pls_library_outlets in the LP and choose Export - Save Features As. Hit the little globe button beside the Selected CRS. In the coordinate system menu, under recently used coordinate systems select the NAD 83 / Rhode Island projection and hit OK (if you don't see it there, use the filter and search for 3438, look under the Projected Systems / Transverse Mercator dropdowns and click on it to select it). Browse and save the file as an ESRI shapefile in your part 3 data folder and call it libraries_usa. Make sure the Add saved file to map box is checked. Then hit OK to save the layer. If prompted with a coordinate transformation screen, just hit OK and proceed.



5. *Re-open our project and add the new layer.* The new libraries_usa layer has been added to our temporary project, and it looks the same as the plotted points - this is a visual trick that QGIS is pulling on us. It's re-drawn

the layer on the fly to match our original layer - but in reality is has transformed the file. Re-open our Part 3 project by going up to Project > Open Recent, and select Part 3. When asked to save the current project, say no and select Discard. Back in Part 3, flip from the Layers to the Browser menu, navigate down to the part3 folder. Select the folder, right click, and hit Refresh. Then drag the libraries_usa layer into the view. Flip the Browser menu back to Layers - the libraries should be drawn on top, and they match our underlying layers.



6. *View the attribute table.* Select the libraries_usa layer in the LP, right click and open the attribute table, to take a look at what's there. You should see all of the data that's affiliated with the libraries. Close the table when you're finished. Save your project.

3.3.2 Commentary

Text Files

Delimited text files, along with spreadsheets, are supported by QGIS as stand-alone tabular data files. You can add them to QGIS and join them to spatial data. Spreadsheets are usually better for table joins, while text files are more suitable for plotting coordinate data to create spatial data. You use the Add delimited text menu and specify whether the data has coordinates (to plot them) or not (because you're going to use the table for a join operation instead).

A text file is a plain document format that is often used for storing and sharing data. Since it is relatively simple and contains no formatting, it is cross platform and historically stable. The attributes of each record are separated by a delimiter to indicate different fields. This allows spreadsheet and database programs to parse the text file into columns when you open or import it into that software. Common delimiters include commas, tabs, and pipes. Files can be saved with the extension .txt or .csv. CSV (comma separated values) files are text files that use commas as delimiters.

While they are stable, cross-platform, easy to create and thus very common, the disadvantage of text files is that the fields are not associated with a specific data type, unlike a spreadsheet where a field can be designated as a string, integer, real, or other type. QGIS will try to make correct assumptions when importing the data and assigning it a data type, but it's possible that it won't. The option to specify data types when importing text and CSV data was added with QGIS 3.28, which greatly simplifies the process of working with this format. When importing text files you need to be careful that columns are designated correctly; strings inadvertently stored as numbers may have zeros dropped, while numbers inadvertently stored as strings cannot be treated mathematically. In the former case, leading zeros dropped from ANSI / FIPS, ISO, or ZIP Codes will be useless as identifiers, and in the latter case numbers stored as text can't be classified numerically or used mathematically.

Other considerations: to preserve values as numbers, make sure you don't have any stray characters or footnotes stored in your numeric column. In order to preserve values as text, a common convention is to surround strings by double or single quotes and upon import they will be recognized as text. Also, values that contain a character that's used as a delimiter (like 'Providence, RI') can be preserved as a single string using quotes (to prevent it from being incorrectly split into two values). Text or CSV files that you download from the web may or may not use this convention. Excel generally does a poor job at working with the CSV format - in order to preserve data when opening a CSV in Excel, always open a blank spreadsheet and import the data using the From Text tool on the Data ribbon (rather than clicking on the file or opening it within Excel). Lastly, QGIS does recognize CSVT helper files, which contain instructions that specify types for a corresponding CSV file. Search the Internet for examples.

Coordinate Systems for Coordinate Data

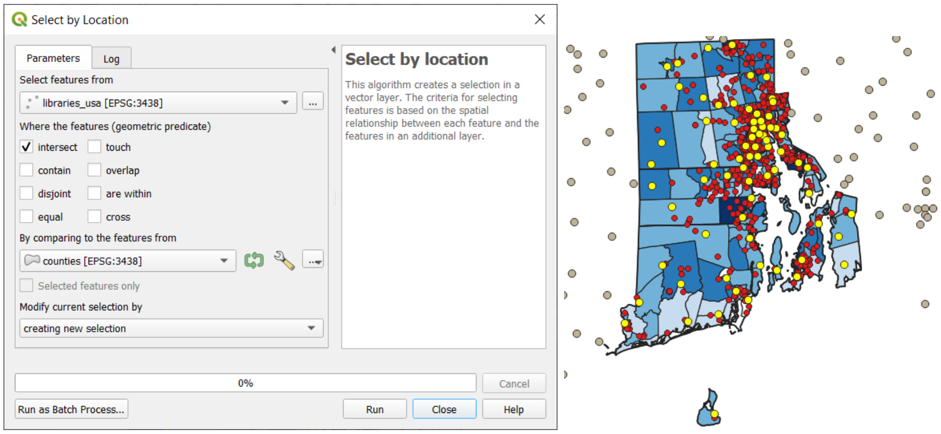
We will cover coordinate reference systems and map projections in chapter 4, but some explanation is needed here. Most coordinate data is published with longitude and latitude coordinates, but longitude and latitude coordinates are not absolutes. When we measure temperature in degrees, we need to know what system they are in (Fahrenheit, Celsius, Kelvin) before we can interpret them. Likewise, in GIS we need to know what system coordinates are in to plot them. The most common, modern system is WGS 84, and if you don't know what system your coordinates are in this is a safe guess (assuming the data is modern). The public library dataset in this example was produced by the Institute of Museum and Library Services, which is part of the US federal government. Most US federal agencies publish coordinate and geospatial data in NAD83, which is a system commonly used in North America. When in doubt, consult documentation and metadata that accompanies the dataset, or follow the general practices for your part of the world (i.e. if a given country or region generally uses a particular system). If the numbers of the coordinates are large (more than three digits to the left of the decimal place) these are probably not longitude and latitude, but represent meters or feet in a projected coordinate system.

3.4 Selecting, Filtering, and Deleting Features

In this section we will begin to prepare our vector files for analysis, and in doing so we'll demonstrate three different ways for creating subsets of features: selecting features to create new layers (by attribute and geographic selection), filtering features to hide them, and deleting features to permanently remove them. Specifically, we need to: pare down the national library layer to include just libraries in Rhode Island, remove libraries that are not buildings (bookmobiles), omit schools that are private schools or preschools (as the libraries must be near a K-12 public school to qualify), and select census tracts that meet the initiative's criteria.

3.4.1 Steps

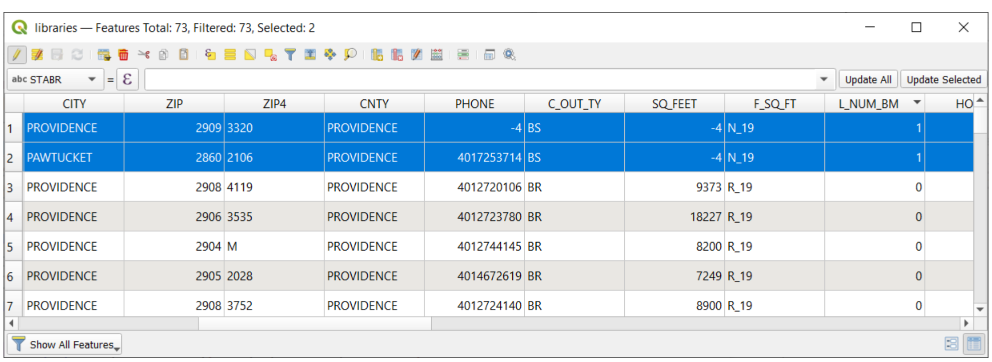
1. *Select libraries by location.* We could create a subset of libraries for Rhode Island by doing either a select by attribute (as the state code is an attribute in the table) or selecting by location. Let's demonstrate the latter. Go to Vector > Research Tools > Select by Location. Select features from the libraries_usa layer in the first dropdown, check the intersect box, and in By comparing to the features from box select counties. Keep the Creating new selection option as the default. Click Run, then close the menu, and you'll see just the RI libraries are selected.



2. *Save selection as new layer.* Select the libraries_usa layer in the LP. Right click and choose the Export - Save Selected Features As. In the Save vector layer menu, save the new layer as an ESRI shapefile. Browse and save it in your part 3 folder simply as libraries. Notice that the new file will be given the same CRS as the current layer, which is in NAD 83 / Rhode Island. Make sure to keep these two boxes checked: Save only selected features, and Add saved file to map. Hit OK.

3. *Tidy up your layers.* As we create new layers we can remove the old ones. We're finished with the libraries_usa layer; select it in the LP, right click and remove it. Drag the new libraries layer so it appears below the schools_pk12 in the LP. Change the symbology of the libraries (double click layer in LP, select symbology tab) to decrease the size of the marker from 2.0 to 1.5, and make sure the libraries have a different color from the schools. Save your project at this point.

4. *Edit the libraries to delete features.* We need to omit bookmobiles from our layer as they are not buildings. Instead of doing another selection and export, we will modify the layer in place. Select the libraries layer in the LP, right click, and open the attribute table. Scroll to the right until you see the L_NUM_BM column which indicates the number of bookmobiles (documentation for the dataset is available in the part3 folder). Click on this column heading to sort it, and click again to sort in the opposite order. There are two records with a value of 1 for bookmobiles. Hit the Edit button on the toolbar. Select these two records by clicking on the row number on the far left; hold down the Control key (command key on Mac) to select them both. On the toolbar select the Delete selected button. Confirm that you want to delete the records. Hit the Edit button again, and when prompted Save your edits. Close the table.


libraries — Features Total: 73, Filtered: 73, Selected: 2

	CITY	ZIP	ZIP4	CNTY	PHONE	C_OUT_TY	SQ_FEET	F_SQ_FT	L_NUM_BM	HO
1	PROVIDENCE	2909	3320	PROVIDENCE	-4	BS	-4	N_19	1	
2	PAWTUCKET	2860	2106	PROVIDENCE	4017253714	BS	-4	N_19	1	
3	PROVIDENCE	2908	4119	PROVIDENCE	4012720106	BR	9373	R_19	0	
4	PROVIDENCE	2906	3535	PROVIDENCE	4012723780	BR	18227	R_19	0	
5	PROVIDENCE	2904	M	PROVIDENCE	4012744145	BR	8200	R_19	0	
6	PROVIDENCE	2905	2028	PROVIDENCE	4014672619	BR	7249	R_19	0	
7	PROVIDENCE	2908	3752	PROVIDENCE	4012724140	BR	8900	R_19	0	

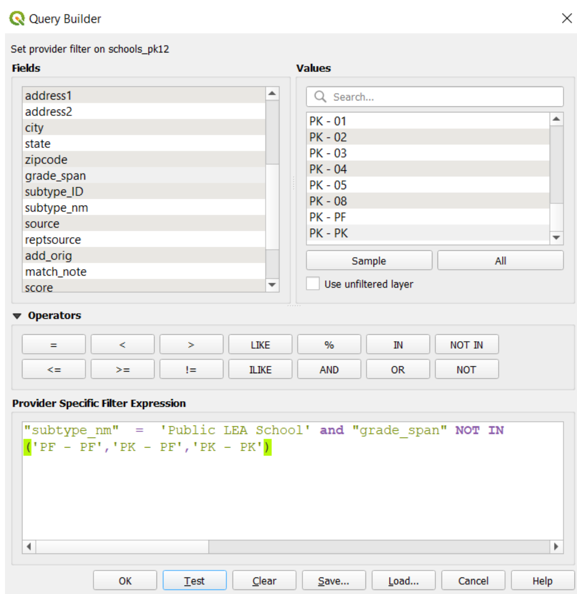
5. *Alternative for editing.* Note that you can also be in an edit mode in the map view, and access tools on the editor toolbar for creating, modifying, and deleting. You would use the Select features button to choose features.

6. *Apply filters to schools.* For the schools layer, we'll apply a filter to hide features that we're not interested in working with, without deleting them or creating something new. Select schools_pk12 in the LP, right click, and

choose Filter. Double click on the the subtype_nm field to add it to the expression. Hit the button for the equals = operator. With subtype_nm selected, click ALL under the Values box to see all possible values. Double click on Public LEA School to add it to the expression. Hit the AND button. In Fields double click grade_span, then ALL to see its Values. Hit the NOT IN button, then type an open parentheses. Add the values for PF - PF, PK - PF, and PK - PK to your expression, separate each with commas, then close parentheses. Your expression should read:

"subtype_nm" = 'Public LEA School' and "grade_span" NOT IN ('PF - PF', 'PK - PF', 'PK - PK')

Hit Test to make sure the statement is correct. Then hit OK to apply the filter. All schools that meet this criteria (they are public schools, and are not solely pre-schools) will pass the filter, while the rest will be hidden in this layer within our project, as long as this filter is applied. Save your project.

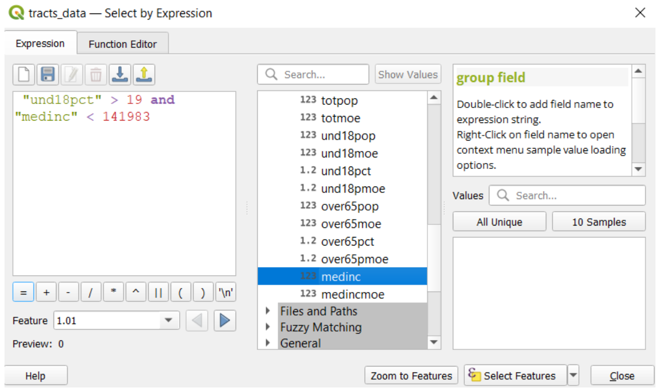


7. *Run some basic statistics.* Ultimately, we only want to consider libraries that are in or quite near census tracts that have a higher percentage of children than the state average, and that are below the 80th percentile for household income for the state. For the latter, we'll take the value directly from the census (table B19080). For the former, we can calculate this from the tracts. Hit the statistical summary button on the toolbar. You can undock the statistics window if you find it's hard to see. Select tracts_data as the layer, and und18pct as the column to summarize. We see that the mean is 19.2085 while the median is 18.6. We'll say that tracts where more than 19% of the population is under 18 will fit our criteria. Close the statistics window.

8. *Select tracts based on attributes.* Select tracts_data in the LP. Hit the Select features by expression button on the toolbar (if you don't see it, look for the Select features by values button and hit the dropdown arrow beside it). Open the Fields and Values options in the center panel. Double click on und18pct to add it to the expression. By clicking on the fields and typing in the operators and values, enter this expression:

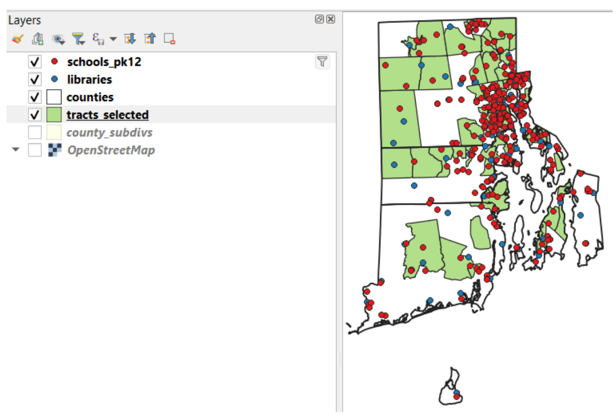
   "und18pct" > 19 AND "medinc" < 141983

   If the expression is syntactically correct, there will be no error message in the preview at the bottom of the window. Hit Select Features, then close the expression window.



9. *Save selected features as a new layer*. Right click on tracts_data in the LP and choose Export > Save Selected Features as. Hit the ellipses button to browse and save the file in the part3 folder as tracts_selected. Make sure to save only the selected feature (check box). Hit OK.

10. *Tidy up and reorder layers*. Remove the old tracts_data layer from the project. Drag the new tracts_selected layer so it appears below the counties layer. Save your project.



3.4.2 Commentary

Selecting, Filtering, Deleting

One of the strengths of GIS is the ability to perform spatial and attribute queries on features. It's pretty common that you'll download geographic data that covers an area that is wider than you need. Since GIS data is malleable, it usually makes sense to grab data for a larger area and select out just the portions you need, if you can't find a layer that consists just of the areas you want. This is something to keep in mind when you search for data on the web; in our example there wasn't a pre-existing public library dataset for Rhode Island, so we used a national one that we narrowed down.

Selecting by attribute allows us to narrow features down by category and quantitative criteria; it's also possible to narrow features down geographically if they have a geographical attribute like a country, state, or county code. Selecting by location allows us to select features based on how they geographically correspond with other features (intersect, touch, overlap, etc), which is useful if the attribute table does not contain geographical identifiers, or if layers don't have a common attribute that can be used for joining them together. The options for geographic relationships are based on specific, standardized definitions for how the interior, boundary, and exterior points of two features relate (for examples see `http://postgis.net/workshops/postgis-intro/spatial_relationships.html`). Intersect is the broadest relationship as it selects all features that share a common interior or boundary point. In some cases imperfections in a file's geometry, or differences in how two layers were created (from a different base layer) can lead to less than perfect selections and may require manual adjustment.

Whether you create selections, filter, or delete features to narrow a layer down depends on your ultimate goal; each method has benefits and drawbacks. Selecting by attribute or location allows us to create new layers from existing ones. It's quick and easy to do, the original layer is preserved if we need to return to it, and both layers can be used in any project. You do create additional files, and if you save them all it consumes space. It is possible to select features in a layer, and then in subsequent operations select additional features just from the initial selection. This reduces the number of files you have to create and thus steps you have to perform, but once the final selection is complete you won't have a saved copy of the initial selection.

Filtering is a second option; in this case we don't create any additional layers and preserve all features in the existing one, but we hide features that we know we don't want to display or include within an analysis within one specific project. If you find you are filtering the same features in multiple projects, it's probably better to create a permanent subset. Deleting features is a third option; this modifies the layer in place and permanently removes features from the file, and thus this change would carry over into other projects where this file is used. In doing so, you won't preserve the original layer and you cannot get the features back unless you saved an original copy.

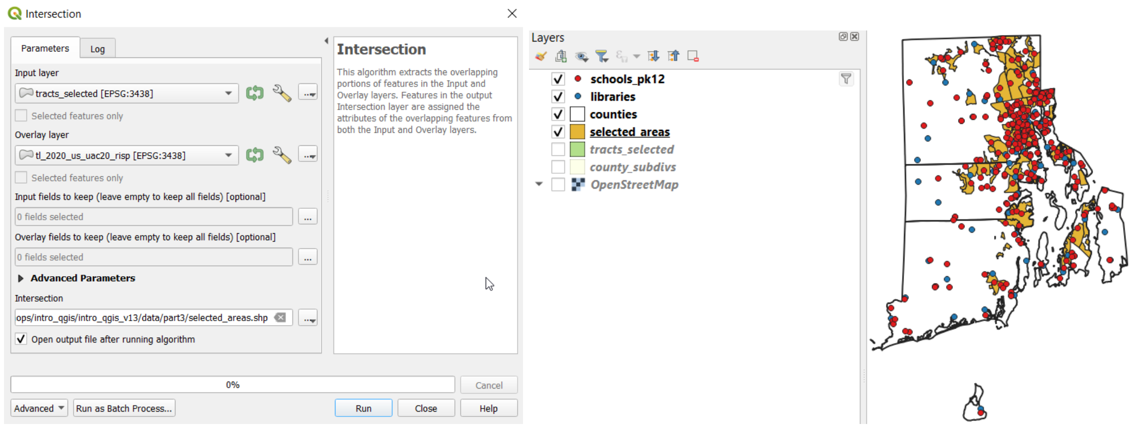
3.5 Geoprocessing

In this section you'll learn how to modify the geometry of a layer based on its spatial relationship with another layer. In the last section we selected census tracts that met our criteria, as libraries need to fall within or near those areas. One of the other criteria is that these areas also need to be urban. In this section we will identify those areas using an urban areas layer from the Census Bureau. We also need to identify libraries that are within a 1/3rd of a mile of a public school, which we can achieve by drawing buffers around the libraries.

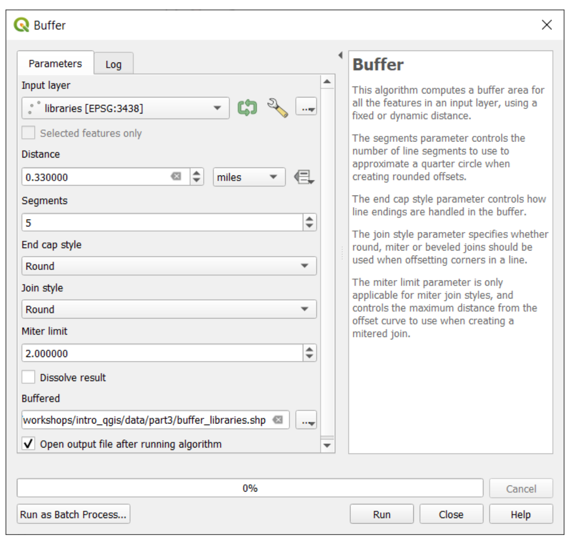
3.5.1 Steps

1. *Add urban areas layer.* Use either the Browser or the Data Manager to navigate to the part3 folder, and add the tl_2020_us_uac20_risp.shp file to the map. This is a national layer that was prepped for this example by transforming it to Rhode Island State Plane. Notice that it overlaps state boundaries; if we were to select by attribute or geography, we would end up with features that fall partially outside of the state.

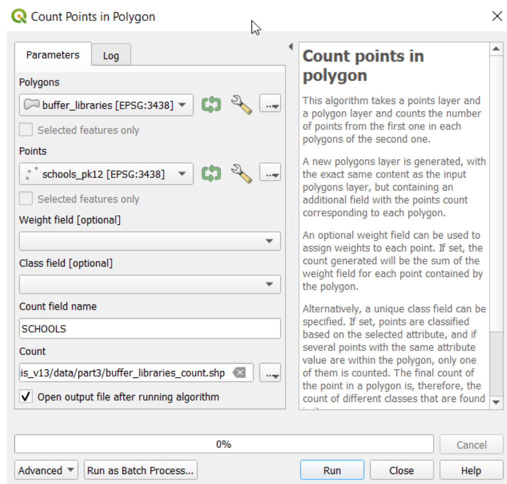
2. *Create intersection between two layers.* Let's isolate urban areas that fall within the census tracts that meet our criteria. Go to Vector > Geoprocessing Tools > Intersection. Select tracts_selected as the Input layer, tl_2020_us_uac20_risp.shp as the Overlay layer, and under Intersection browse to the part3 folder and save a new shapefile as selected_areas. Hit Run, and close the window. The new selected_areas layer consists of the areas where urban areas and tracts_selected coincide. In the LP, drag both of these layers underneath the counties layer. Remove tl_2020_us_uac20_risp.shp from the project, and uncheck tracts_selected to turn that layer off. Save your project.



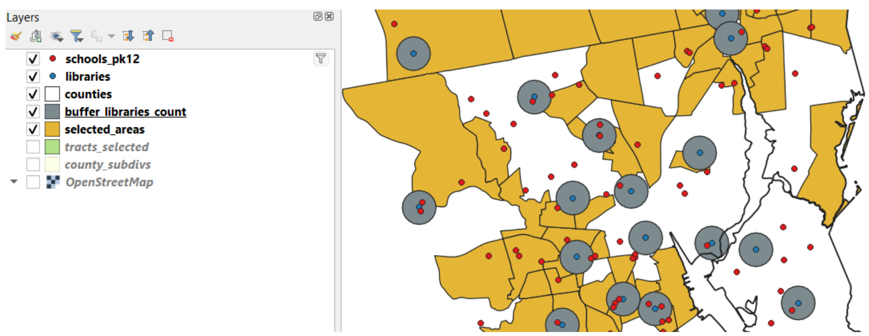
3. *Create buffers*. On the menu bar, go to Vector > Geoprocessing tools > Buffer. Specify libraries as the input layer. For the Distance, change the dropdown from feet to miles and type .33 for 1/3rd of a mile. Keep the segments and style settings the same. Under Buffered browse and save a new shapefile in your part 3 folder as buffer_libraries. Hit Run. Close the menu when finished. Zoom in a little, and you will see circular buffers drawn around each library.



4. *Count points in polygons*. Let's count the number of schools that fall within each buffer. On the menu bar, go to Vector > Analysis tools > Count points in polygons. Specify buffer_libraries as the Polygons layer and schools_pk12 as the Points layer. Under Count field name, change the name to SCHOOLS. Under Count, browse and save a new shapefile in your part 3 folder as buffer_libraries_count. Hit Run. Close the menu when finished. Open the attribute table for the new buffer_libraries_count, scroll to the right, and you'll see the SCHOOLS column with a count of the libraries.



5. *Tidy up and save.* Now that we have library buffer layer with the school count, we don't need the first one we created. Select the first buffer_libraries layer in the LP, right click and remove it. Make sure the buffer_libraries_count appears in the LP just below the libraries layer. Save your project.

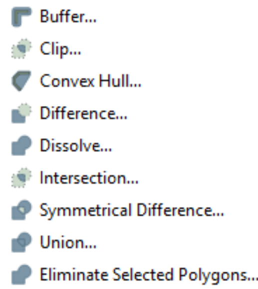


3.5.2 Commentary

Geoprocessing and Distance Measures

Geoprocessing is a GIS operation to manipulate the underlying geometry of GIS data. In the broad sense it includes layer overlay, feature selection, data conversion, and geometry processing. In a more narrow sense that we're using here, it refers specifically to geometry processing; modifying the actual geometry (points, lines, and areas) of features and files. In contrast, the selection methods we used in the previous section simply affect the number of features in the layer, and when making selections we're selecting entire features to create a new subset of features, as opposed to changing their shapes to create wholly new features. We demonstrated Intersection and Buffering. Under the Vector menu, QGIS has the following Geoprocessing tools (running each tool creates a new layer; it does not modify existing layers):

- Buffer - creates zones around specific features at a certain distance
- Clip - cuts a layer based on the boundaries of another layer
- Convex Hull - creates the smallest possible convex polygon enclosing a group of objects
- Difference - subtracts areas of one layer based on the overlap of another layer



- Dissolve - merges features within a single layer based on common attributes in the attribute table
- Intersection - creates new layer based on the area of overlap of two layers
- Symmetrical Difference - creates new layer based on areas of two layers that do not overlap
- Union - melds two layers together into one while preserving features and attributes of both
- Eliminate Selected Polygons - merges left-over or misformed geometry with neighboring features

In addition, there are also some geoprocessing tools under the Geometry Tools menu in Vector that convert or break polygons apart into simpler features (like lines or points) and under the Data Management Tools menu (for merging many vectors into one file; the opposite of the selection / subset process). Geoprocessing for raster layers is available through the Raster menu. Lastly, several extensive collections of processing tools for both vectors and rasters are available in the Toolbox under the Processing panel.

When creating buffers, the layers you are working with must be in a projected coordinate system that uses meters or feet. You cannot perform these operations in a geographic system that uses degrees, as a degree is not a constant unit of measurement. In our example we chose to keep the boundaries of each buffer distinct, so buffer features can overlap and they retain the attribute information of the original features (libraries in this case). We need to do this in order to determine whether an individual library falls near a school. Alternatively, if we were interested in generally identifying total areas that fall within a certain distance of a library, we could check a box to dissolve the boundaries of the buffer to form one large polygon.

While appropriate for this example, a less computationally demanding alternative to creating buffers would be to use the Vector > Analysis > Distance Matrix tool to measure the distance from a feature to the closest feature, closest N features, or all features. The results could be joined back to the original vector layer in a table join (as the distance matrix would include unique IDs of origin and destination features). Additionally, a new Select Within Distance tool was added in QGIS 3.34, and is available under the Vector > Research Tools.

3.6 Final Analysis

It's not unusual that you will spend more time preparing to do an analysis compared to actually doing it! Now that our layers in place, we can identify libraries that meet all of our geographic and attribute criteria. We're going to use the library buffer count polygon layer to make selections rather than the library point layer itself. Census tract boundaries are often delineated by roads, and libraries are along roads, so we want to capture tracts that are reasonably close to a library (those just across the street) in addition to ones that contain a library.

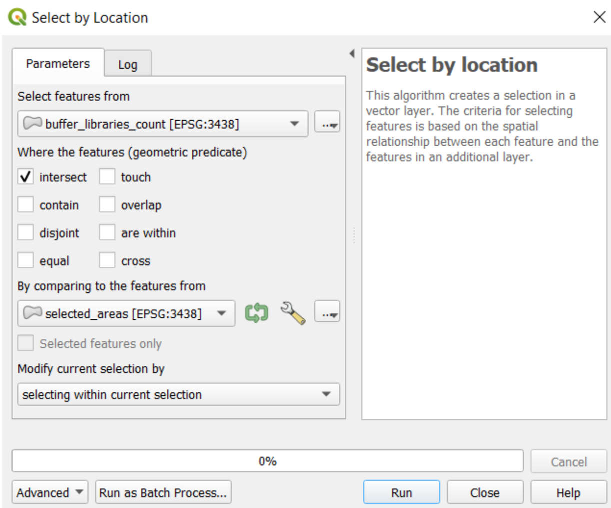
3.6.1 Steps

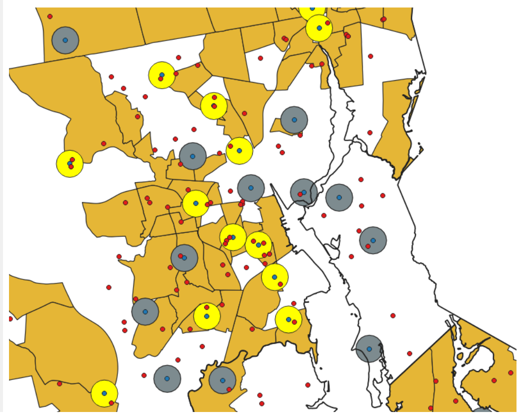
1. *Select libraries by attributes.* Select `buffer_libraries_count` in the LP to activate that layer. Then hit the select features by expression button on the toolbar. By using the Fields box in the center pane and typing in operators and values, enter the following expression to select libraries that are within 1/3rd of a mile of a school and that are sufficiently large enough to accommodate the program:

"SCHOOLS" > 0 AND "SQ_FEET" >= 2000

Hit the Select Features button, then close the window. In the status bar at the bottom of the screen, you will see 26 libraries were selected out of 71 (you could also open the attribute table to see this summary).

2. *Select libraries by location.* Go to Vector > Research Tools > Select by Location. In the Select features from dropdown choose buffer_libraries_count. Keep the intersect option checked. In By comparing to the features from dropdown select selected_areas. Under Modify current selection by, change the dropdown to Selecting within current selection. Hit Run, and close the menu. At the bottom of the screen, 16 libraries are now selected out of the 71. Pan and zoom around the map to see the results. Save your project.

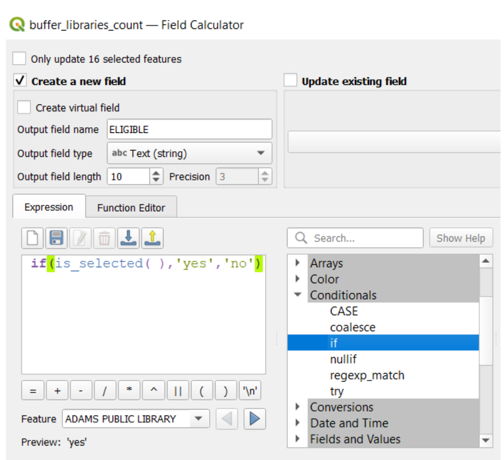




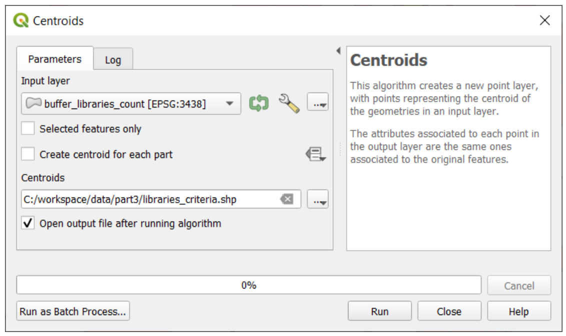
3. *Mark features as selected.* Open the attribute table for buffer_libraries_count. Hit the field calculator button. At the top, check the box to Create a new field. Uncheck the box that says Only update selected features. In the output, name the field ELIGIBLE. Designate the output type as Text (string), keep the output length as 10. Build a compound expression in the expression box; in the center pane, the if function is a conditional expression while is_selected is a record and attributes function:

if(is_selected( ),'yes','no')

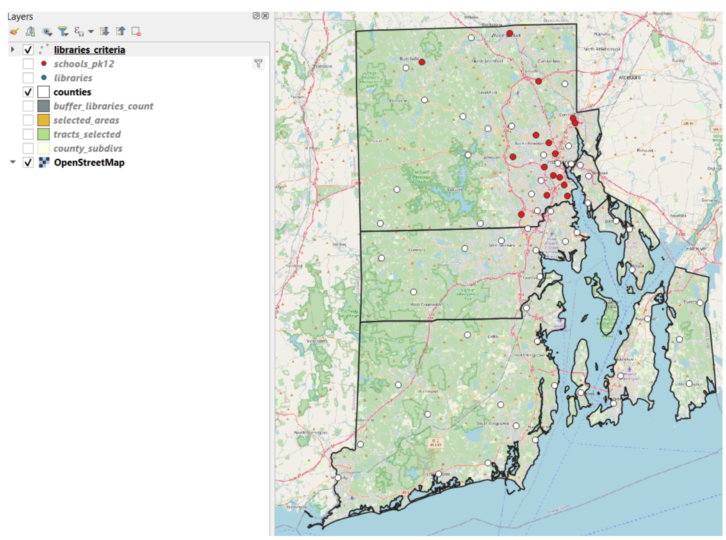
Then hit OK. This creates a new attribute column in the layer that marks selected features with 'yes' and unselected features with 'no'. Hit the edit button to exit the edit mode and save changes. Close the table.



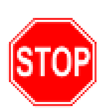
4. *Convert library buffers back to points*. We'll want to display the eligible libraries as points on our final map, as opposed to buffers. Let's convert the buffers back to points. Go to Vector > Geometry Tools > Centroids. Select buffer_libraries_count as the Input layer. Browse to the part 3 folder and save the Centroids layer as libraries_criteria. Hit Run, then close the menu. Open the attribute table for libraries_criteria, scroll to the right, and you'll see attributes from the buffers were carried over to the new point layer. (Alternatives to this step would be to do either a tabular join or a Data Management Tools > Join Attributes by Location between buffer_libraries_count and libraries).



5. *Tidy up and visualize*. Uncheck the original libraries layer as well as buffer_libraries_count, selected_areas, and schools to turn them off. Drag libraries_criteria to the top of the LP. Doubleclick on libraries_criteria and go to the Symbology tab. Change the dropdown from Single Symbol to Categorized. For Value select ELIGIBLE and hit Classify. Change the colors as you see fit; make yes stand out more than no (click the individual symbol for each category to change its style). Hit OK. Turn the OpenStreetMap layer back on to gain more context. Congratulations on completing your first GIS analysis! Save your project.



6. *Take a screenshot of your map*. Use the zoom tools and hand tool to center the map view. If you'd like, you can use the text annotation button to add a title. On the menu bar, go to Project - Import / Export - Export Map to Image. Keep the default settings, and browse to your data folder for part 3 and save the image there as map_screen. Change the Files of Type dropdown to PNG file. Click Save. If you navigate into the folder in your file explorer (outside of QGIS) you can open the file to view it. We'll cover layouts and creating formal, finished maps in the next chapter.

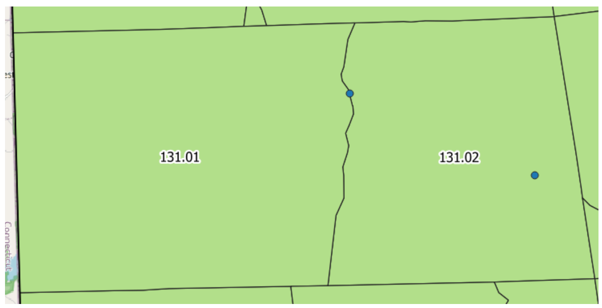


3.6.2 Commentary

Selection Criteria and Assumptions

Since the goal of our exercise is to demonstrate the capabilities and possible uses of GIS, we're not adhering to strict criteria in our selection process; the example is merely illustrative. Sometimes there will be hard criteria that you'll need to apply (if defined externally) and in other cases you must make your own decisions. The basic statistics tool for descriptive stats and the data classification tools for symbolizing your data can help guide you.

All analyses incorporate certain assumptions. In our example, we opted to select and include tracts in our analysis if they fell within 1/3rd of a mile of a library, rather than simply selecting tracts that contained a library. The rationale was that it doesn't make sense to exclude an area as being "served" by a library or as "characterizing" the area around a library when it is simply located across the boundary, which in this case means across the street (see example below). If that boundary denoted a service area beyond which a population could not be served (such as a state boundary) than it would have made sense to exclude those areas.



We also selected tracts and urban areas separately, selecting entire tracts that met the criteria for age and income, and then urban areas within those selected tracts. Our population (for age and income) may not necessarily fall within the urban areas of those tracts, but assumptions of uniformity are often made (and are more likely to be true the smaller an area is). Our hypothetical criteria didn't specify that we needed to account for this. If we had to account for this, we could have chosen a population area smaller than a tract (a block group) which would have given us geographical specificity in exchange for lower quality data. Or we could have omitted tracts where the percentage of area that was urban was below a specific threshold.

One factor that we discounted was the margin of error for the ACS estimates. These estimates represent the midpoint of a possible range of values, and the margin of error indicates the range. In selecting areas where the percentage of children was more than 19%, we excluded areas where the midpoint of the estimate fell below this value, but where the actual range may have been above it. For example, an area that was 18.5% +/- 1% would be within the 19% threshold. We did not consider this in order to focus on introducing GIS concepts, but this could be a serious point to consider in your analysis.

In short, there are many assumptions that are built into doing GIS analysis, and there are compromises you need to make. You should consider this and test different criteria to see how they impact the results. Given the number of steps we had to follow for conducting this analysis, it would be beneficial to automate the process, so we can easily adjust the model and test different criteria. QGIS has a graphical Model Designer (under Processing > Model Designer) that allows you to build a process, specifying data to input into tools and then directing the output to subsequent tools as input. See the QGIS documentation for details: `https://docs.qgis.org/3.34/en/docs/user_manual/processing/modeler.html`

Some Basic SQL

The Select features using an expression menu allows you to build complex queries for selecting features. QGIS, and most GIS packages, use the Standard Query Language (SQL) that's used when working with databases. Some tips:

- The boolean operator AND is exclusive; use it to select features that meet all of the criteria; the statement SCHOOLS > 0 AND SQ_FEET >= 2000
  will only select features where both criteria are met.
- The boolean operator OR is inclusive; use it to select features that meet one of the criteria; the statement SCHOOLS > 0 OR SQ_FEET >= 2000 will select features that meet the first criteria, or the second one, or both.
- Your statements must be explicit; for every operation you must include the field that is part of the operation: SCHOOLS > 2 AND SCHOOLS < 4 is a correct statement. SCHOOLS > 2 AND < 4 will yield an error, because you didn't specify the field for the second operator.
- Statements can be written more than one way. SCHOOLS > 3 and SCHOOLS >= 4 would yield the same result, since the number of schools is saved as an integer.
- If your query includes text rather than numbers, all text must be surrounded by 'quotes', otherwise you'll get an error. COUNTYFP='007' will return all tracts in Providence. You can also use wildcards. COUNTYFP LIKE '00%' will return all the tracts in Rhode Island, as the codes for the five counties range from '001' to '009'. COUNTYFP IN ('007','009') would return all tracts in Providence and Washington counties.

Many of the functions in the calculator are based on Python expressions; if you're not familiar with Python the calculator displays examples on how to build the expression and what it returns in the right pane, if you have an expression selected in the center pane.

File Management

As we've moved through this exercise, we've created many shapefiles along the way; every time we made a selection or performed a geoprocessing function we ended up with a new file. There are two things we should note here.

First, this can get pretty confusing. With each new file you create, it's easy to lose track of what each one represents. You can mitigate this by giving your files names that clearly indicate what they are. Documenting your progress in a notebook or text file can help you keep things straight. You may also decide to delete files that were created during the middle of the process. This is fine as long as you think you won't need to go back and re-do a step, either because the parameters of your project have changed or you've spotted an error.

Second, it's not always necessary to create a new file with every single processing step. Some menus will give you the option to select features or perform operations on features that are ALREADY selected. This allows you to work with just the features you need from one layer to create a new one, skipping the interim step of creating a new file of just the features you want to work with.

We have used underscores instead of spaces when naming files, i.e. `buffer_libraries_count.shp`. When naming files it's best practice to use underscores instead of spaces and to avoid using any punctuation in file names. This helps to insure compatibility of data across operating systems and to prevent possible errors when loading or reading data in the software. You should follow the same rules when creating folders to store data. The name of your file should reflect what it contains; you could include the geographic area it covers, the type of feature, and possibly a date or number to indicate different iterations of the data.

3.7 Considerations and Next Steps

For more practice, some things to try:

- Try modifying some of the criteria to see how the results are impacted; try a 1/4 of a mile instead of 1/3 for distance, or the 60th percentile of income (92,470) instead of the 80th percentile.
- As an offshoot of the project, measure the distance from every school to the closest public library using Vector > Analysis Tools > Distance matrix, to identify schools and potentially areas that would benefit from having a new library.

Chapter 4

Thematic Mapping

The goal of this chapter is to introduce you to coordinate reference systems and to map layout and design. Coordinate reference systems and map projections are central underlying components of GIS; these systems are what allow you to overlay and interact with GIS data from any source. Cartographic representation and design encompasses principles that pre-date GIS, and are important to grasp in order to create maps that are effective for communicating data.

The goal of this particular exercise is to create a stand-alone thematic map to show voter participation by state in the November 2020 elections in the United States. The data we'll use was collected as part of the Current Population Survey published by the US Census Bureau at `http://www.census.gov/topics/public-sector/voting.html`.

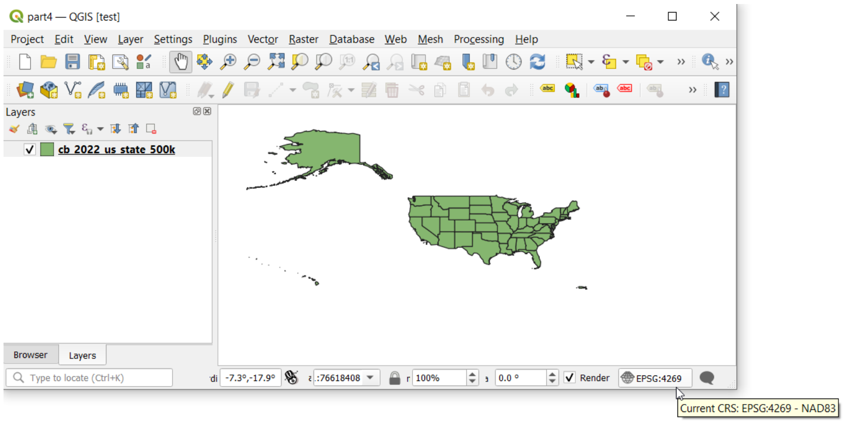
4.1 Transforming Map Projections

This section will show you how to transform a file from one coordinate reference system (CRS) to another, and will generally cover what CRS are and how they work (both in general and in QGIS in particular). Choosing a CRS for your layers is of critical importance; all layers in a project need to share the same system in order to work together, and the choice of a system is influenced by the type of analysis you're doing and what your final map will depict.

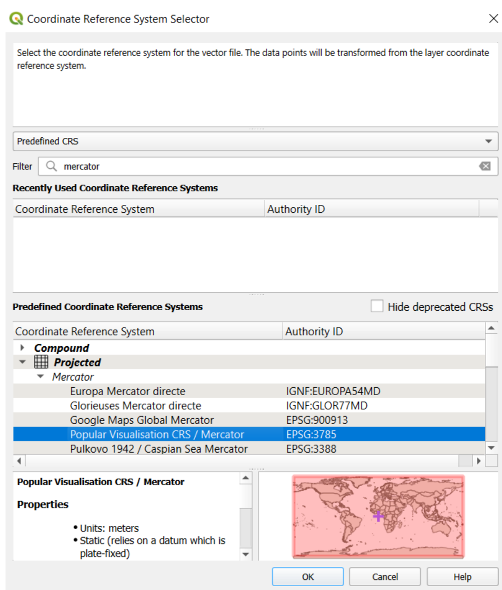
4.1.1 Steps

1. *Create a new project.* Open QGIS to an empty, blank project. Note the small box in the lower right hand corner of the screen, indicating that our default system for the map window is EPSG 4326 (a unique identifier for CR systems). Hover over the box, and we see the name of the system is WGS 84, which is the basic geographic longitude / latitude system for the world. Hit the save button. Browse to your data folder for part 4 and save the project as part4.qgz.

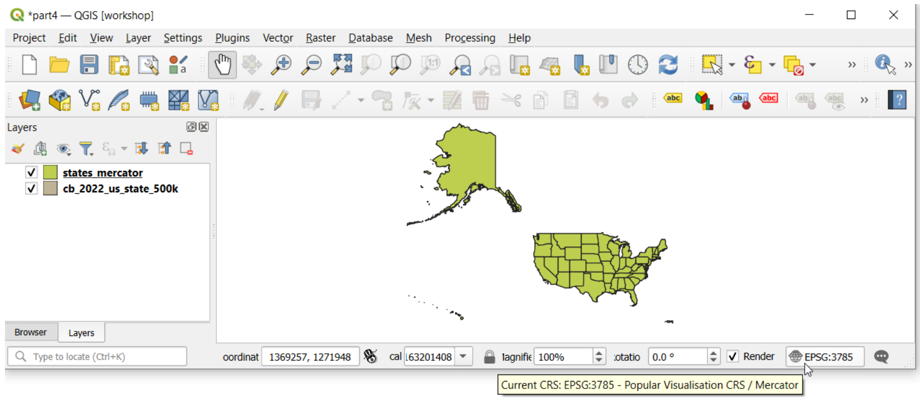
2. *Add the states shapefile.* Tab to the Browser, browse to the part 4 folder, and add the cb_2022_us_state_500k shapefile (select it in the browser and drag it into the project, or select, right click, and choose Add selected layer to canvas). Tab back to the Layers menu. Hover over the map and at the bottom of the screen in the Coordinate window note how the coordinates change. Given the size of these numbers we can tell that these are in degrees. To the right of the coordinates, hover over the text that says EPSG:4269. A little window appears that tells us the current CRS for the map window is now NAD83. Since the map window was empty, it was reset to take the system of the first layer we added. Select the states layer, right click, choose Properties, and hit the Information tab, and there is a section that indicates the CRS of the layer. Save your project.



3. *Transform the projection.* Let's transform the layer to something that's more suitable for a thematic map. Select the states layer in the LP. Right click and hit Export - Save Features As. In the Format dropdown choose the GeoPackage format (rather than a shapefile). Browse and save the file in your part 4 folder as states_mercator. GeoPackages have a layer name that's distinct from the file name; change the layer name to match the file: states_mercator. Next, hit the globe CRS button beside the CRS entry. In the CRS Selector window type in Mercator in the Filter Box at the top. This filters the entire CRS database by name. In the bottom window, hit the dropdowns under Projected and Mercator and select Popular Visualization CRS / Mercator (EPSG: 3785). Hit OK. Back on the the Save As menu, hit Save, then OK.

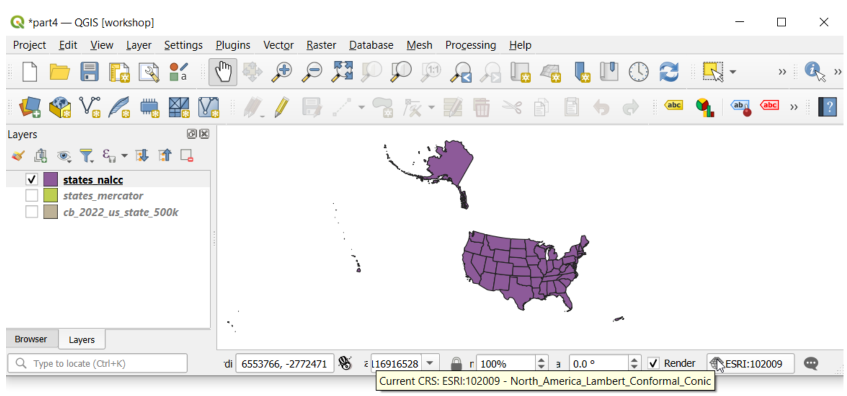


4. *Reset the CRS in the window.* Seemingly nothing has happened. Our new states layer in the Mercator projection looks exactly the same as our NAD83 states layer, and in the lower-right hand corner the EPSG Code is still 4269 for NAD83. Why is this? By default, the QGIS map window takes the CRS of the first layer that's added to the project, and will attempt to reproject all layers on the fly, so if they are in a different CRS they will draw together. To overcome this, select states_mercator in the LP, right click, and choose Layer CRS - Set Project CRS from Layer. This renders our Mercator file correctly and changes the window to EPSG 3785 (note the updated EPSG number in the lower right-hand corner). Notice the coordinates in the window at the bottom as you mouse over the map; they're in meters. Hit the Zoom to re-center the map window.



5. *Re-project the states layer again*. For practice let's try this one more time. Uncheck the states_mercator layer in the LP. Select cb_2022_us_state_500k in the LP, right click, and choose Set CRS - Set Project CRS from Layer. Right click on the layer again, and choose Export > Save Features As. Hit the Change button beside CRS. In the CRS Selector window type North America Lambert in the Filter box (if nothing appears, insert underscores: north_america_lambert). Under the Projected and Lambert windows select North America Lambert Conformal Conic (ESRI 102009) in the results and click OK. Back in the Save as window, Browse and save the file as a new geopackage in the part 4 folder as states_nalcc. Also assign states_nalcc as the Layer name. Hit OK to create the new file.

6. *Reset the projection for the project*. Uncheck the original states layer in the LP. Select states_nalcc in the LP, right click, and choose Set CRS - Set Project CRS from Layer. Then hit Zoom to Full Extent. We should see our newly projected layer in the North America Lambert Conformal Conic projection, and the EPSG Code for the window is 102009. Save your project.



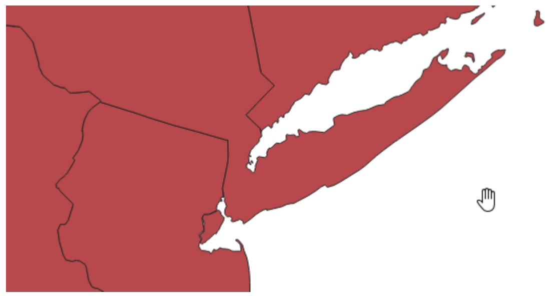
7. *Examine the differences between the systems*. The original file in NAD83 was intended for locating objects on a sphere, not two dimensions; in this format distances are preserved along certain lines, but shape and area are not which makes it look distorted. The Mercator projection is commonly used in web mapping applications, but also does not preserve area, and as a cylindrical projection it adds significant distortion to shapes in the northern hemisphere. The Lambert Conformal Conic projection was centered specifically over North America so that angles and shapes are preserved and the distortion in area is minimized (though not preserved) for the continent. That's the layer we'll use.

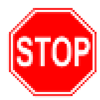
8. *Remove first two layers*. Remove states_mercator and cb_2022_us_state_500k from the project (right click - remove layer). Save your project.

9. *Coordinate transformation window*. From time to time, when you add a file to an existing project a coordinate transformation window may pop up if that file doesn't match the CRS definition of the project. You can usually just click OK to close the window, but take it as a warning and check the system of the file against the project (add the file, go to its properties, Information tab, CRS section). If it's different, consider transforming it.

10. *Alternative to Exporting*. Under Vector - Data Management Tools there is an option called Reprojecting layers. This is a simple alternative that let's you save a new layer in a new system. The Export - Save Features As option that we've been using gives you a little more flexibility for specifying additional options when saving.

11. *Inspect the states layer.* Zoom in to the northeastern US, to the area around New York City. You'll see that Manhattan and Long Island appear joined to the mainland. This shapefile is from the Census Cartographic Boundary Files, which are census TIGER files that have had their boundaries simplified so they appear less jagged at small scales (viewing the US as a whole) but are not appropriate for large scale maps (viewing a small area like the NYC metro). Hit the Zoom to full button to re-center the map window.





4.1.2 Commentary

Census TIGER Files, Generalization, and Scale

The US Census Bureau creates and maintains legal, statistical, and administrative boundaries for all geographic areas that it publishes data for. It also creates and maintains geographic features such as water, roads, and landmarks that are used when creating boundaries. These files were originally in a vector format called TIGER which the Census now provides in shapefile format. The TIGER files are in the public domain and can be downloaded from:
`https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html`

The Census Bureau makes minor updates to boundaries and issues new TIGER files each year, but major changes occur at the beginning of each decade as the decennial census is released. There are often minor changes to statistical areas (like census tracts and ZCTAs) to correct errors and improve accuracy, but generally these areas change very little until the next ten-year census. In contrast, updates to legal boundaries (like states, counties, or municipalities) are made on an annual basis. The files we've used in these exercises are from the 2021 or 2022 TIGER / Line Shapefiles, which are based on 2020 Census geography.

TIGER files represent precise legal and statistical boundaries that incorporate land and water. The Census Cartographic Boundary Files (CBF) (`https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html`) we are using in this chapter were generated from the TIGER files. CBF boundaries have been generalized to depict land areas, to smooth coastlines and borders, and to remove small islands. This makes the boundaries appear smoother and cleaner at smaller scales (national, regional, state-level), while sacrificing geographic accuracy that wouldn't be visible. The files we've been using are the least generalized, at 1:500,000. The image below illustrates the difference between TIGER and the three CBF scales.

TIGER Line File | Cart Bndy 500k | Cart Bndy 5m | Cart Bndy 20m

When choosing vector files for thematic mapping you need to make sure that the generalization for the file is appropriate for the scale you're working at. If you were creating a map of the NYC metro area, you would not want to use these boundary files as the generalizations become apparent at this larger scale and will make your maps appear inaccurate. But these files are perfectly suitable for a map of the US, as these imperfections are not visible at that scale.

You can identify whether a layer is appropriate by looking at the metadata and seeing if an optimal scale is indicated. Scale is a proportion of units of measurement on the map versus the actual distance in reality. A scale of 1:500,000 indicates that one measurement unit on the map represents 500,000 units in reality. Small scale maps cover large areas while large scale maps cover small areas; this may seem counter-intuitive, but remember that scales represent fractions: 1/50,000 is a larger number (and thus larger scale) than 1/500,000.

All of the Census TIGER files use the coordinate reference system (CRS) NAD83, which is identified with the code EPSG 4269. This is a basic longitude and latitude system that is common throughout North America. In most cases you will want to reproject this system to a projected coordinate system that's appropriate for your project.

Understanding Coordinate Reference Systems

All GIS layers are created using a specific coordinate reference system (CRS). The reason that we can take data from different sources and overlay them in GIS is because they share the same system; likewise, we can plot coordinate data and create layers because there's a coordinate system under the hood of our map window. In order for everything to work, your layers must share the same system and the map window must be defined to use that system. GIS software can be used to transform layers from one system to another. Each CRS is composed of at least three or four parts:

Spheroid or Ellipsoid: We typically imagine the earth as a perfectly round sphere, but in reality the earth is rather lumpy and uneven, with protrusions in some areas and indentations in others. Models of this true shape of the earth are called geoids. For most mapping purposes, the shape of the earth is approximated using spheroids, round three dimensional models of the earth, and ellipsoids, which represent the earth as being more oval than sphere-like in nature.

Datum: When you use a particular spheroid or ellipsoid it must be fitted or applied to the true shape of the earth, and there needs to be a method for creating a coordinate grid and accurately attaching it to that surface. Mathematically, where does one draw the prime meridian and equator on a particular spheroid in order to accurately represent their location? The instructions for attaching a spheroid / ellipsoid to the surface of the earth is called a datum.

Coordinate System: This is the reference grid used for locating places on the earth and measuring distances. Longitude and latitude is the most common system, but there are other systems with different grid cells and units of measure; for example, the Universal Transverse Mercator (UTM) system uses a distinct grid.

Collectively, when you have these three elements: a spheroid or ellipsoid, a datum, and a coordinate system, you have something called a Geographic Coordinate System (GCS), which uses a three-dimensional spherical surface to define locations on the earth. The terminology is confusing, as a coordinate system is one part of a geographic coordinate system, and some systems are named based on the datum they use. For example, WGS 84 (World Geodetic System of 1984) is the most common GCS and uses the WGS 84 spheroid, WGS 84 as a datum, and latitude and longitude as a coordinate system. WGS 84 is used by the Global Positioning System of satellites and thus by individual GPS units as a default, and is commonly used by online mapping applications. It is so common that it is often referred to a THE Geographic Coordinate System. There are other systems; in North America NAD 83 (North American Datum of 1983) is widely used by government agencies in the US and Canada. It uses GRS 1980 as a spheroid, NAD 83 as the datum, and lat and long as the coordinate system.

If you add a map projection as the fourth element to the spheroid/ellipsoid, datum, coordinate system trio, you have a projected coordinate system (PCS), which is defined on a flat two-dimensional surface:

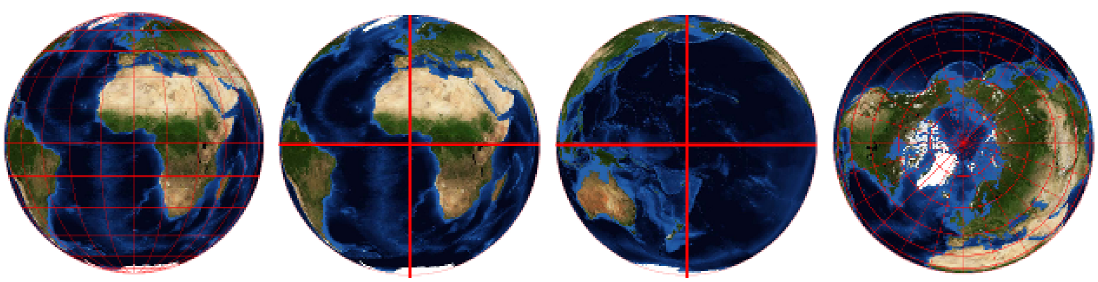
Projection: Map projections are mathematical systems for taking the three dimensional earth and transforming it to a flat two dimensional surface. There is no way to take a 3D shape and accurately represent it on a 2D surface, so map projections are designed to preserve one property of the earth - area, shape, or distance/direction, or are created as a compromise to make the earth appear the way we expect it to appear on a flat surface.

In most GIS software, libraries of GCS and PCS system definitions are stored or organized separately, under their own menus or tabs.

Longitude and Latitude

The most common coordinate system is longitude and latitude, a grid system that covers the earth and uses a unit of measurement called a degree. Lines of latitude, called parallels, run east-west. The origin of latitude is the equator, which is zero degrees latitude. The equator bisects the earth and along this line there are twelve hours of daylight and twelve hours of darkness each day, throughout the year. Lines of latitude run 90 degrees to the north pole and 90 degrees to the south pole. One degree of latitude is equal to approximately sixty-nine miles, and since they are parallel lines they never converge.

Lines of longitude, called meridians, run north-south. Unlike the equator, which is the defacto line of latitude based on natural phenomena, the selection of an origin for longitude is arbitrary. The Prime Meridian, zero degrees longitude, was designated as the origin parallel in the 19th century. It runs through the center of the astronomical observatory in Greenwich, UK. There are 180 degrees of longitude to the east and to the west of the prime meridian. The meridian that is opposite the prime meridian on the far side of the globe, 180 degrees longitude, is the International Date Line (approximately). Unlike latitude, longitude converges at the poles to a single point at zero degrees. Since lines of longitude converge there isn't a uniform distance between them - the distance decreases as you move away from the equator. At the equator one degree of longitude is approximately 69 miles across, while at the poles it is zero miles.



There are two conventions for recording coordinates: in degrees, minutes, and seconds (DMS) or as decimal degrees (DEC). Take a look at the following coordinates for Philadelphia, PA from the USGS GNIS gazetteer:

39 deg 57' 08" N 75 deg 9' 50" W (DMS)

39.952335, -75.163789 (DEC)

The DMS notation is similar to the notation for telling time - there are 60 minutes in one degree and 60 seconds in one minute. DEC notation is preferable for computer processing; if you're plotting coordinates in GIS they should be in DEC. In DEC, latitudes south of the equator and longitude west from the prime meridian to the international date line are recorded as negative numbers. It is crucial that DEC coordinates indicate direction, otherwise you'll be confusing your point with a different place:

39.952335, -75.163789 is Philadelphia, PA USA

39.952335, 75.163789 is a remote area in western China near the Kyrgyzstan border

In a coordinate pair, longitude is always the X coordinate and latitude is the Y coordinate. This is confusing, as latitude is often expressed as the first number in the pair, i.e. lat, long, as in the examples listed above. In this example, if we plotted latitude as X and longitude as Y this point would appear in Antarctica, directly south of Africa, instead of in Philadelphia.

Map Projections

Most people today would agree that the earth is round. Most maps, whether they're on paper or a computer screen, are flat. When you take a three dimensional sphere and flatten it to two dimensions, you get fair amount of distortion. Imagine removing the peel from an orange and laying it out flat - you can't do it without tearing the peel. A map projection is a method for taking the three dimensional earth and transforming it to a flat surface.

For a nice overview, visit `http://www.radicalcartography.net/?projectionref` - Radical Cartography's projection page and note the common projections (marked in pink). Projections can be classified based on how the grid is applied to the earth's surface - a grid laid flat on top (azimuthal), wrapped as a cone on the top half of the earth (conical), wrapped around the earth as a cylinder (cylindrical), etc. They can also be organized based on which property they preserve:

Area (Equal-Area) - areas that are the same size on the globe appear as the same size on a map. Examples: Mollweide projection for the earth, Albers Equal Area for continents.

Shape (Conformal) - preserves angular relationships and shapes for small to medium areas (but distortion of shape occurs for larger areas). Examples: Mercator for the world, Lambert Conformal for continents.

Distance (Equidistant) - maintains accurate distances from the center of the projection along specific lines; a straight line on the map will give you the shortest distance between two points, the same distance as a great circle on a globe. The Geographic Projection, also known as Plate Carree or Equirectangular, is the most common.

Direction (Azimuthal) - maintains accurate directions (and thus angular relationships) from a given central point. Azimuthal Equidistant and Gnomic are examples.

Other projections:

Interruptions - these projections show tears in the earth's surface and try to mitigate them to create something readable. Goode's Homolosine is good for showing land areas, but poor for showing oceans (as these are interrupted).

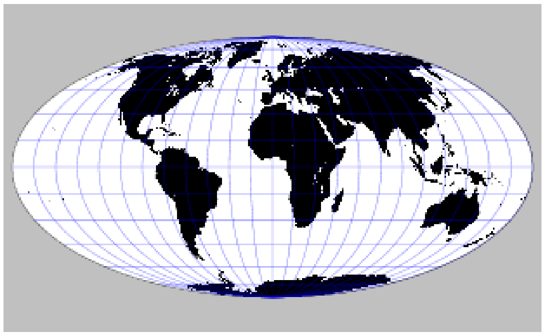
Compromises - these projections don't preserve any quality of the earth exactly, but they compromise to make a map of the earth that "looks right". Good compromise projections of the earth include Robinson and Winkel Tripel.

You can compare maps that use different projections to get a sense for how they distort different areas (in particular, observe Greenland). Common map projections for the world for general reference or thematic use include Robinson, Mollweide, Goode Homolosine, and Winkel Tripel. In general, projections that appear oval-like, showing the curvature of the earth at the edges, are best for general or thematic use.

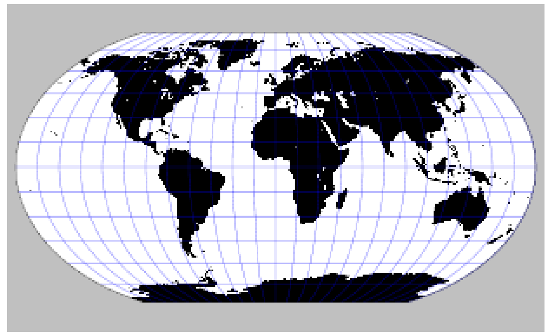
Every continent and country has a preferred map projection or set of projections that is appropriate for each area based on its size and shape. Look at atlases or pre-existing maps to get an idea of what these are. Albers Equal Area, Lambert Equal Area, and Lambert Conformal are common and are adjusted to focus on specific continents or countries. Orthographic projections are used to map polar areas.

CRS Definitions

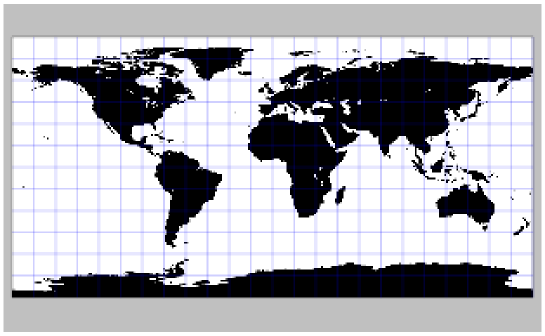
Geographic reference systems have been classified with codes, which makes them easier to identify and retrieve. QGIS uses a CRS library called the European Petroleum Services Group (EPSG). This library originally contained most of the primary GCS systems, such as WGS84 and NAD83, and local PCS systems like State Plane. For example, EPSG 4269 is the code for NAD83, and EPSG 4326 is the code for WGS84. The original EPSG library lacks most of the PCS



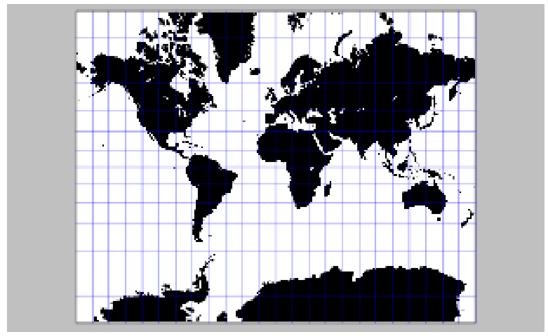
Mollweide



Robinson



GCS (Equirectangular)



Mercator

systems for global and continental map projections (like Mollweide, Robinson, and North America Lambert Conformal Conic), but the QGIS developers have augmented the library to include these projections from ESRI. A brief list of common projections and definitions is included in the appendix of this tutorial.

Several formats have been created for recording the definition of a CRS (alternatively known as spatial reference systems or SRS). There's the Open Geospatial Consortium's Well-Known Text Format (OGC WKT) (shown below), the Proj4 format, and .prj file format created by ESRI. The major difference is that the OGC format includes the EPSG codes for each system, while the ESRI format does not. To look up CRS information, you can use this website: `https://epsg.io/`. In the WKT definition, you'll see the elements that make up the CRS (projection, datum, spheroid) as well as units of measurement and origin information:

```
PROJCS["North_America_Lambert_Conformal_Conic",
    GEOGCS["GCS_North_American_1983",
        DATUM["North_American_Datum_1983",
            SPHEROID["GRS_1980",6378137,298.257222101]],
        PRIMEM["Greenwich",0],
        UNIT["Degree",0.017453292519943295]],
    PROJECTION["Lambert_Conformal_Conic_2SP"],
    PARAMETER["False_Easting",0],
    PARAMETER["False_Northing",0],
    PARAMETER["Central_Meridian",-96],
    PARAMETER["Standard_Parallel_1",20],
    PARAMETER["Standard_Parallel_2",60],
    PARAMETER["Latitude_Of_Origin",40],
    UNIT["Meter",1],
    AUTHORITY["EPSG","102009"]]
```

From this definition, we can see that North America Lambert Conformal Conic projection uses GRS 1980 as a spheroid, NAD 83 as the datum, and meters as the unit of measurement for the coordinate system. The system is centered at -96 degrees longitude, which runs north to south approximately through the center of the lower 48 states.

QGIS Projection Handling

When you open a new, blank project in QGIS the default CRS is WGS84. Then, when you add your first layer, your project automatically takes the CRS from that layer. If you add subsequent layers that don't share the same CRS, QGIS will attempt to reproject them on the fly to match your other layers. Even if the software is successful at rendering the layers, many selection and geoprocessing operations won't work as the files don't have a matching CRS, and any distance or area calculations you make could be erroneous. In this tutorial, and in general, I suggest that you know what CRS your layers are in and make sure all of the files you're using share the same CRS. Don't create projects where you're using a mix of files that are in different systems. This cuts down on confusion and helps avoid errors caused by mis-aligning data layers and using systems of measurement that don't match.

It's important to remember that, if you have added files to the map window, removed them, and then added new files that don't share the same CRS as the original ones, you need to reset the CRS of the map window. Select a layer in the LP, right click, and choose Set Layer CRS - Set Project CRS From Layer.

4.2 More Table Joins

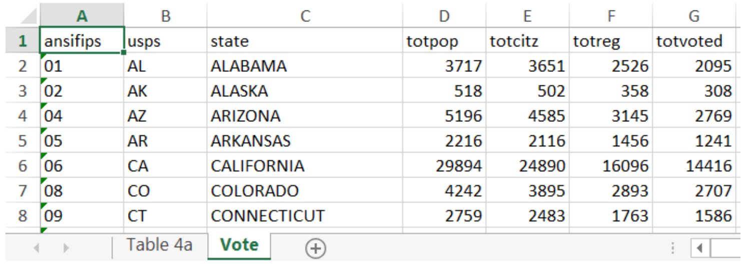
In this section you will get more practice joining a data table to a vector file, and will see before and after examples of data tables than can't and can be readily joined.

4.2.1 Steps

1. *Count features for your layer.* Select the states_nalcc layer in the LP, right click and check the Show features count box. It tells us there are 56 features. That's 50 states plus DC and the US territories.

2. *Examine the voter data table.* Minimize QGIS. Use your file browser to go to the part 4 data folder. Find the file called Table04a.xlsx. Double click on the file to open it in Excel. The workbook has two sheets. The first sheet, Table 4a, is the original data. The second sheet, Vote, has been reformatted from the original so that it's appropriate for importing and joining in QGIS. Take some time to note the differences between the two sheets. The tables show the total population, the total population who are US citizens, the total number of registered voters, and the number of people who voted in Nov 2020. All totals are rounded to thousands (i.e. the population of Alabama is recorded as 3,769 which is 3,769,000). Different proportions were calculated for each group, and since the data is sample-based a margin of error is provided for the percentages (confidence interval is 90%). Close the file when finished.

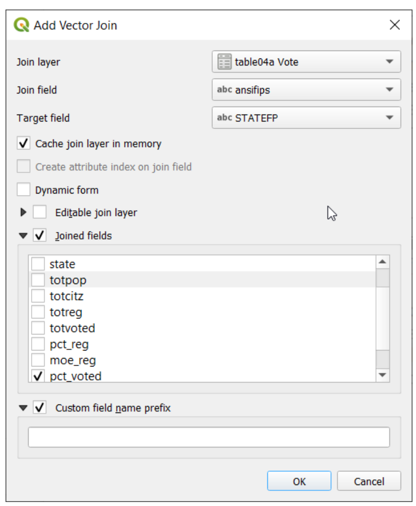
	A	B	C	D	E	F
2	Table 4a. Reported Voting and Registration for States: November 2020					
3	(In thousands)					
4						
5						Registered
6	STATE	Total population	Total citizen population	Total registered	Percent registered (Total)	Margin of error [1]
7	UNITED STATES	252,274	231,593	168,308	66.7	0.4
8	ALABAMA	3,769	3,716	2,527	67.0	3.1
9	ALASKA	528	516	383	72.6	3.2
10	ARIZONA	5,638	5,075	3,878	68.8	2.5
11	ARKANSAS	2,283	2,195	1,361	59.6	3.4



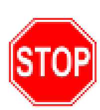
ansifips	usps	state	totpop	totcitz	totreg	totvoted
01	AL	ALABAMA	3717	3651	2526	2095
02	AK	ALASKA	518	502	358	308
04	AZ	ARIZONA	5196	4585	3145	2769
05	AR	ARKANSAS	2216	2116	1456	1241
06	CA	CALIFORNIA	29894	24890	16096	14416
08	CO	COLORADO	4242	3895	2893	2707
09	CT	CONNECTICUT	2759	2483	1763	1586

3. *Add the Excel sheet to your project.* Maximize QGIS. You can add table04a.xlsx to the project either by dragging it from the Browser or via the Data Manager and vector option (browse to the part 4 folder, if necessary change the file type dropdown to show all types of files, select Table04a.xlsx, hit Open, then Add). When prompted to choose a sheet, select just the Vote sheet that has 51 records. Hit OK. This will add the Vote table to the LP as table04a_Vote. If you select it in the LP, right click, and open the attribute table, you can verify that the data has been imported properly. Notice the column ansifips has the two-digit state codes. If you close this table and view the attribute table for the states layer, you'll see it has a column called STATEFP that contains the same code.

4. *Join the data to the shapefile.* Select states_nalcc in the LP, double click, and open the Joins tab in the properties menu. Hit the green plus sign to add a new join. table04a_Vote is the join layer, ansifips is the join field in that layer, and STATEFP is the target field in our shapefile. Instead of taking all of the columns, check the Joined Fields box to choose fields, and take these two: pct_voted and moe_voted. Check the Custom field name box, and delete the text so that the option is blank. Click OK. Then make sure to click OK again on the properties menu. Select states_nalcc, right click and open the attribute table. You'll see that the data that we've selected has been added. The vote columns for the US territories are NULL, as there were no records for them in the voting spreadsheet. Close the table. Save your project.



5. *To save as or not to save as?* If we wanted to permanently fuse the voter data table to our layer, we would need to take the next step of selecting the features in the LP and doing an Export > Save Features As, to create a new file with the data permanently attached to it. In this case, since we're simply going to symbolize and map the data, we don't need to take this extra step. The dynamic join will be saved within this specific project, and as long as the data table and states features are both present in the project, the data will remain joined. For whatever reason, if you do find that you're having problems classifying and symbolizing the joined data, then just take the extra step - create the new vector layer with the data fused to it, and any problem will likely go away.



4.2.2 Commentary

Calculated Fields

In this example, a well-formed version of the original voting table was created in advance in our Excel spreadsheet. It includes a number of derived fields that showed percent totals. QGIS does have a field calculator that will allow you to create new fields or modify existing ones, as we saw in the previous chapter. For example, you can add a new field and calculate a ratio or percent total for other values within QGIS, if your data table lacked that information. You can also use special functions that will calculate the coordinates, length, area, or perimeter of features.

Generally speaking, there are some circumstances where it may make sense to map values as whole numbers - cities by number of crimes, states by total population, counties by number of renter-occupied housing units, etc. But in each of these examples a particular place could have a higher value simply because it has more people or is a larger place. In order to make more meaningful comparisons it's often necessary to do a little math:

Percentage - (value of subset / total value)*100: (3,000 renter units / 10,000 renter units)*100 = 30% units are rentals

Rate - (value / total value) * multiplier: (400 robberies / 50,000 people)*100,000 people = 800 robberies per 100,000 people

Ratio - (value 1 / value 2): (4000 cars / 3000 people) = 1.33 cars per person

Density - (value / land area): (800,000 people / 2500 sq miles) = 320 people per sq mile

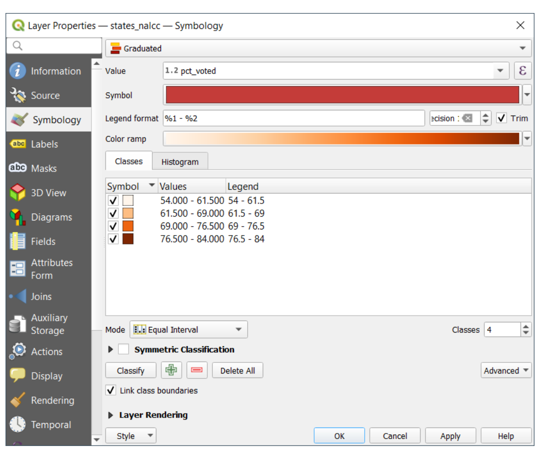
Percent Change - [(recent value / older value)-1] * 100: [(10,000 people / 9,000 people)-1] * 100 = 11.1% change

4.3 Classifying and Symbolizing Data

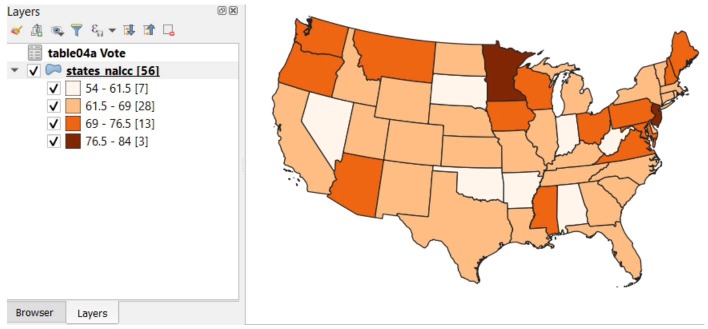
In this section you'll learn about the different methods for classifying data and the best approach for choosing color schemes to symbolize your data. These are important concepts to grasp, as they have a direct impact on how successful your map will be in communicating your data.

4.3.1 Steps

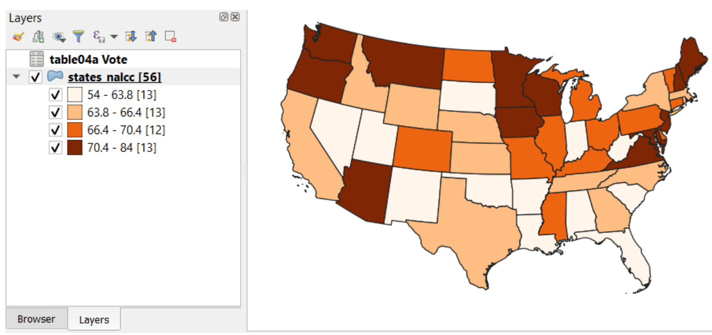
1. *Classify your data.* Select `states_nalcc` in the LP and double-click to open the Properties menu. Go to the Symbology tab. In the drop down at the top of the menu, switch the option from Single Symbol to Graduated. In the Value drop down (the field you're classifying) select pct_voted, which is the percentage of US Citizens who voted in 2020. Check the Trim box to drop trailing zeros from the values. Choose one of the default color ramps - for quantitative data with only positive values you should choose a color scheme that uses a single color value from light to dark - DO NOT choose a multi-color or random scheme. Underneath the Classes box change the number of classes from 5 to 4. Change the mode to Equal Interval. Hit the Classify button below the classification window. Then hit OK.



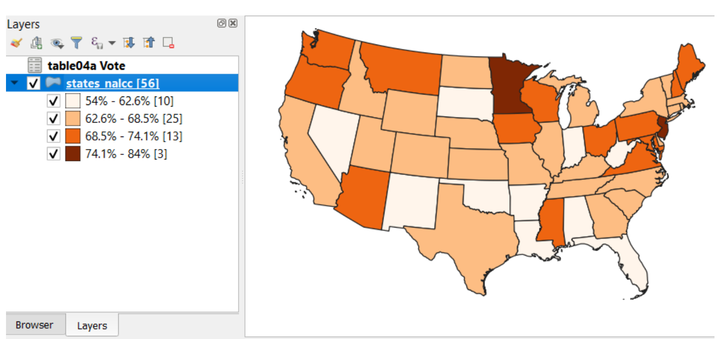
2. *Examine the Equal Intervals map.* In the LP, expand the menu for states_nalcc to see the classes. Equal intervals took our four classes of data and divided it so that each class has an equal range of values; with a min value of 54 and a max value of 84 our data has a range of 30 - divide by four and each class covers a range of approximately 7.5% from lowest to highest. Right click on states_nalcc in the LP and check the Show feature count option. You'll see the number of states in each class varies, but the range of values in each class is constant. Note that the features for the territories do not appear, as their values are null.



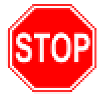
3. *Map data using Equal Counts.* However, we could use an alternate classification method called equal counts (also referred to as quantiles). Double click on the states_nalcc layer to go back to the Symbology tab under the Properties menu. Change the classification mode to Equal Counts (Quantile) and hit Classify. Hit OK to re-map your data in this scheme, and take a look at the result. Compared to the equal intervals map, equal counts shows us a greater range of colors since each class has the same number of features; data is divided into classes that have an equal number of data points. Since we have 51 data points we have about 13 states in each class sorted from low to high, as you can see in the feature count.



4. *Map data using Natural Breaks.* We have another option. Double click on the states_nalcc layer to go back to the Symbology tab under the Properties menu. Change the classification mode to Natural Breaks (Jenks). In the Legend format box, put a percent symbol % after the numbers 1 and 2, so this symbol appears in the legend. Hit OK to re-map your data in this scheme. The natural breaks method classifies data based on the location of clusters of values, or conversely in gaps or breaks in the data range, which is less arbitrary than equal intervals or counts.



5. *Save your project.* At this point save your project. For our map we'll stick with the natural breaks method, but read the commentary below for an explanation of each method and it's advantages and disadvantages.



4.3.2 Commentary

Data Classification and Color Schemes

The purpose of a thematic map is to communicate a message about the data. If a map uses too few classes, then the data is too generalized and meaningful patterns can be hidden. If a map uses too many classes, then a pattern becomes difficult to detect because there is too much detail. It is difficult for the human eye to distinguish between too many colors or variations of color. Generally speaking, it is a good idea to use 3 to 6 classes, and ideally 4 or 5. When choosing the number of classes you should consider the number of data points, the range of the data, the purpose of the map, and the color choice based on the output. While a certain number and range of colors may look good on a color printed map, they may appear washed out if the map is shown on a projector or blurred together if printed in black and white. You should design with the final output in mind.

After ranking the data from lowest to highest values, there are a number of classification methods:

Equal Interval - each class has the same range of data values. Easily understood by map readers, but does not account for data distribution and can result in categories with few or even no values.

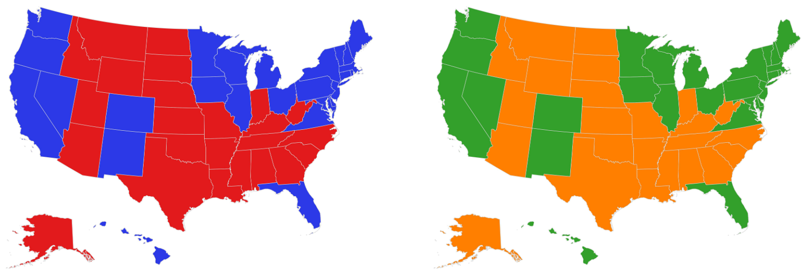
Equal Counts - each class has the same number of data points. Always produces distinct map patterns, but can be arbitrary and often creates categories that have an inconsistent range of values.

Natural Breaks - classes are created based on the location of gaps in the data. Since the data is divided based on its distribution it is good for distinguishing patterns, but like the Equal Intervals method it is sensitive to outliers.

Unique / Manual - classes created based on some external criteria. Should only be used when justified, otherwise the classification is completely arbitrary.

It's often necessary to make some common sense adjustments to any classification scheme, such as creating unique classes for values of zero or missing values, and adjusting classes so they don't contain a mix of negative and positive values. In QGIS you have the ability to adjust classes or create manual classes. To do this, you classify the data using one of the standard methods in the Symbology tab for the layer, then select the class that you want to change and double click on the range. You'll be able to type the values in by hand.

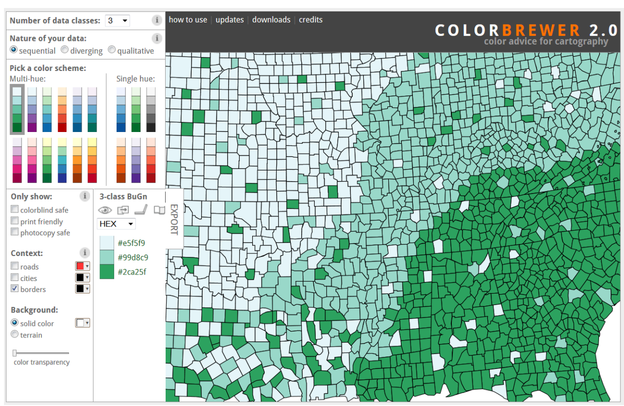
Color schemes for displaying quantitative values on shaded area maps should show a logical progression of values. The progression from light to dark helps convey the change in data values from low to high, and most map readers can infer this without even looking at the Layers Panel. Creating a mixed, fruit salad of colors will defeat this natural inference and will confuse the map reader - so don't do it. When comparing qualitative values (categorical data instead of ranges of values), a map should use colors that reflect those values. For example, it makes sense to use reds and blues to show which political party a state voted for, as these colors have become associated with the US political process. The average American will instantly understand what this map is about. Depicting the same data with greens and yellows doesn't make much sense, and results in confusion. Conversely, in our voter participation map we may want to avoid reds and blues, so we don't give viewers the wrong impression that the map depicts votes by party.



While we're not considering it for this exercise, the unit of geography used to map phenomena can profoundly affect the interpretation of a distribution or pattern and the ultimate message that your map sends. Mapping populations of states is fine if you are interested in seeing which ones have the most people. But these maps tell you very little about how the population is distributed across the country, since there is considerable variation in the concentration of people in each state. Using a smaller unit of geography, like a county, can give you a better idea of the distribution of the population. Oftentimes you'll be limited to using certain geographic units based on the availability of the data, making it necessary to compromise.

ColorBrewer

ColorBrewer is an online tool for choosing good color schemes for thematic maps. QGIS has integrated many of the schemes from ColorBrewer, and you can access them by hitting the drop down Color Ramp drop down menu in the Style tab, choosing the New Color Ramp option at the bottom, and selecting ColorBrewer. It's still worth visiting the site at `https://colorbrewer2.org/` for color selection advice. The tool let's you choose the number of classes and class options like sequential (for quantitative data), categorical (for nominal or qualitative data), and others. You also have the ability to filter color schemes based on desired output. In the lower-right hand corner of the map, you can click on a scorecard that shows whether your choice is ideal for the color blind, color printing, photocopying, and viewing on an LCD screen. You should always choose color schemes based on what your final output format will be.



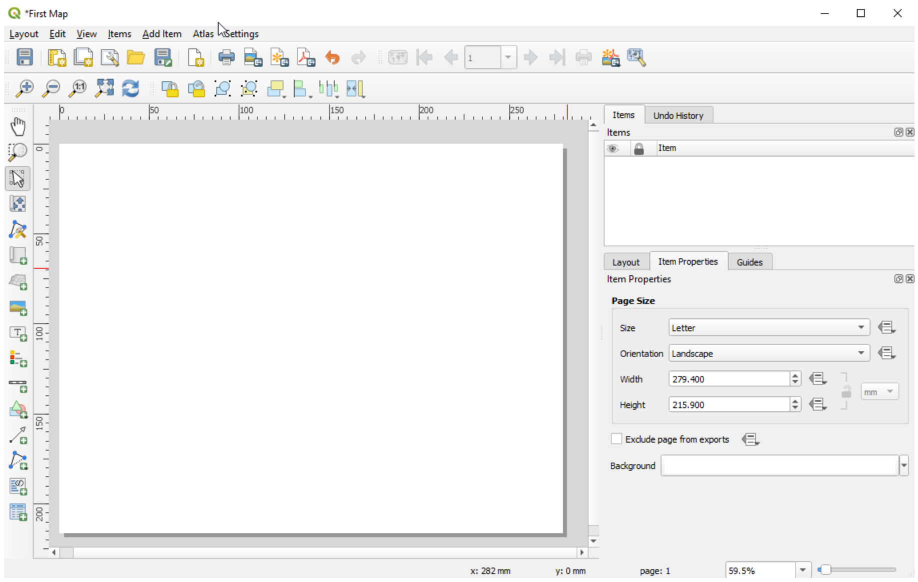
4.4 Designing Maps

In this section you'll learn how to create a finished map that includes typical map elements: legend, title, scale bar, and source information.

4.4.1 Steps

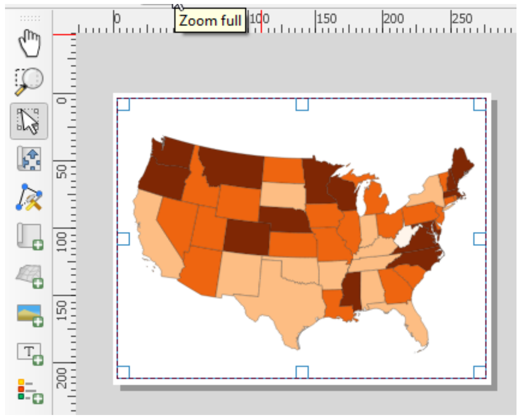
1. *Adjust zoom.* Use the  Zoom in button and draw a box around the lower 48 states, so they fit perfectly within the map window. This will help insure that our map will initially be well placed in the print composer.

2. *Set the environment for the print layout.* Hit the New Print Layout to enter the print layout screen, and when prompted give the composition a title called First Map. Right click on the page and select Page Properties. On the Item Properties tab that appears on the right, change the Page Size from A4 to Letter (the standard North American paper size). The initial layout and item properties tabs provide you with options for the map canvas as a whole. Once you add individual items (a map, label, legend, etc) and select them, these tabs will change to reflect the unique properties of that item. Each tab has collapsible menus for editing various elements.

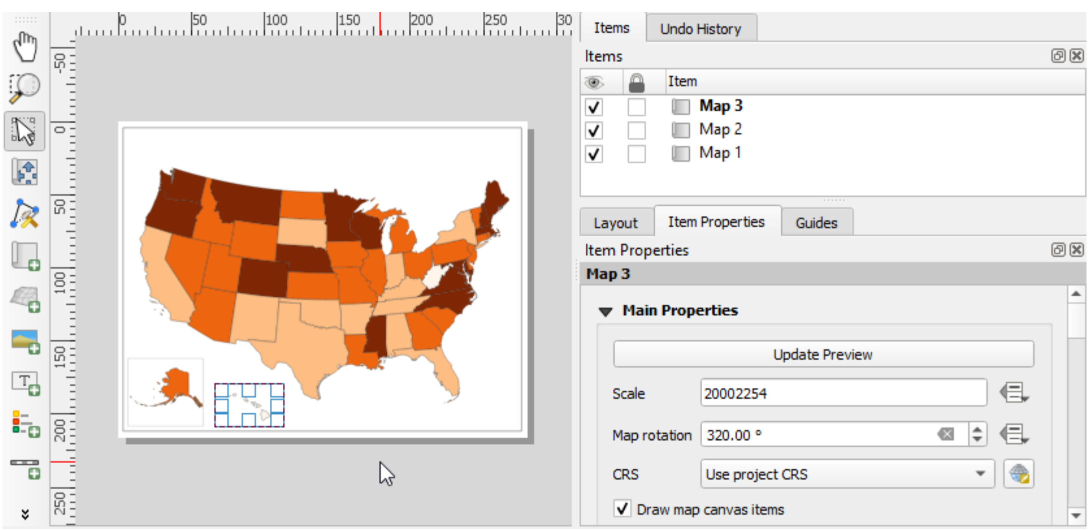


3. *Add your map and configure zoom.* Hit the add new map button in the toolbar. Then draw a box on the map canvas (click on upper-left hand corner, hold down left mouse button, drag box), leaving an even amount of space on each side so there is a gap between the map and the edge of the canvas. If you don't get it right on the first try, you can always hover over an edge of the map, hold down the left mouse, and drag the edge to change the size. Or, to shift the entire map on the page, use the select move button. This button moves the entire map box. To shift the geography *inside* the map box, use the adjacent move item content button. Move the map around so that the lower 48 states are roughly centered in the box. With the move item button selected, you can also change the zoom of the map by using the mouse wheel, or by clicking on the Item properties tab on the right and experimenting with the scale under the Main properties menu - a lower number will zoom in and a higher number will zoom out. Above the Scale in the Main Properties, there is also a series of small buttons that allow you to adjust the map zoom in the layout to match the map zoom in the Map View. In Item properties, scroll down to the bottom and check the frame box to add a frame around your map.

4. Experiment with the canvas zoom. The regular zoom buttons on the toolbar will NOT effect the zoom of the geography; these zoom buttons just zoom you closer and further from the map canvas, similar to taking a piece of paper and holding it closer or further from your face. Experiment with them and see. If your map looks blurry from resizing a window, just hit the refresh button. When you're finished, with the map selected go to the Item properties tab, and check the box beside Frame to turn the map frame on. Save your project.

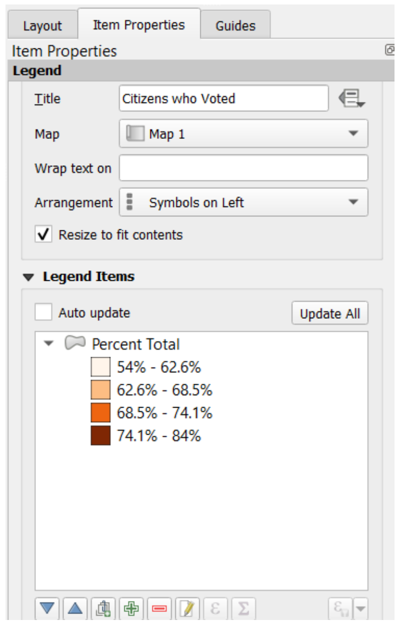


5. *Add additional maps for Alaska and Hawaii.* Given the vast distances between the lower 48 states, Alaska, and Hawaii, it doesn't make sense to include them in the same map window at the same scale. In most maps of the US, Alaska and Hawaii appear in separate boxes so that optimal scale can be achieved for all three areas; we'll do the same with our map. Hit the add map button and draw a smaller box in the lower left hand corner. Use the move item button to shift the focus of the map to Alaska, and with this button selected use the map wheel to change the zoom. If you have trouble getting the zoom "right"; with the new Alaska map box selected, open the Item properties tab on the right. Under Main properties watch how the scale changes as you zoom in and out with the mouse wheel. In the scale box type in a number that's somewhere in-between. Alternatively, you could minimize the composer, and back out in the data view zoom in to Alaska so it's centered in the view. Then return to the composer, and just above the Main Properties, click the Set to map canvas extent button. Right below the scale in the menu is Map rotation, which is currently set to 0. You can type values here to rotate the items in the map from 0 to 359 degrees clockwise. Since Alaska looks a little skewed (since we're using a map projection for the whole continent and AK is on the edge) change the rotation to 330 to straighten Alaska out (see commentary for alternatives). Under Item properties, scroll down to the bottom and turn on the Frame for the box but decrease the thickness. Once you're finished, repeat the same steps for Hawaii: add another map, zoom in to focus on the main eight islands, rotate it by 320, and turn on the frame. Save your project.



6. *Add a legend.* Hit the add new vector legend button and click on the lower right-hand corner of the map, then click OK. With the legend selected, go to the Item Properties tab and the Main properties menu. For Title type Citizens who Voted. In the Legend items box, uncheck the Auto update box. Select the Vote table, and hit the red minus sign to remove it. Then select the `states_nalcc` layer, hit the edit legend button, change the name to Percent Total, then hit the back arrow button. Then select the Percent Total title and hit the Sigma button

to turn the feature count off. Then, under Fonts, hit the dropdown under the Title font and change it from 16 to 14. Scroll down to the bottom of the Item properties menu and turn the Frame on, but decrease the thickness. The final step is to move the legend to an ideal position in the corner of the map (which may require you to shift the map around a bit).



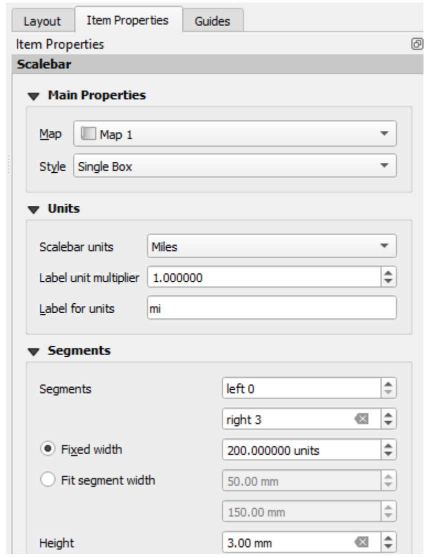
7. *Add a title.* Hit the add label button. Click on the top of the map, click OK, and a generic label is added. In the Main properties under the Item properties tab, change the default label to Voter Participation in the 2020 Election. Under the Appearance menu change the font to 20 using the font drop down. Click on the label in the map, and using the select move button, move the label to the top center of the map, and expand the size of the label box so the title appears on one line.

8. *Add a label with source information.* Hit the add label button. Click on the bottom of the map to add the generic label. In the Label menu on the Item tab, change the label to read: Source: US Census Bureau CPS, Voting and Registration in the Election of November 2020 - Detailed Tables. Change the font to size 8. Click on the label in the map, and using the select move button, move the label to the bottom center of the map, and expand the size of the label box so the text appears on one line.

9. *Add a label with author information.* Repeat the same step above to add a label with your information - Map created by [insert your name / organization] [insert date]. Move this label underneath the source label.

10. *Add a north arrow.* Hit the add north arrow button. Draw a box somewhere to the right of the US in the map and hit OK. This creates a default arrow. Move the arrow around on the map to get it centered, and resize it to make it a bit smaller. If you want to choose a different arrow style, in the Item properties tab look under the Picture - SVG Browser options and the arrows subfolder.

11. *Add a scale bar.* Hit the add scale bar button, click on the map and hit OK. In the Item properties tab for the scale bar under Main properties verify the Map is set to Map 1 (the map for the lower 48 states). Under the Units menu change the units from kilometers to miles. Under Segments increase the right segments to 3. Under Display change the font size to 10. Use the select move button to position the scale bar on your map. If the bar is too short or long, you can modify it by changing the Fixed width unit value. For example, if you want a bar length of 600 miles, and there are 3 bar segments, each segment will be 200 miles. Type this value in the Fixed width box under segments. Save your project.

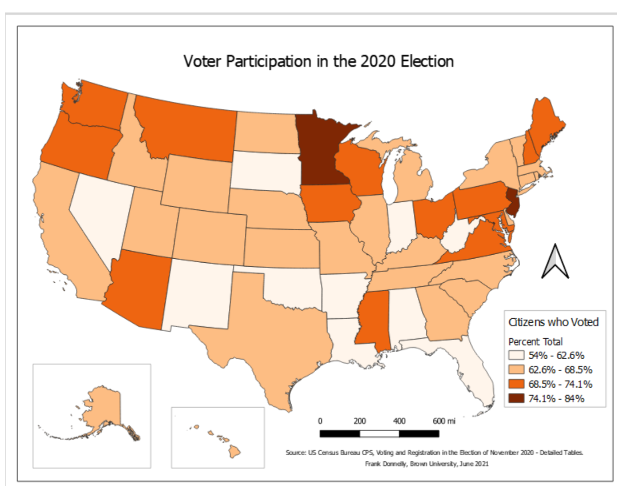


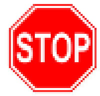
12. *Balance your map elements.* At this point you should have all of your map elements in place. You may need to resize and shift elements around in order for the map to appear balanced. If you want to insure that boxes are lined up properly, you can hit the select move button and click on individual features while holding down the CTRL key to select multiple items. You can use the various align buttons to arrange elements in a certain way, and you can use the group button to bind several features together so you can move them in unison. Save your project.

13. *Print to PDF.* PDFs are good for stand-alone maps. Before you export, make sure you don't have any map elements selected and return to the Composition tab for the map. Hit the export to PDF button and save your map as a PDF file, voters_2020.pdf, in your part 4 data folder. You have the option to save the map in the PDF as a vector; if you leave the box unchecked it will be saved as an image.

14. *Export as PNG.* You can also save your map as an image file like a JPG or PNG, which would allow you to insert it into documents. Normally we would want to design the map to be the size of the desired image, and we'd want to adjust the DPI quality to reduce it's size. Hit the export as image button. Browse to your data folder for part 4 and save the map there as voters_2020.png. Make sure that you specify PNG as the file type.

15. *Take a look at your maps.* Minimize QGIS and use your file browser to go to your part 4 data folder. Double click on the PDF file to open it in your PDF viewing software, and on the PNG file to open it in your default image viewing program (or open it with your web browser). Congratulations on creating a finished map!





4.4.2 Commentary

QGIS Map Composer Details

In some GIS software packages the current view in the map window and the print layout are dynamically linked, and a change in one (such as adjusting the zoom) affects the other. This isn't the case with QGIS; the two are separate. If you do change something in the map view, such as reclassifying the data, you can update the map composer under the item properties tab for the map by hitting the Update Preview button.

The print composer allows you to customize minute details of the canvas, map, and legend. The composer also gives you the ability to draw shapes or add portions of an attribute table directly to a map. You can also store more than one map in a single project. From the map view, you can use the new layout button to create new, individual maps, and the layout manager button to manage your maps and choose a particular one to show or edit.

While units can be converted on the fly for the scale bar (from meters / kilometers to feet / miles in this example), it's still best practice to use a map projection that uses feet or meters and not degrees. Converting degrees to other units of measurement, particularly at this scale, is complicated and should be avoided. Most projected global and continental systems are in meters, while regional and local projections like UTM and State Plane are published both in meters and feet. In our map we created a scale bar that's just for the US; conventional practice would require us to create separate bars for Alaska and Hawaii since they are not at the same scale. On the other hand, scale bars and north arrows are only crucial on reference maps (i.e. street, property, or topographic maps), where the emphasis is on depicting direction or distance; for many thematic maps they are considered optional and we could have omitted them from our map.

In our example, we applied a graphic design trick to rotate Alaska and Hawaii to make them appear "correctly", as they are at the periphery of the North American continent and thus appear tilted in the NA Lambert Conformal Conic projection (which is "correct" within that system). Alternatively, within the Map Layout we could have assigned a different CRS to each map window in its Item properties, one that would be more appropriate for each of those areas. This approach would be best if we wanted or needed to apply scale bars and arrows for these maps.

General Map Design

When creating maps you need to design with the end use, format, and audience in mind. If you're designing a map that you're going to embed as an image in a document or web page, you should change the size of the canvas and design the map to the specifications for the document. Creating a full size 8 1/2 by 11 map and scaling or cropping the final image is a bad idea; you'll introduce distortion into the map and text will become illegible. You also need to think about page orientation; it's appropriate to map the United States using a landscape page layout, but if you were mapping an area that was taller rather than wider (South America) you'd want to flip the page to portrait.

Individual map elements (maps, title, arrow, legend, source text) should be balanced on the page to achieve some harmony; avoid lumping too many elements together or having large areas of white space. The title and legend should concisely and accurately describe what the map is about and what you are mapping. The amount of detail you provide and the terminology you use should vary with your audience. You should always include the source of your data in the map. The fonts, north arrows, and other elements should also be tailored to the map content; a title in calligraphy font and an ornate compass rose may look good if you're recreating a Renaissance-era chart, but it would look rather silly on our US voter participation map.

Maps are a form of communication, designed to send a message. Like a book or article that is poorly written, maps that are poorly designed will fail because they do not effectively communicate their message to their audience. Some reasons why maps can flop:

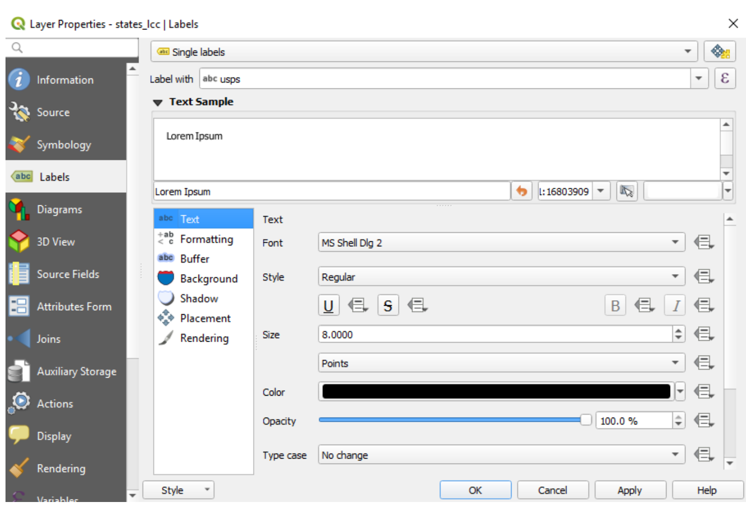
- Poor layout - map elements (map, legend, title, text) arranged in an uneven or sloppy way
- Poor use of symbols - circles too big or small, not enough dots per person, etc
- Improper data classification - too many or few classes that obscure patterns, illogical scheme for dividing data
- Violation of basic cartographic convention - improper conventions for labels and color
- Poor figure-ground relationship - inability to clearly distinguish land from water or foreground from background
- Poor color scheme - random schemes for quantitative data, color that's improper for final format (color print, photocopy, screen projection, etc.)
- Information overload - too much information (several variables or map elements) or noise (unnecessary information)
- "Chartjunk" - concept coined by the graphic designer Edward Tufte, refers to kitschy or gimmicky elements that add nothing to the message of a map or graphic
- Factual errors - mistakes with labels, data, or geography
- Violates expectation of the user - simplification or generalization is too much for the user to accept
- Offends culture of the user - the message or how the message is communicated (text, colors) violates taboos that a user or group cannot accept

4.5 Adding Labels

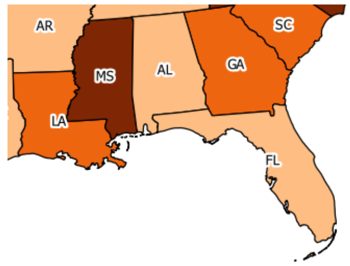
In this section we'll go back and add some labels to our map. The labeling system can be accessed via the Labels tab under the Properties menu for a particular layer.

4.5.1 Steps

1. *Turn labels on.* Close the print composer and go back to your QGIS map view. Select `states_nalcc` in the LP and double click on the layer to go to its properties, then labels tab. Change the drop down box value from No labels to Single Labels. In the Label with drop down, choose STUSPS as the label field (this field has the two-letter postal code for each state). In the Text menu change the size of the text to 8.0. In the Buffer menu check the box to Draw text buffer. Hit OK to apply the label settings.

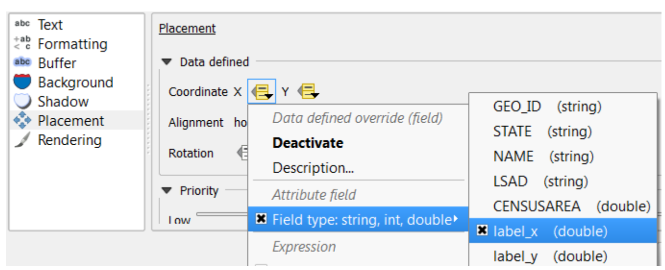


2. *Inspect the labels*. At first glance the label placement looks pretty good. There are a few small issues; the labels for Florida and Louisiana look a little off center. And if you're zoomed out so the lower 48 states fill the screen, the labels for Rhode Island and Washington DC might be omitted, as they overlap with labels of neighboring states. With a little extra work we can fix that.

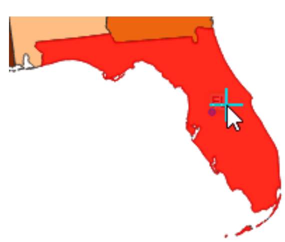


3. *Add new columns to the attribute table*. The labels are automatically placed in the center of the state. In order to define and store a specific position for them, we have to add some new columns to the attribute table. And to do that, we'll have to enter an edit mode so we can actually modify our file. Open the attribute table for states_nalcc. Hit the edit button at the top of the table. Hit the New field button. In the Add field window name the new field label_x. Assign it a Decimal number type. Hit OK, and the new column gets tacked on at the end of the table. The label_x column will hold the X (longitude) coordinates for our label. But we need a second column to hold our Y coordinates (latitude). Repeat the previous step to add a second column called label_y. Once you've added it, hit the edit button to save the changes, and the columns become permanent. Close the attribute table.

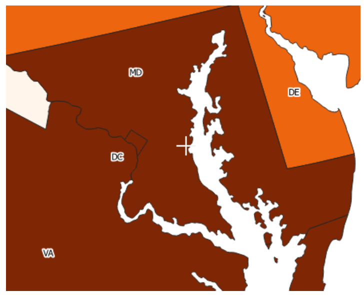
4. *Update label menu settings*. Before we can start moving labels we have to tell QGIS to store their positions these new fields. With the layer selected in the LP go back to the Labels tab in the properties menu, and go to the Placement menu. Scroll down to the bottom of Placement to the Data defined menu. In the dropdown for X Coordinate, select Field type and choose the label_x field. In the dropdown for Y coordinate, select the label_y field. As you make the selections the drop down icons will turn from white to yellow. Hit OK to save the settings.



5. *Move the label for Florida*. With states_lcc selected in the LP, right click on it and hit the toggle edit button to enter an edit mode. You'll see that the move label button on the toolbar is now active. Hit the button, and you'll see a crosshairs as you move across the map. Adjust your map so that Florida (FL) is visible and centered. Click on the Florida label to select it. Then move the crosshairs over to the center of the state, and click again to move the label to that spot.

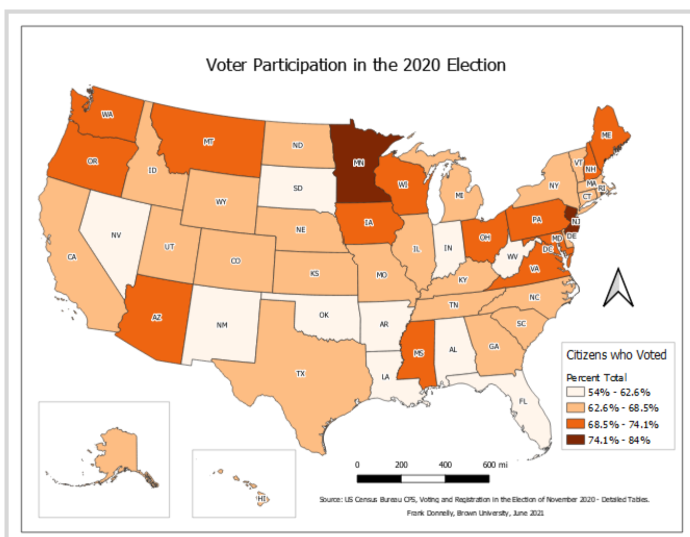


6. *Adjust additional labels.* Do the same to center the labels in Louisiana (LA) and California (CA). If need be (this could vary based on your zoom and may not be necessary) use the pan tool to move to the northeastern US, then reactive the move label button. Move the label for Maryland (MD) to the north and the label for DC to the south so that both will be visible. The labels aren't going to look right at this scale, so zoom out back to the continental US to make sure the labels look OK at that scale. Lastly, if necessary, adjust the label for Massachusetts (MA) but make sure that the label for Rhode Island (RI) still draws. Once you're satisfied, hit the edit button for the layer to stop editing and save your edits. You may have to enter the edit mode, move labels, and exit a few times until you get the labels right (as it may be difficult to see their placement in the edit mode). When you're finished, you can open the attribute table for the layer, scroll to the right, and you'll see that coordinates are stored in the x_label and y_label fields for the labels you moved. Close the table. Save your project.



7. *Update your map layout.* Hit the Layout manager button, select First map and hit Show. You should see all your map labels - don't worry if they appear overlapped; they should turn out fine in the export. If you don't see the labels, hit the refresh button, or select each map in turn and under Item properties hit the Update preview button.

8. *Save and export.* Export your map. Print your map out as a PDF or save it as an image. Save it in your part 4 data folder as voters_2020_labels.pdf (or .png). Minimize QGIS, go to your part 4 data folder, and take a look at your final map.



4.5.2 Commentary

Labeling in QGIS

QGIS has good support for labeling, in terms of automatic placement and the ability to move and customize labels. There are some additional options beyond what we've demonstrated here:

- You can also add columns to your attribute table that allow you to specify label details for each feature such as font type, size, color, placement, and rotation.
- The text annotation tool allows you to add call out boxes directly in the map view. This is practical if you only need to place a few labels.
- You can also use the add label feature within the map composer. This can be a little cumbersome since you cannot copy and paste labels, but must create each one from scratch; ok if you only need to add a few labels.

Generally, features can be displayed and differentiated from each other using text. For example, the standard cartographic convention for labeling bodies of water is to use an italic font and, when possible, a dark blue color. The size of a label indicates the hierarchy of the feature - oceans have larger fonts than seas, which have larger fonts then rivers, larger than streams, etc. Land features are labeled in black, or anything that isn't blue, and are never written in italics. Larger features, land or water, may be written in all capital letters, while smaller features are in lower case.

ATLANTIC OCEAN GULF OF MEXICO Lake Ontario Hudson River

UNITED STATES NEW JERSEY Philadelphia Trenton

Thematic Maps and Symbols

In this tutorial we worked through an example for creating a shaded area or choropleth map. However, there are a number of other techniques that you can use to create a thematic map. New symbolization improvements include heat maps and rule-based placement. QGIS also supports graduated symbols for point and line layers, where the relative size of the symbol (a circle, square, line, or image) represents a value (if you look at the style tab for a point layer, you can change the legend type to graduated symbols). If you have a polygon layer that you'd rather map as graduated circles (instead of shaded areas) you have to convert it to a point layer first (you can do this under Vector > Geometry Tools > Polygon Centroids).

Symbols are used to show qualitative data (name or feature type) or quantitative data (proportions or numbers) and are often divided into four types:

Nominal - qualitative measurements like the name or type of feature, shown using unique symbols.

Ordinal - quantitative measurements with a general order of size, like small, medium, or large, shown using symbols of different sizes or colors.

Interval - quantitative measurements with a specific beginning point and range of specific values (distance, temperature, elevation), shown using a variety of symbols (isolines, shaded areas, graduated symbols).

Ratio - a type of interval measurement that shows the relationship between the area and some phenomena (time to cover a distance, population density).

Symbols are often designed to mimic the features they represent, i.e. airplanes for airports, little buildings with flags to represent schools, etc (these are all examples of nominal symbols). In some cases, features may be represented with geometric shapes (circles, squares, triangles) that can be easily distinguished on small scale maps. Some features may be represented using a standard convention for classifying them, i.e. mining maps may label minerals based on their abbreviation in the periodic table - Sn for tin, Pb for lead, Cu for copper, etc.

A single symbol can be used to identify a feature. Varying the size or color of the symbol can indicate quantity. The width and color of roads on a map is highly standardized to show the type of road and volume - in the US thick blue roads are interstate highways, thick green roads are toll highways, thinner red roads are US highways and thinner black roads are state or local roads (all ordinal symbols).

4.6 Considerations and Next Steps

Now that we have mapped this data - what does it mean? How would you interpret this map? Are there any spatial patterns to the data (clustering) or does it appear more or less random? Maps have the ability to answer questions but also raise new ones. In order to understand what's going on, we have to become familiar with the underlying dataset. What influences voter participation and registration, and how might that explain the distribution across different states?

For more practice, some things to try:

- In addition to shaded areas, we can also create graduated circle maps. Edit the table join to add the citizen population that's registered to vote (pct_reg) attribute to the shapefile. Convert the states_nalcc polygon layer to a point layer (using Vector - Geometry Tools - Centroids) and in the Style tab under Layer properties symbolize the point layer based on pct_reg. Click on each of the circle symbols one by one, and modify their size so that larger values are depicted with a larger circle. Hop into your map layout and create a bi-variate (two variable) map that shows the percentage of the population that's registered (as circles) and percentage of citizens who voted (as shaded areas). You'll have to turn the labels off, as the map will look too busy.
- Since this is sample-based data, most of the values have a margin of error (MOE) associated with it in an adjacent column. For example, in Alabama in 2020 we're 90% confident (the confidence interval for the entire dataset) that 60.5% of US Citizens voted (in the pct_vote column), plus or minus 3.3% (in the moe_voted column). How could you communicate this information about the margin of error in your map?
- In the part 4 folder is a subfolder called extra that has data for countries and shapefiles from Natural Earth, a good source for global boundaries and features. Start a new project, and using what you've learned take the data file un_internet.xlsx stored in the part 4 folder and join it to the countries layer to make a thematic map (hint - use the adm0_a3 column in the shapefile as the unqiue ID). The data represents the percentage of a country's population that uses the Internet, and was download from UN Data at `http://data.un.org/`. Re-project the countries shapefile from WGS84 to a projected coordinate system like Mollweide of Robinson. There's an additional shapefile that will give you a bounding box (ne_50m_wgs84_bounding_box) that you can underlay on your map - make sure you re-project it to match your countries layer. Note that data for some countries is missing.

Chapter 5

Data and Educational Resources

This tutorial has provided you with a basic introduction to GIS concepts and applications using QGIS. This chapter will cover the next steps you can take on your own.

5.1 Finding Data

Throughout this tutorial you've been provided with data that you've used to work through various exercises. Once you're working on your own projects, you'll need to find or create the data you need. There is a lot of free GIS data available on the web, created by various government agencies, academic and non-profit organizations, and private companies. You can try a search engine or look at an academic map / GIS library website for a list of helpful links (a list of suggestions is included in the following section). To be strategic about your search, it helps to understand who creates and provides the data:

- *Global / international:* Look at supra-national agencies, like the United Nations (such as the UN's Environment Programme) or academic / non-profit organizations who have enhanced and updated public domain data such as the Global Administrative Areas (GADM), Natural Earth, and OpenStreetMap projects. If you need satellite imagery the best sites to visit are the USGS and NASA.
- *Country level:* In some cases you'll want to visit a few of the international sites to get basic country-level datasets like state or provincial boundaries. But in many instances you may want to visit a mapping agency website or data repository for the specific country you're interested in; you'll find more country specific layers and they will be processed in a way that is readily compatible for mapping attribute data from that country. Most countries have one or two agencies that will provide the bulk of the country's GIS data - a statistical agency responsible for the census, or a mapping agency responsible for surveying. In the US you could go directly to the US Census Bureau or the USGS to download data, or you could visit the central data.gov repository. In Canada, you could visit Statistics Canada directly or visit the Geogratis repository.
- *State / Provincial:* You may be able to visit a country level source, like the US Census Bureau to get state, county, or postal code boundaries for the entire state, or you can visit a state level agency to get more specialized datasets for that state. Some states will have state government portals where you can access all data for a state, others may cooperate with a college or university located in that state to provide data via the university's portal. In addition to centralized portals, individual departments or agencies may also provide data directly; road and transportation layers may be provided by a state department of transportation or may be provided through the state's central portal. State agencies are also the most likely source for aerial photography.
- *County / City / Local:* Local governments may have portals where they provide administrative boundaries, transportation data, and real estate or tax parcels, and datasets that would be of local interest (such as neighborhood boundaries that may not be formally defined elsewhere). You can also look at the geography one step above (state level) to see if data is available for the local area.
- *Gazetteers and Geocoding:* If you can't find an existing GIS dataset, you can always try to create one from an online gazetteer that provides longitude and latitude coordinates for point-based features; the USGS has a US

level gazetteer, while the NGA has an international one. Geonames is a global, crowdsourced gazetteer that draws from official sources. Do you have a list of addresses but no coordinates? You can use the geocoding tool that's part of the mmqgis plugin, or you can upload them to a free geocoding service at the US Census Bureau's Geocoder which will translate your addresses into coordinates.

- *Universities and non-profits:* These sources often provide data within a specialized area of interest. Some universities provide data for the geographic areas where they reside, and special labs or research groups produce data based on their research interests. There are a number of universities who collaborate to create meta-sites that provide geographic data at a variety of scales across the globe; the Open Geoportal and Geoblacklight communities are two such groups. Non-profit projects like Natural Earth and OpenStreetMap exist to produce free and open data, while other organizations produce data as a byproduct of their larger missions.

Regardless of where you download your data, you'll want to examine the metadata for the layers. Metadata can be formally or informally described on the website where you downloaded your files, in narrative documentation that is included with the files you downloaded, or in special XML files that accompany each of your GIS files. There are a few well-defined standards such as the FGDC and ISO 19139 that data creators use to document data, and include elements that explain who created the data, when it was last updated, what the file contains, what the intended purpose of the file is, if it was created for a specific optimal scale, the coordinate system and map projection it was created in, and copyright and use restrictions. You'll want to check the metadata to verify that the data is going to meet your needs and that you can use it for your intended purpose. For example, you wouldn't want to use a generalized boundary file if you're mapping at a large, local scale, and if you are going to use the data for a commercial purpose you need to verify that that's permitted. In any event, you should cite the source of your data in any maps, tables, or reports you create from it.

If you are looking for a particular GIS file and it's provided by several sources, which source should you use? For example, if we wanted census tracts for a particular city, we could download them from the city's GIS page, from a state-based site, from a college or university repository, or from the Census Bureau itself, via the TIGER page or the generalized boundary page. To answer this question, you'll have to examine the download page, and even download the files to view them and their metadata. Here are some things to consider:

- How are the files packaged for download? Do I have to download them one place at a time, or could I get the entire area in one download?
- Who created the files originally? Is it better to go with the original source? Or has a secondary source added some value that makes their files more desirable?
- Can I trust the source? Is there metadata? How did they create the data?
- How old are the files? Are they timely enough to suit my needs?
- For vector files, are the layers generalized or not? What scale are they appropriate for?
- For vector files, are the polygons saved as single or multipart layers?
- For vector files, what attributes are available in the attribute table? Are there ID codes that I can readily use to join data? Are there place names that I can readily use as labels?
- For raster files, what is the resolution of the data? Is it appropriate for my intended use?

- What format is the file in? Is it a format I can use, or at least one that I can easily convert?
- Are there any copyright or use restrictions with the data?

Finally, remember that GIS data is often just one piece of the puzzle. It represents the geographic features, but if you need attributes to go with these features (demographic data, weather data, sales data, etc) you'll have to download this data from someplace else (or create it yourself) and process it to make it usable with your GIS data.

5.2 Data Sources

Meta

- *Geolode* `http://geolode.org/`: This resource contains search-able records with links to hundreds of websites from around the world that provide freely available GIS data. It was created by the Geospatial Librarian at Cornell University and is maintained by volunteers who work in GIS positions in universities.
- *GeoblackLight Repositories* `https://geoblacklight.org/`: This is collaborative community of geospatial professionals (primarily at universities) who share data and metadata from their individual institutions through an open suite of software that the group develops. This link brings you to a list of university repositories; some data is restricted, but there is much that is public (scanned, georeferenced historical maps in particular).

Global

- *Natural Earth* `https://www.naturalearthdata.com`: Generalized raster and vector data for countries, available at three different scales.
- *GADM* `https://gadm.org/data.html`: Global administrative boundaries for every country, for national and sub-national units several levels deep (states, provinces, counties, districts, sub-districts).
- *United Nations environment programme* `https://unepgrid.ch/en/platforms`: Several platforms for accessing global and continetal GIS layers. The GRID core datasets platform is an accessible, searchable database.
- *OpenStreetMap*: Features on the OSM can be extracted as vector layers in QGIS by enabling the third-party QuickOSM plugin. The plugin allows you to create extracts of specific features within a given area. To grab data in bulk (by country or region) there are several sites with downloadable extracts such as Geofabrik `https://download.geofabrik.de/` and BBBike.org `https://download.bbbike.org/osm/`.

Canada

- *GeoGratis* `https://geogratis.gc.ca/`: Canadian government GIS repository provided by the Earth Sciences Sector of Natural Resources Canada.
- *Statistics Canada, Maps and Geography* `https://www.statcan.gc.ca/`: Select the Geography tab and search for boundaries, road networks, and place name files from Canada's statistical agency.

United States

- *TIGER Line Shapefiles, U.S. Census Bureau* `https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html`: Extracts of the bureau's TIGER Line files for several legal, administrative, and statistical areas in the US, updated annually.
- *Cartographic Boundary Files, U.S. Census Bureau* `https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html`: Generalized extracts of the bureau's TIGER Line files for several administrative areas (i.e. states, counties, zip codes) and census (i.e. tracts, block groups, metros) areas in the US.
- *National Historical Geographic Information System* `https://www.nhgis.org/`: The NHGIS is a project at the University of Minnesota that compiles and provides historical census boundaries and data for the United States from 1790 to present. New users must register, but there is no cost and downloads are free.

- *USGS National Map* `https://apps.nationalmap.gov/downloader/`: This federal agency provides imagery, digital topographic maps (DRGs), elevation data, and some boundary files.
- *HIFLD Open Data* `https://hifld-geoplatform.hub.arcgis.com/pages/hifld-open`: National infrastructure data from the Department of Homeland Security, includes hospitals, schools, transportation, utilities, law enforcement, and more.

State of Rhode Island

- *RIGIS* `https://www.rigis.org/`: The Rhode Island Geographic Information System is maintained by government agencies throughout the state and is hosted by the University of Rhode Island. It contains a variety of vector and raster layers of local interest. Most states and many universities host similar repositories.
- *Ocean State Spatial Database* `https://github.com/Brown-University-Library/geodata_ossdb`: A geodatabase for Rhode Island that includes water bodies, legal, administrative, and statistical areas, census data, and basic point features like schools, hospitals, and libraries. Produced by the Brown University Library's GeoData@SciLi team.

5.3 Other Tutorials and Tools

In this tutorial you've learned what GIS is, what it looks like, and generally how it works. You've learned how to work with vector-based GIS data to do some basic geoprocessing and analysis, and you've learned the basics of thematic mapping and map design. There is still a lot more to explore. As you begin to use GIS more frequently, you will likely begin using more specialized tools that are relevant within your specific field. The processing toolbox contains many such tools, and QGIS interfaces with some more specialized packages like GRASS and SAGA (which are bundled with the QGIS download). Spatial databases like Spatialite and PostGIS have ready interfaces with QGIS through the Database Manager, and will give you additional power for querying and organizing data.

The QGIS website `https://qgis.org/en/docs/` and the OSGeo foundation `https://www.osgeo.org/resources/` have links to additional manuals and tutorials for learning QGIS. The QGIS Tutorials blog (not affiliated with the QGIS project) has detailed tutorials for individual tasks at `http://www.qgistutorials.com/en/`. There are lots of QGIS tutorials on YouTube. In print, *Learning QGIS* by Anita Graser provides a good introduction with clear steps and plenty of screenshots, while *Mastering QGIS* by Menke et. al. provides a fuller treatment. Both are published by PACKT, which specializes in books on open source software and programming languages.

Stuck and need help? Take a look at Geographic Information Systems StackExchange, an extensive question and answer forum at `https://gis.stackexchange.com/`.

If you think you're going to become deeply involved in GIS, you may want to consider experimenting with the major proprietary packages such as ESRI's ArcGIS Pro, which includes a deeper catalog of specialized tools. Check with your geography / earth sciences department, campus library, or GIS lab to to see if your campus has a site license (Brown University has a site license that's available for members of our campus). Once you're familiar with QGIS, the leap to a package like ArcGIS isn't too great because they use a similar interface and operate under the same basic principles. ArcGIS Pro is well documented; there are many books and online tutorials. On the flip side, the software is more resource intensive, is only available for the Windows operating system, and is expensive enough that it's not a viable option for an individual user (although basic low-cost options are available for students and non-profits).

How about web mapping? ArcGIS Online `https://www.arcgis.com/index.html` is a basic web-mapping platform, and free public accounts are available. You can use it to create interactive web-maps. You can use QGIS to prep your data and upload your vector files and data to ArcGIS Online, where you can create the interactivity and have a place to host your map.

Appendices

Appendix A

Web Mapping Services

To add web base maps to QGIS: use the Browser Panel, select the type of service (WMS / WMTS, or XYZ), right click, and add a new service. Provide a name and the URL. If you use these base maps in your maps, cite them as a source.

A.1 OpenStreetMap

These are global XYZ Tile services. For basic reference showing terrain and streets; the OpenStreetMap (left image below) is already included in QGIS.

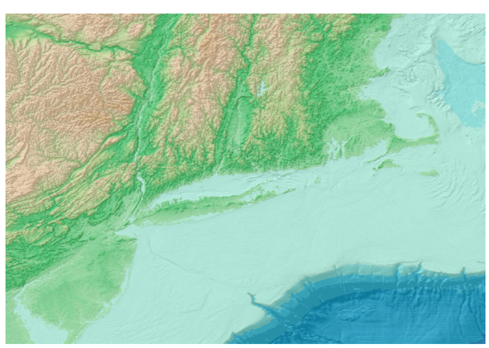
OpenStreetMap : http://tile.openstreetmap.org/z/x/y.png

OpenTopoMap : https://tile.opentopomap.org/z/x/y.png

A.2 GEBCO

These are global WMS services. General Bathymetric Chart of the Oceans, also includes shaded elevation and relief.

GEBCO : https://wms.gebco.net/mapserv?



A.3 USGS National Map

These are global WMTS services, but fine detail at large scales is only available within the US. A mix of satellite imagery and topography.

Imagery :
https://basemap.nationalmap.gov/arcgis/rest/services/USGSImageryOnly/MapServer/WMTS/1.0.0/WMTSCapabilities.xml

Imagery & Topo :
https://basemap.nationalmap.gov/arcgis/rest/services/USGSImageryTopo/MapServer/WMTS/1.0.0/WMTSCapabilities.xml

Shaded Relief :
https://basemap.nationalmap.gov/arcgis/rest/services/USGSShadedReliefOnly/MapServer/WMTS/1.0.0/WMTSCapabilities.xml

Topographic :
https://basemap.nationalmap.gov/arcgis/rest/services/USGSTopo/MapServer/WMTS/1.0.0/WMTSCapabilities.xml

A.4 US Census Bureau

These are US WMS services for current census geography; see their website for links to older vintages.

Current TIGER :
https://tigerweb.geo.census.gov/arcgis/services/TIGERweb/tigerWMS_Current/MapServer/WMSServer

Current physical features :
https://tigerweb.geo.census.gov/arcgis/services/TIGERweb/tigerWMS_PhysicalFeatures/MapServer/WMSServer

Appendix B

Common CRS Definitions

Visit `http://www.radicalcartography.net/?projectionref` for visualizations and descriptions of map projections and `https://epsg.io/` for a catalog of CRS definitions.

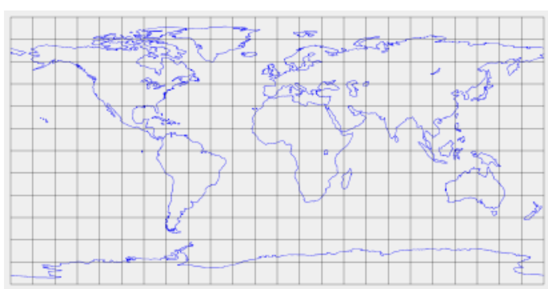
B.1 Geographic Coordinate Systems

WGS 84 (EPSG 4326): World Geodetic System of 1984, commonly used by organizations that provide GIS data for the entire globe or many countries and used by many web-based mapping engines.

NAD83 (EPSG 4269): North American Datum of 1983, commonly used by US and Canadian federal government agencies (the US Census Bureau in particular) that provide GIS data.

NAD27 (EPSG 4267): North American Datum of 1927, used prior to the adoption of NAD83. The earliest GIS files (up until the mid 1990s or so) employed NAD27, and coordinate data for North America was recorded in this format for much of the 20th century.

Since WGS 84, NAD83, and all geographic coordinate systems are unprojected, they all look like the Equirectangular or "Plate Caree" projection.



B.2 Projected Coordinate Systems for Local Areas

NAD83 / Rhode Island (ft US) (EPSG 3438): The State Plane zone that covers Rhode Island. An alternate projection, EPSG 32130, uses meters instead of feet. Many city, county, and state agencies in the US produce data in their specific state plane zone. Larger states have multiple zones.

NAD 83 / UTM Zone 19N (EPSG 26919): An alternative to State Plane for larger regions that is applicable around the world; satellite or ortho imagery is often provided based on the UTM zone where the tile is located. UTM Zone 19N covers eastern New England and Atlantic Canada. Alternatively, EPSG 32619 uses WGS 84 as a datum instead of NAD 83.

Visually the difference between State Plane (middle) and UTM 19 North (right) is relatively modest when focused on RI, but both are clearly distinct from the unprojected NAD83 GCS (left):

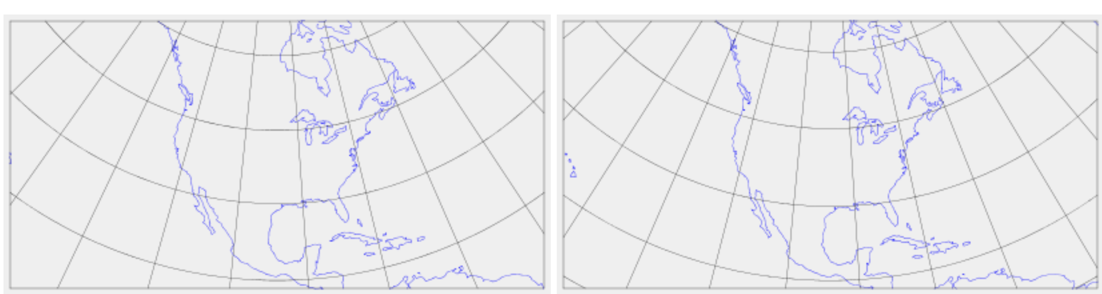
B.3 Continental Projected Systems

US National Atlas Equal Area (EPSG 2163): Commonly known as the Lambert Azimuthal Equal-Area projection, this CRS preserves equal areas and true direction from the center point of the map. It was the best CRS in the original EPSG library for thematic mapping of the entire US, but since QGIS has expanded the CRS library there are better options.

North America Lambert Conformal Conic (EPSG 102009): Perhaps the most common map projection for North America, a conformal map preserves angles. LCC can be modified for optimally displaying specific countries (i.e. USA and Canada), other continents (i.e. South America, Asia, etc.), or other ellipsoids and datums (WGS 84).

North America Albers Equal Area Conic (EPSG 102008): An alternative to LCC, all areas in an AEAC map are proportional to the same areas on the Earth. Can also be modified for specific countries and continents with alternate datums.

Although difficult to see at this scale, visually Albers Equal Area Conic (on the right) looks more compact east to west versus Lambert Conformal Conic (on the left):

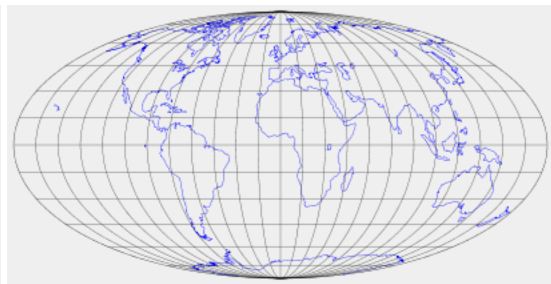


B.4 Global Projected Systems

Robinson (EPSG 54030): A global map projection used by National Geographic for many decades. The Robinson map is a compromise projection; it doesn't preserve any aspect of the earth precisely but makes the earth "look right" visually based on our common perceptions.

Mollweide (EPSG 54009): A global map projection that preserves areas, often used in the sciences for depicting global distributions on small maps.

Visually the difference between Robinson (on the left) and Mollweide (on the right) is apparent:

Appendix C

ID Codes

- *ISO Country Codes:* `https://www.iso.org/iso-3166-country-codes.html`
- *US ANSI / FIPS Codes:* `https://www.census.gov/library/reference/code-lists/ansi.html`

ANSI FIPS Codes for US States

Name	ANSI/FIPS	USPS Code	Name	ANSI/FIPS	USPS Code
Alabama	01	AL	Montana	30	MT
Alaska	02	AK	Nebraska	31	NE
Arizona	04	AZ	Nevada	32	NV
Arkansas	05	AR	New Hampshire	33	NH
California	06	CA	New Jersey	34	NJ
Colorado	08	CO	New Mexico	35	NM
Connecticut	09	CT	New York	36	NY
Delaware	10	DE	North Carolina	37	NC
District of Columbia	11	DC	North Dakota	38	ND
Florida	12	FL	Ohio	39	OH
Georgia	13	GA	Oklahoma	40	OK
Hawaii	15	HI	Oregon	41	OR
Idaho	16	ID	Pennsylvania	42	PA
Illinois	17	IL	Rhode Island	44	RI
Indiana	18	IN	South Carolina	45	SC
Iowa	19	IA	South Dakota	46	SD
Kansas	20	KS	Tennessee	47	TN
Kentucky	21	KY	Texas	48	TX
Louisiana	22	LA	Utah	49	UT
Maine	23	ME	Vermont	50	VT
Maryland	24	MD	Virginia	51	VA
Massachusetts	25	MA	Washington	53	WA
Michigan	26	MI	West Virginia	54	WV
Minnesota	27	MN	Wisconsin	55	WI
Mississippi	28	MS	Wyoming	56	WY
Missouri	29	MO			

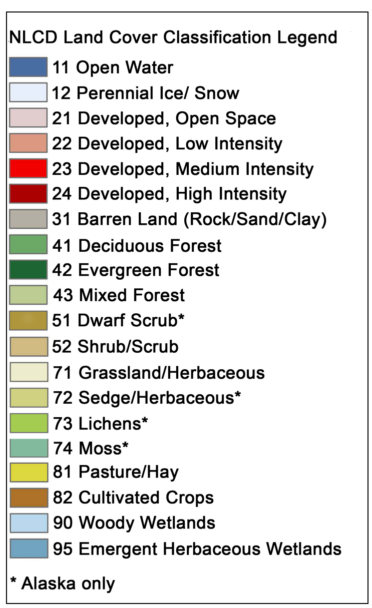
ANSI FIPS Codes for US Territories

Name	ANSI State Numeric Code	USPS Code
American Samoa	60	AS
Guam	66	GU
Northern Mariana Islands	69	MP
Puerto Rico	72	PR
U.S. Minor Outlying Islands	74	UM
U.S. Virgin Islands	78	VI

SGC Codes for Canadian Provinces and Territories

Name	SGC Code	Alpha Code
Alberta	48	AB
British Columbia	59	BC
Manitoba	46	MB
New Brunswick	13	NB
Newfoundland and Labrador	10	NL
Northwest Territories	61	NT
Nova Scotia	12	NS
Nunavut	62	NU
Ontario	35	ON
Prince Edward Island	11	PE
Quebec	24	QC
Saskatchewan	47	SK
Yukon	60	YT

National Land Cover Database Classification



Source: `https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description`

