LAB 1 – DATA HYGIENE AND MANAGEMENT

What you'll learn: Understand common spatial data formats, how they're offered as compressed (aka zip and archive) files online, and how to use Windows software to manage these files. You'll also learn how to set up QGIS projects with good data management practices.

Data: link provided in Step 2.

What you'll submit: Via Canvas, a word document in .doc or .pdf format that answers the questions in Step 6. It should be titled "[YourLastName]_Lab1", e.g., "Spangler_Lab1".

Background: New GIS users are often confronted with a steep learning curve when they begin learning QGIS and similar software, like ArcMap. The most heard exclamation in the lab when trying to first map new data is, "What is it and where did it go??" Good data hygiene is the practice of keeping your data orderly, in a known location, and unpolluted with irrelevant data. The goal of this lab is to show you how to own your data in all steps of map making.

Lab naming conventions: Tools that you click will be bolded, e.g., **QGIS Menu > File > New** to create a new QGIS project file. Text that you'll type will have quotes around it, such as "MyNewProject.qgs" and names of existing datasets and directories will be italicized, e.g., *DataToUse.zip*. Key terms will be underlined. **Questions that you must answer for the lab or instructions for screenshots you must capture will be in bold green font. Important tips and key instructions will be in bold red font.**

Note: As QGIS runs on multiple OS's (operating systems), be aware that my screenshots may not always look identical to what you see on your screen.

STEP 1: OWN YOUR DATA

Let's begin by defining where you will work on your GIS projects. Whether you're working on your own machine or a computer in the lab, you must always begin a GIS project with data discipline.

The folder where you will work on your GIS projects is called your GIS Root Folder. To define it, first open your File Finder app (e.g., Windows Explorer, Finder) and navigate to the *T:* drive. Find the *users* directory. Open that directory and **Right+Click** the white space to bring a context-specific menu (see Fig. 1).

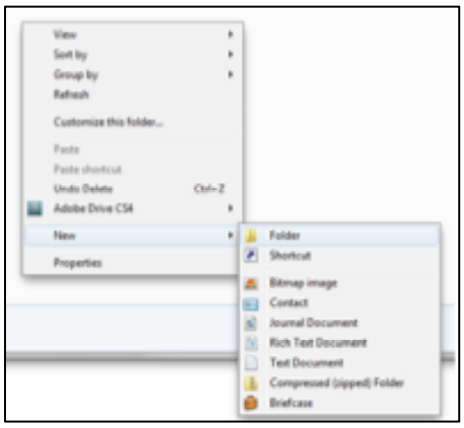


Figure 1

Create a new folder called “[YourName]GIS” (e.g., “SpanglerGIS”). This is your GIS Root Folder because you will store all of your GIS files and data here.

Note that this assumes you are working on a lab computer, in a Windows OS. You may complete this lab on your personal machine, but instead of creating the “[YourName]GIS” folder in the *T:* drive, you will save it to another folder (e.g., a personal folder for “Intro to GIS”).

Inside that folder, **Right+Click** again to create a new folder called “Lab1.” This is called a GIS Project Folder.

Next, inside your *Lab1* folder, save a word document (please use .doc format, no .pages) titled “[YourLastName]_lab1” – for example, “spangler_lab1”. Include your name and the title of the class in the header. This will be the document in which you answer questions, paste screenshots, and submit your Lab 1 assignment.

Finally, inside your *Lab1* folder, **Right+Click** again to create a new folder called “downloaded-data”. This is called your Data Folder, and will be where you move, store, and save all datasets related to the current GIS project. You should now have a file path that resembles the following:

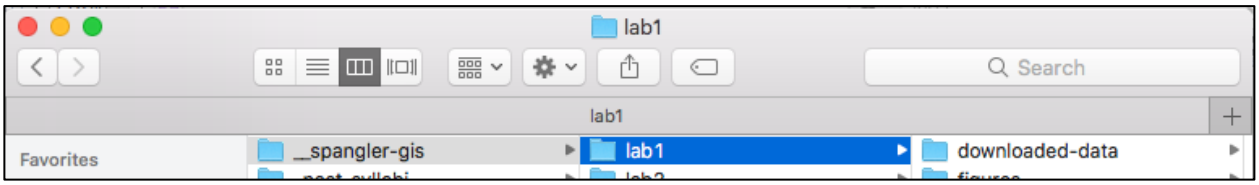


Figure 2

Each week, you will create a new GIS project folder (e.g., *lab2, lab3, and so on)*. It will be where you store data and files relating to that week’s lecture and exercises, including the word document you will submit as your lab assignment.

STEP 2: DOWNLOAD AND EXTRACT SPATIAL DATA

Now that we’ve established a Root Folder, Project Folder, and Data Folder, we’re going to download some data and make sure it gets into the right place.

Most spatial data offered online is compressed into a single file, an archive often called a zip file. Because spatial data is typically large and consists of multiple files, it is often compressed into a single file for easy online service and file sharing.

Open a web browser and navigate to https://data.lexingtonky.gov/dataset/zip_code. You should see a screen that resembles Figure 3:

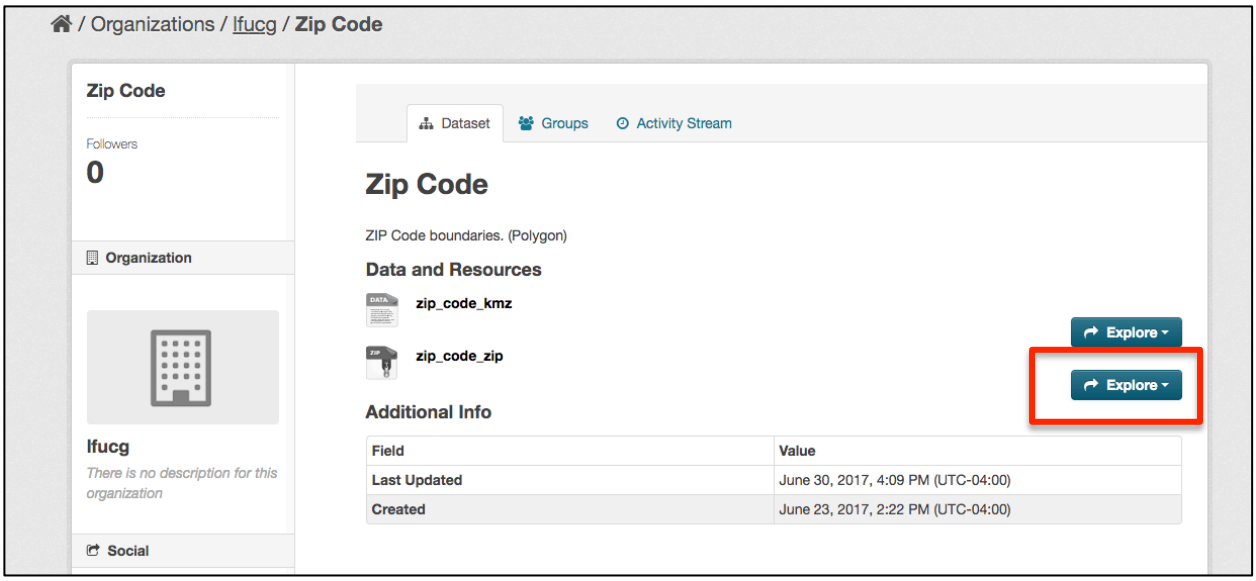


Figure 3

This is the open data portal for the Lexington-Fayette Urban County Government's geospatial datasets. We want to download some zip code data, and we have two options: "zip_code_kmz" and "zip_code_zip." They both refer to the same spatial data, but have different file extensions (.kml and .zip). We want the .zip extension.

Remember: a .zip extension is simply a collection of one or more files that has been compressed into a single file for data storage/transmission. Think of it like a suitcase that has been packed with a number of distinct but related items (e.g., travel items like a toothbrush, shampoo, change of clothes).

As indicated in Figure 3, click **Explore.** You should see two options: "More Information" and "Go to resource." **Right+Click** "Go to resource" and select **Save Link As…** (it might also show as **Save Target As…**). A new window should open, in which you'll be prompted to save the file in a particular location. Navigate to your Lab 1 data folder (e.g., **Root Folder > Project Folder > Data Folder**) and save the zip code data in your data folder. It should be titled "ZipCode.zip".

NOTE: When you work on larger projects, it is best practice to separate raw data from data you create or alter. In those cases, you may find it useful to create multiple Data Folders.

The dataset should now be saved in your Data Folder with a .zip extension. When the file is zipped, we cannot easily discern what it contains – so let's unzip it, or unpack the suitcase, so to speak.

To extract the file on Windows, **Right+Click** on the file and select **Extract All** or **7-Zip > Extract Here**, depending on your version of Windows. To extract the file on a Mac, just **Double-click**.

After extraction, a new folder will be created. Inside it, you should see a series of files with different extensions (Figure 4).

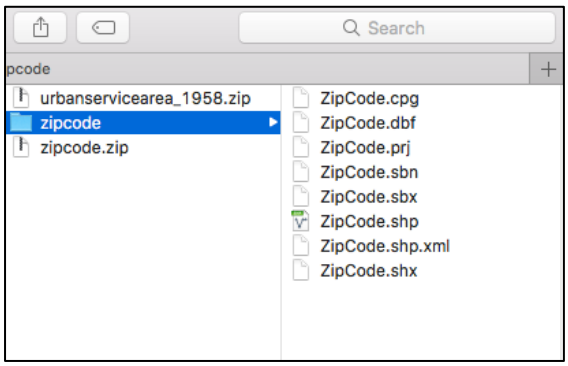


Figure 4

Altogether, these files define a Shapefile format. **Do NOT delete, move, or edit these files outside of a GIS environment.** Each one is essential to the Shapefile's functioning. Think of this assortment of files as team members on a basketball team. Rules dictate that you need five players on the court: no more, and no less. Although the shapefile obviously contains more than 5 files, the same principle applies. When you view this dataset in a GIS, you'll only see *ZipCode.shp*. Later in the course, we'll explore what these various files contain.

You can delete the *ZipCode.zip* file, if you wish, since we have transferred the file and extracted its contents.

Repeat this Step 2 process with one additional file from http://data.lexingtonky.gov/dataset.

STEP 3: CREATING A MAP DOCUMENT IN QGIS

Now that we have acquired two spatial datasets, let's set up a map document in QGIS where we will eventually visualize them. If you're on a lab computer, search for QGIS in the Windows Task Bar and open the program. If you are on your own machine and need to install the software, refer to pp. 5-24 in your *Learning QGIS* textbook. QGIS can be installed at the following link: https://qgis.org/en/site/forusers/download.html

When selecting a version of QGIS to install, choose QGIS 3.4, the stable long-term release.

Once the program has opened, you should see something that resembles Figure 5 at the top of the screen:

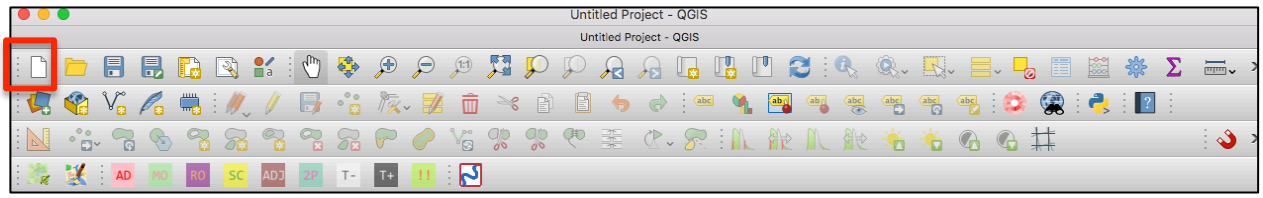


Figure 5. "New Project"

As indicated in Figure 5, **Click** the “New Project” button. You should now see this screen:

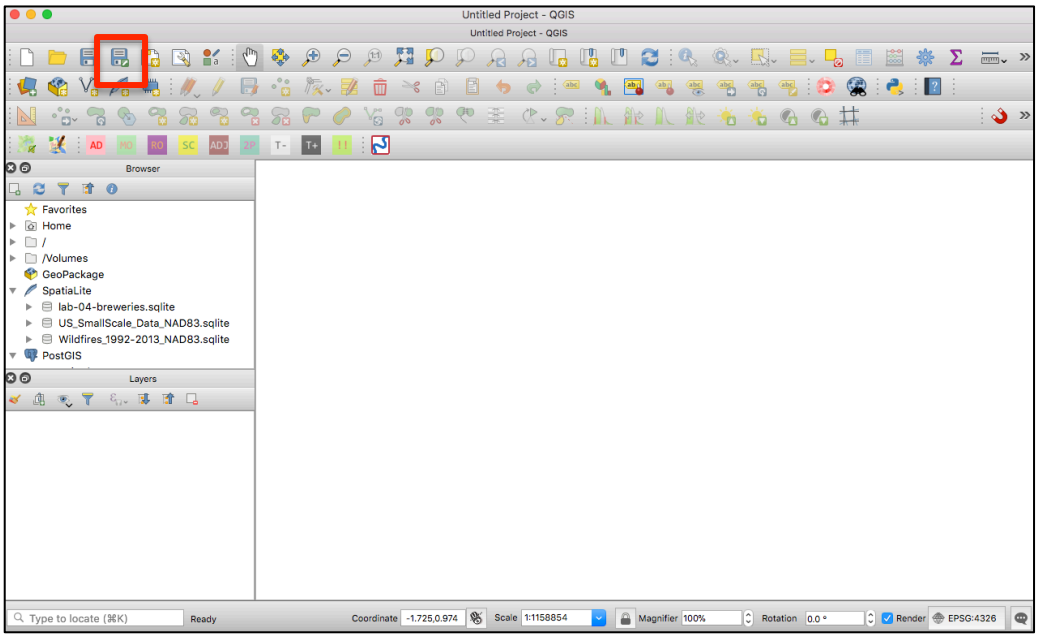


Figure 6: “Save Project As…”

Click “Save Project As…” and navigate to your *Lab1* folder. Save the map document in your *Lab1* folder with the .qgs file format and the proper title (e.g., “Lab1-zipcode-map”).

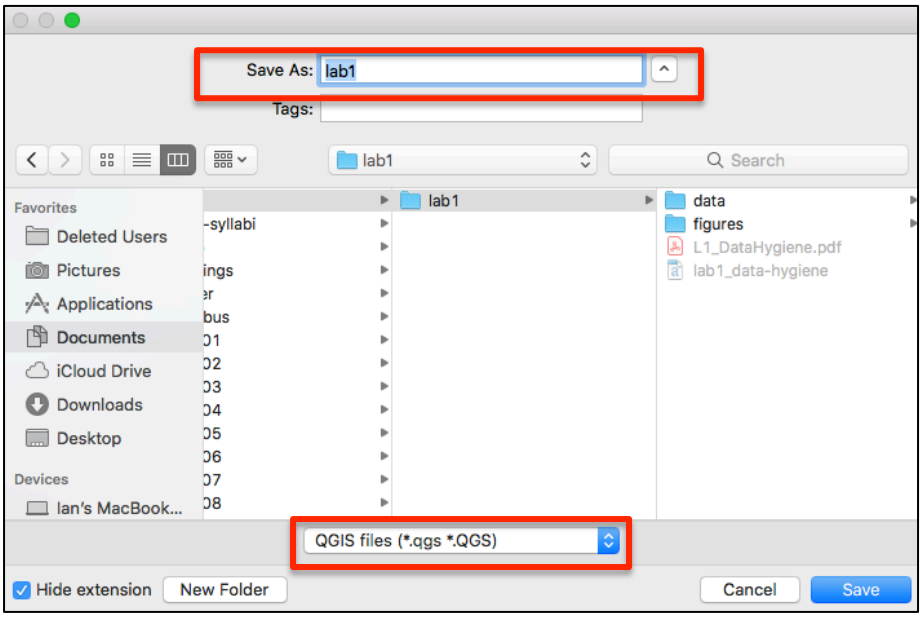


Figure7: properly titled and saved with appropriate file format

Never include spaces in file names, folder names, and downloaded data. Note the naming convention best practices at the end of this document. We will go into greater detail on this in lecture.

Let's briefly take stock of what we've done so far, and review some of the conceptual reasons for these decisions. We have:

- Established a series of folders, including a Root Folder, Lab Folder, and Data Folder (these folders are nested in one another)
- Downloaded 2 spatial datasets and saved them to our *downloaded-data* Folder
- Opened a new map document in QGIS and saved it in our *Lab1* folder

This has all been in the interest of creating a logical workspace in which our data is sensibly stored and our tools for mapping are easy to find. Imagine the alternative: saving data in random folders, titling our workspace with forgettable or unclear names… I shiver just to think of it.

In the next step, we'll finally add our data and get briefly familiarized with the QGIS Graphical User Interface (GUI).

STEP 4: ADD SPATIAL DATA TO THE DATA FRAME

When we look at QGIS, we see a GUI: a way of visually interacting with a computer program through intuitive windows, graphics, and buttons, as opposed to issuing command line prompts (e.g., https://en.wikipedia.org/wiki/Command-line_interface). The QGIS GUI includes a few different components:

- Project name
- Menu
- Toolbars
- Browser / Data
- Layers / Table of contents
- Processing toolbox
- The Map / Data frame

Note where each of these is located in Figure 8, and read through pp. 15-20 in *Learning QGIS* to familiarize yourself more deeply with this interface.

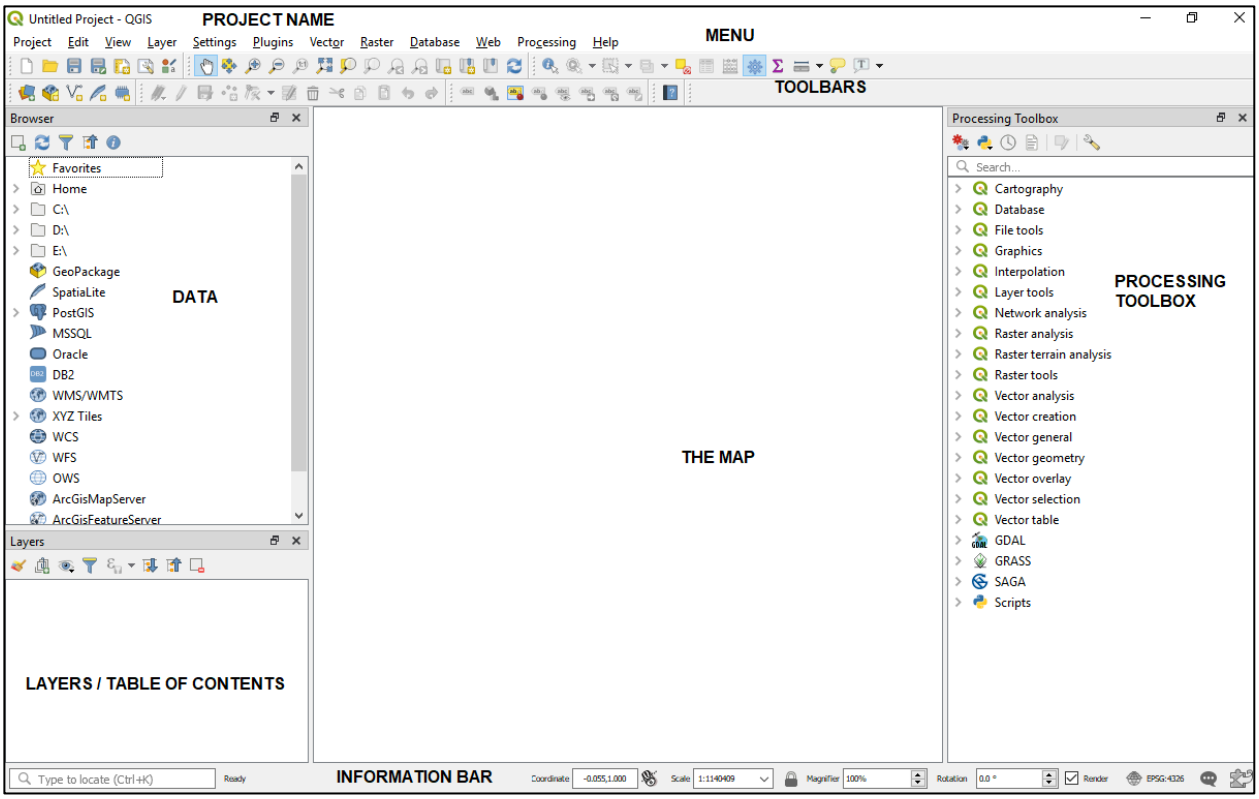


Figure 8: the labeled QGIS GUI, from Cutts and Graser 2018, p. 16

We can add our downloaded data to the map frame in a number of ways:

- Navigating to our data in the *Browser* and **Double-clicking** when we find it
- Dragging and dropping it from our *data-downloaded* folder into the *Data Frame*
- Using the GUI to **Add Vector Data**

Let's start by using the GUI. In the **QGIS Menu**, select **Layer > Add Layer > Add Vector Layer…**

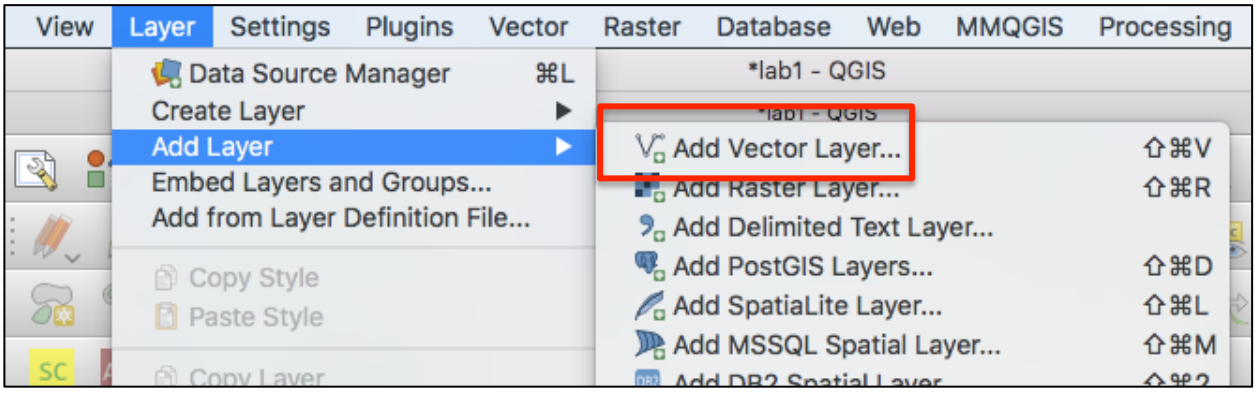


Figure 9

You should then see a screen that looks like figure 10. **Click** the ellipsis, and navigate to your **Root Folder > Project Folder > Data Folder**, selecting the *ZipCode.shp* file.

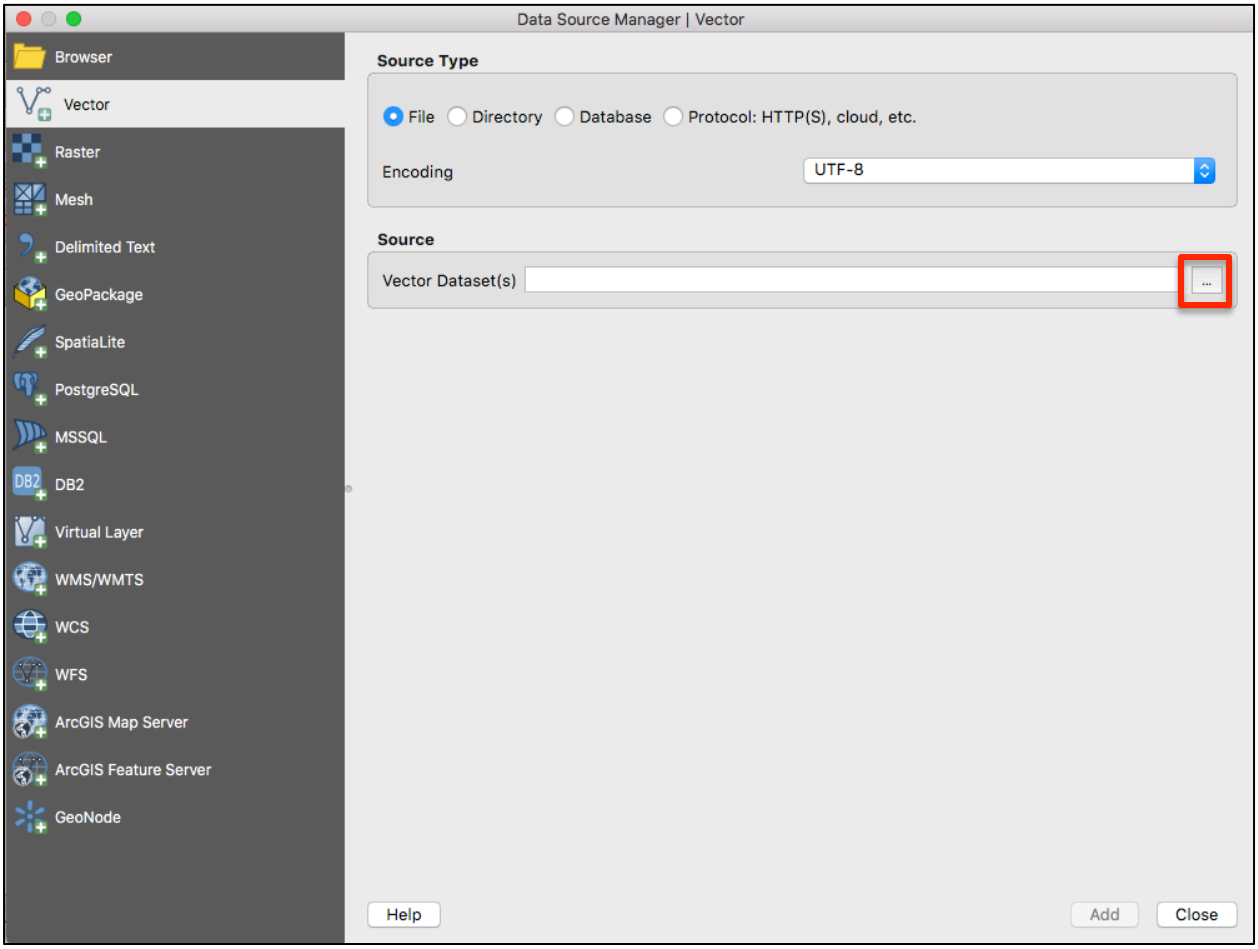


Figure 10. Ellipsis button highlighted

This will add the downloaded data to your Data Frame, resulting in a visualization of Lexington's zip codes. Nice! Get a feel for the GUI by clicking around the data frame a bit. Note that you can drag the data, zoom in and out, and **Right+Click** on *zipcode* in the Layers panel to see what kinds of options are available to you. If you ever get "lost" in the data frame (e.g., you accidentally scroll away from your data and can't find it again), **Right+Click** on the layer and select **Zoom to Layer**.

Note that as an alternative to adding the file with a .shp extension, QGIS can handle loading .zip files. This is not possible in ArcMap.

This is your first map, but it has some problems. First, it is in the wrong projection (we'll explore this thoroughly in the next lab) and the data frame is not appropriately named. Let's fix it.

STEP 5: PROJECTING YOUR DATA IN QGIS

The process of Map Projection, or "transforming the curved, three-dimensional surface

of the planet into a flat, two-dimensional plane" (Monmonier 2018), inherently distorts the earth's surface. And yet, projection is essential to mapmaking. As cartographers, you're not only encouraged to be liars – you *must* be! As the Floating Sheep collective writes: "All maps lie, but lies are the source of maps' power."

In ArcGIS, map projection is a somewhat complicated technical process (and we will discuss that at greater length in Lab 2). In QGIS, however, we can do map projections quite easily "on the fly."

Every spatial layer in a GSI has an inherent coordinate system (you can read about coordinate systems at this link – https://pro.arcgis.com/en/pro-app/help/mapping/properties/specify-a-coordinate-system.htm – but we will discuss them in greater detail next week). The layer's coordinate system is used by a GIS or mapping application to determine how to display the spatial data. Within QGIS, both individual layers and the project as a whole will have a spatial reference setting.

Click on the **Project menu > Project Properties**. This is where the CRS for the entire project is set. Choose the **CRS** tab (left-hand side). It should currently be set to "EPSG: 4326," and it should look like this:

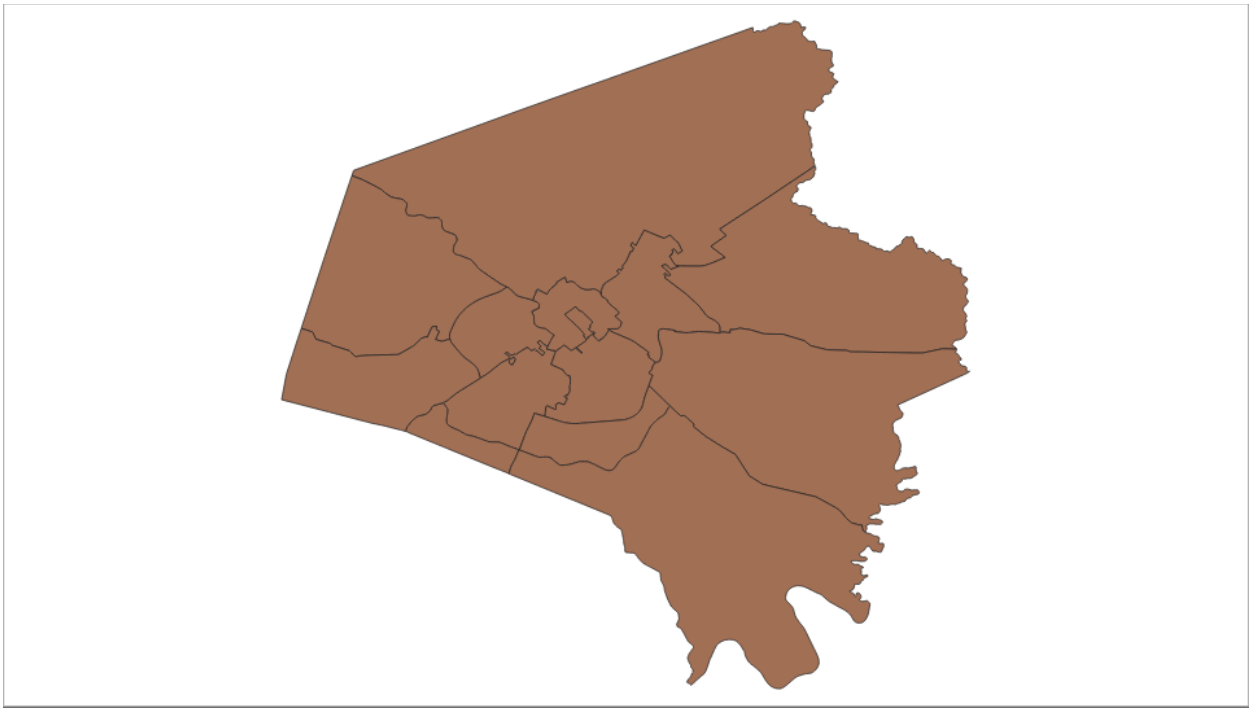


Figure 11: zip code data projected in EPSG:4326, or WGS 84

Find "EPSG: 3089" CRS, which is the official CRS for Kentucky. According to QGIS 3.4 documentation: "QGIS supports "on the fly" CRS transformation for both raster and vector data. This means that regardless of the underlying CRS of particular map layers in your project, they will always be automatically transformed into the common CRS

defined for your project." That's right: you don't necessarily need all your layers and project's spatial reference systems to match!

In the **Project Properties > CRS** window, type "3089" into the filter, and select "NAD83 / Kentucky Single Zone (ftUS)" from the list. Then **Click** "OK."

Ahh – much better! This process should have re-projected your data to look like this:

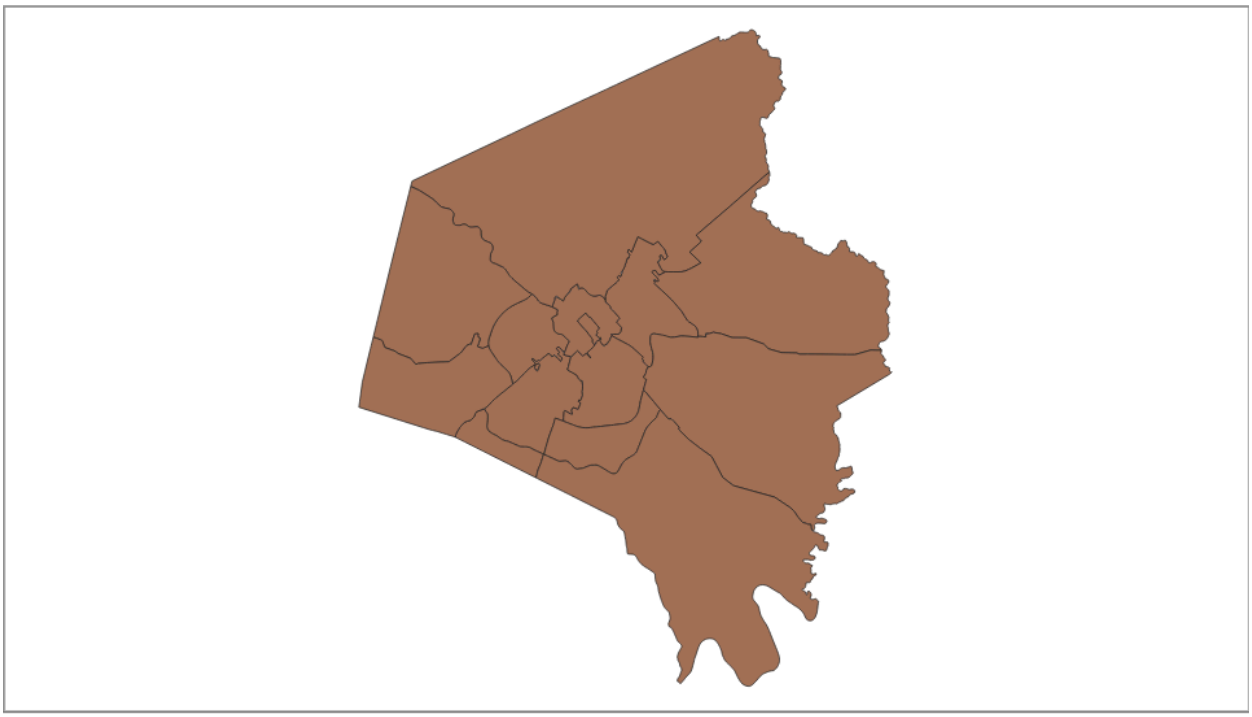
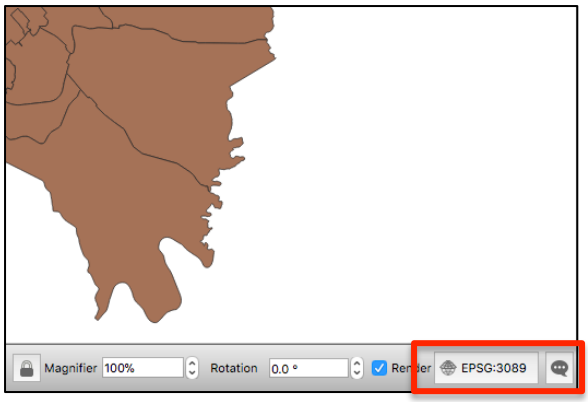


Figure 11: zip code data projected in EPSG:3089

Feel free to test out other projections and see what they look like. Again, we'll work with projections more directly in Lab 2, so don't worry too much about understanding the definitions and processes going on here. Note that on the bottom right hand of the screen, you can directly access **Project Properties > CRS** by clicking this button:



Complete Step 5 by adding the second spatial dataset that you downloaded from Lex's open data portal. Save your work.

STEP 6: QUESTIONS AND SUBMISSION

A few important rules should be observed when naming files, data, and descriptions in GIS environments and when sharing files over the Internet:

1. Never start with a number.
2. Never use spaces. Use the underscore "_" or dash "-"to indicate space.
3. Avoid special characters like "!@,.#$" which might have reserved meaning
4. =Stick to letters, numbers, dashes, and underscores.

Your files should always follow these conventions – not doing so will create more work for you in the long run.

In your *Lab1* word document, answer the following questions. All answers should be written in complete sentences and formatted in a readable manner.

QUESTION 1: Using figure 2 as an example, capture a screenshot of your file path and paste it in your *Lab1* word document. In 2-4 sentences, discuss why it is important to maintain consistent data hygiene practices and follow appropriate naming conventions.

QUESTION 2: What is a GUI? In the QGIS GUI, what is the difference between the browser, the layer panel/table of contents, and the map/data frame? (Feel free to reference your textbook to answer these, and if you do, please indicate as such by citing.)

QUESTION 3: Capture a screenshot of your QGIS GUI, with both datasets loaded and properly projected, and paste it into your *Lab1* word document. **You must include the whole screen!** Please indicate what secondary dataset you downloaded, and in 1-2 sentences, discuss something that you notice about it.

QUESTION 4: Notice that you can click the layers in the **Layers panel** and drag them so that one is on top of the other. This is a key feature of GIS. In 1-2 sentences, why do you think this is important?

Submit your *Lab1* word document, titled "[YourName]_Lab1," via Canvas by Monday 1/27 at 11:59pm.

References and links:

Cutts, Andrew and Anita Graser. 2018. *Learning QGIS*. 4th Edition.
Monmonier, Mark. 2018. *How to Lie with Maps*. The University of Chicago Press: Chicago and London. 3rd Edition.
https://docs.qgis.org/3.4/en/docs/