3. Making a Choropleth: Demographic variation in the Southern U.S.

# Overview

We all know what a map should look like, but there are a surprising number of choices that go into map making and it is easy to make those choices bad choices. We will begin with the most arcane piece, projections, and cover just enough to make you aware that you need to think about it. From there we will go through the basic mechanics of making a choropleth map in QGIS. To make this map we need to think about how to classify our data, how to show this classification in terms of colors that make sense, and finally how to add the basic, but essential pieces of context that make a map informative. We will be working today initially with data for the entire world, and then shifting over to just counties in the southern U.S. that we used yesterday. This tutorial draws heavily on Chapter 4 of Introduction to GIS Using Open Source Software by Frank Donnelly and available in full [here](https://www.baruch.cuny.edu/confluence/display/geoportal/GIS+Practicum). That document goes into significantly more detail than we have time for here, but will serve as a good reference if you are looking for more information or want to take some of these concepts further.

# Coordinate Systems and Projections

This topic is the bane of lower level geography courses and has probably turned off more people to the art of mapmaking than any other topic. We are going to tread dangerously close to oversimplification here, but if you remember one thing, remember that getting this wrong can turn a reasonable analysis into garbage, so do try to make at least some basic reasonable choices. We will try to help you make those choices as painlessly as possible.

We are essentially dealing with two kinds of problem here:

* The first has to do with ***coordinate systems***. The earth is not flat, but it is not a perfect sphere or even a perfect ellipse. To work with geographic information at all we need to nail down three things: the geoid we are using to represent the earth, where its equator and prime meridian are drawn and what units describe distance from these starting points. These three things together make a Geographic Coordinate System (GCS). A common GCS is WGS 84 used by GPS systems around the world.
* The second has to do with the difficulty of transforming coordinates that are meaningful in three dimensions to a two dimensional map representation. It is impossible to make this conversion without distorting some of the relationships among points in space. ***Projections*** deal with this problem by preserving some important aspect of the true relationship among points represented in two dimensions; area, distance, or direction. Some projections work by being as accurate as possible in specific parts of the world and sacrificing that accuracy in others. Others may be perfectly accurate in one dimension at the expense of others. Finally some work towards a compromise that preserves enough accuracy in each dimension so that the world looks 'the way it should.' When we add a projection to a GCS we get a Projected Coordinate System (PCS). Albers Equal Area (area preserving), and Mercator (shape preserving) are two common examples.

***Task: Explore coordinate systems and projections in QGIS***

* Using your file explorer navigate to the folder "world" and open the file with the .prj extension (using any basic text editor)
  + Is this a projected or geographic coordinate system?
* Open QGIS and add the layer (Layer🡪Add Layer🡪Add Vector Layer, navigate to the class folder for today and add the file with the .shp extension)
  + Observe the text in the bottom right corner where it says "EPSG 4326" this is a shorthand code for the current coordinate system. Hover over it and it will reveal that this code refers to WGS 84
* Project this map layer into a projected coordinate system
  + Select the countries layer. Right click and hit Export - Save Features As
  + Give your file a new name: "world robinson"
  + Use the globe button to search for a new CRS
    - In "Filter" type "Robinson"\
    - Select "World Robinson" and continue
  + Make sure to "Add layer to map"
* So, nothing happened. Why?
  + By default QGIS sticks to whatever the first CRS added to the project was. To see our changes we will have to change the projection being visualized.
  + FWIW it is bad form to work with layers in multiple projections. QGIS will do its best to convert things on the fly, but it is a bad habit. If something goes wrong and the transformation isn't working well, you might not know for some time.
  + To change the displayed projection click on the Globe icon in the bottom right where it says EPSG.
    - Under "Recently Used Coordinate Reference Systems" you should find World\_Robinson. Select and apply
    - Try some others: Use the filter to find Lambert, Mercator, and whatever else you can think of.

The exercise up to this point has been about big geographic areas where the projection makes the most obvious difference. In the next section we will project the southern counties data used yesterday (and available in today's files as well).

* Close your project and start a new one
  + Add the southern counties with bvap layer to the project
  + What is the projection? What kind is it? GCS or PCS?

We are going to choose a strong compromise projection that emphasizes maintaining area. This tends to be good for visualization, it also means that if you happen to be doing some sort of areal interpolation on your data that it is likely to minimize your error.

* Right click on your data layer and Export🡪Save As
  + Search for EPSG 1002003 This is the Albers Equal Area projection for the contiguous US
  + Save this file as southern aea and add it to your layers box.
  + Remove the original (unprojected) layer and change the display projection to match your new data.

# Classifying Data

# Joining tabular and spatial data

In this section we perform a very simple data operation; loading spatial and tabular data into QGIS and joining them together based on a shared variable.

* 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 today and then the folder for 'southern counties'
  + In the Files of Type dropdown at the bottom of the window make sure the ESRI shapefiles option is selected. Select the layer.
* To add the tabular data simply drag the Excel file into the "Layers" box on the QGIS display
  + Right click on the properties of this file and open up Properties
  + Under "Source Fields" note the data types of the imported Excel data. Check to make sure all of your columns that are numbers imported as numbers. If not check the instructions on formatting in the Excel section above.
* Close the Data Source Manager
* Select the Southern Counties layer and right click on "Properties"
  + From Layer properties pick "Joins" from the left column
  + Hit the green plus button to add a join.
  + The join layer will be the excel file. The Join field in that table is FIPS. The Target field in the tract layer is GEOID.
  + At the bottom of the menu, check the box that says Custom field name prefix, and delete the text so that you just keep the names from your Excel file

# Mapping data

Tomorrow we will get into making maps in more detail, but spend the rest of our time exploring the various ways you can interact with the data. In particular, "Symbology" with "Graduated" and making sure to add a Classification will help you see some of your work.