Making a Choropleth with Census Data: Black Voting Age Population in the Southern United States

Adapted from 2019 Voting Rights Data Institute activities by Ruth Buck and Chris Fowler

Overview

This assignment will introduce you to working with Census data, Quantum GIS (a free and open source mapping program), through making a map that shows Black voting age population in the Southern United States. The core GIS techniques you will learn and demonstrate are: joining tabular and spatial data together, setting a map projection, classifying data by color in what is known as a choropleth, and adding essential information in a simple design to make a map informative.

You will likely run into questions – about Census data, mapping, and more. This is to be expected (and great)! You are encouraged to take note of them, use the resources linked to below, and otherwise do your best to make educated guesses without agonizing too much over perfect solutions. You can also reach out to your peers on the Slack channel to ask and answer questions – the GDBC organizers will check it occasionally, but it is mostly for you to support one another.

Before beginning this assignment:

- We highly recommend reading through Chapters 2 and 4 (skip section 2.5) of Frank Donnelly's *Introduction to GIS Using Open Source Software* before beginning this assignment (the entire tutorial is available here). We will be pointing to sections of it throughout, but a preliminary read-through will likely ultimately make it easier to complete the assignment.
- Download and install the latest long-term release of Quantum GIS (hereafter referred to as QGIS).

After completing this assignment, turn the following in here:

- Mandatory: A .pdf file of your map. The file name should be "FirstnameLastname-map.pdf" (with your name filled in!).
- Optional: Any pressing questions you ran into or thought of while going through this tutorial. We will not directly respond to questions before the event, but we will take note of them and make sure to address them during GDBC.

Accessing Census data

The Census has a number of ways that we can access their data. There are also several providers of repackaged census data designed for particular uses. The national historic GIS project (NHGIS) is one such provider. Associated with the University of Minnesota, NHGIS is an entity that has worked hard to improve the quality of the spatial data produced by the Census and to make data for older censuses available for use, facilitating longitudinal analysis. NHGIS is designed specifically for GIS users and is packaged in a way to make mapping relatively straightforward.

This tutorial uses NHGIS data tables, which you will gather, and spatial data files that we provide.

Note: The Census Bureau itself has an interface, data.census.gov, for distributing data. However, it is relatively new and much more difficult to use at this point, so this tutorial focuses on NHGIS.
NHGIS is designed specifically for GIS users and is packaged in a way to eliminate some of the problems

that we might otherwise find when working with census data. Today we will actually make things a bit more challenging by not using the GIS files from NHGIS, but for general use it is a great resource.

Task: Get county-level data to calculate Black Voting Age Population for Southern United States.

- Navigate to nhgis.org and select "Get Data."
- Set Filters:
    - Geographic levels = County
    - Years = 2010
    - Topics =Age AND Race (topic filter only)
    - Datasets = Decennial Census, 2010_SF1a
- Select Source Tables
- …Do not select GIS files (but take a look at what is there)
- Select the table "P10. Race for the Population 18 years and over"
- Continue through the rest of the process, making sure to select "include additional header row (best for spreadsheets)" You will need to sign up for a free account, and then you will get an email when your files are ready.

Cleaning data in Excel

Now that we have the data we will be using we need to create the measures we want to represent: Black Voting Age Population. To get there though we need to sort through a lengthy list of both geographic and population variables. In the previous step, you downloaded data on the population over 18. In this step, you will select only those persons who name Black as their only or at least one of their races.

Notes: part of the reason we are going through this exercise is to familiarize you with how the Census' race variables actually work. For some applications you may end up using more simplified race categories, but for other uses you may need something more complex (race and ethnicity together will double the number of categories for example).

- Check out the codebook first. Look at all of the categories for spatial and population data.
    - Open the csv file in Excel (or equivalent).
    - Narrow down to just the columns we need: GISJOIN, STATEA, COUNTYA, Total (H74001), Black or African American alone (H74004), multiple races including Black or African American (H74011, H74016-H74019, H74027-H74030, H74037-H74042, H74048-H74053, H74058-H74061, H74064-67, H74069, H74071). You can delete the other columns!
    - Create a new variable "BlackVAP" by summing all of the population variables together (except H74001).
    - Delete population variables other than total and your new variable.
    - Create a new variable BVAPPct by dividing BlackVAP by Total (H74001).
    - Create a new variable "FIPS" equal to STATEA*1000 + COUNTYA.
    - Delete the 2nd row of the table so that we just have one header row. Make sure all column names are meaningful, are less than 10 characters, and have no spaces or special characters (e.g., periods and other punctuation).
    - Convert the new variables from formulas to fixed values by selecting each column, copying, and then Paste□Values.
    - Convert all of your columns to "Numeric" format with no zeroes after the decimal for most columns, but two numbers after the decimal point for BVAPPct.
    - Save your file as an Excel workbook (.xlsx).

Joining tabular and spatial data

In this section we begin mapping by performing a very simple data operation: loading spatial and tabular data into QGIS and joining them together based on a shared variable. Note that if you mouse over buttons in QGIS descriptions will appear for many of them.

- Download and unzip (extract) the Southern_counties files available here. (Also included as an attachment in the instructions email.)
- Open QGIS.
- 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 folder where you put the southern counties files.
  - 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 tabular data layer and select "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 (e.g., integers). If they didn't, double-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 GEOid2.
  - 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. Then hit "OK."
  - Right click on the layer again and open the attribute table. Scroll to the right to confirm that your join worked: you should see the columns from your tabular data!
- Export the layer to make the join permanent by going to File □ Save Feature As… and saving it in the folder you are working in (save it as an Esri shapefile for now), using a name you'll remember (e.g., "southern_counties_joined"). After exporting, the joined layer should automatically be added to the map.

Coordinate Systems and Projections 1

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 go from the Earth to a sphere and then a sphere to a flat map. The first of these steps is done through a **Geographic Coordinate System**, and the second is through a **projection**. Understanding coordinate systems and projections is often one of the most challenging pieces of learning to make good maps. We make some simplifications that are spelled out in more detail in section 4.1 of *Introduction to GIS Using Open Source Software*, but some important features of coordinate systems and projections are as follows:

- A Geographic Coordinate System (often abbreviated to a GCS, or sometimes just called a coordinate system) is defined through three things: the shape (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.

Some common GCSes are WGS 84 and NAD 83.

- All projections necessarily distort (a common visual for projections is the challenge of flattening an orange), but with knowledge of projections you can make thoughtful decisions about which spatial relationships to preserve and which it might be more okay to distort. Projections preserve some important component of the original relationship among points: area, distance, or direction. Some may be perfectly accurate in one component at the expense of others. Others work towards a compromise that preserves enough accuracy in each dimension so that the world looks 'the way it should.' Also, some projections work by being as accurate as possible in specific parts of the world and sacrificing that accuracy as you go farther away. 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 set a projection in QGIS

- Using your file explorer navigate to the unzipped set of "southern_counties" files, and open the file with the .prj extension (using any basic text editor).
    - Consider: Is this a projected or geographic coordinate system?
- Back in QGIS, observe the text in the bottom right corner where it says "EPSG 4269". This is a shorthand code for the current coordinate system. Hover over it and it will reveal that this code refers to NAD 83.
- We are going to choose a projection that prioritizes maintaining the area of shapes (e.g., counties or states), but doesn't do a bad job at maintaining the shapes themselves. This tends to be good for visualization, and it also means that if you happen to be doing some sort of areal interpolation on your data that you can minimize the error.
- Right click on your data layer and Export□Save Features As
    - Search for EPSG 102003 This is the Albers Equal Area projection for the contiguous US.
    - Save this file as a shapefile (using a name that makes sense to you, such as "southern_counties_joined_proj") and make sure it gets added to your layers.
    - Remove the original (unprojected) layer and change the display projection (the button in the lower-right corner) to match your new data.
- Now, return to your file explorer and open up the .prj file of the projected layer. The projection should be different than that of the unprojected one. Is it?

Classifying Data

A choropleth map typically uses color to visualize difference in some data value for different subregions of a map. Our maps are going to examine demographic and voting data in the southern U.S. at the county level.

The starting point for a choropleth map is understanding what kind of variable you are hoping to represent. For most purposes we will have some variant of categorical, discrete, or continuous data. If our data is categorical, classification is already built into the variable, at least to some extent. Otherwise, we need a system to break our data into groups. QGIS offers a number of options, including quantiles, equal interval, and manual settings. Defaults are not at all aware of what message you are trying to convey with your map, so you won't want to use take them at face value -- think about what kind of difference yields a meaningful message and work from there. One of the easiest ways to "lie with maps" is in the choice of data classification scheme. For more information about the strengths and weaknesses of various data classification methods, see section 4.4.2 of *Introduction to GIS Using Open Source Software*, or take a look at Axis Maps' Basics of Data Classification.

Task: Experiment with data classification options

- Double click on the southern counties with bvap layer and select "Symbology" ( ).
- Start by changing the "Single Symbol" line at the top to "Categorical"
    - Under column select "STATEFP" and then "Classify." You should see a list of symbols appear.
    - Hit "Apply" to see how they look on the map.
    - Note the way colors are selected such that they emphasize difference among categories (in this case, states) without implying order.
- Change "Categorized" to "Graduated" and change the column to "BVAPPct." Hit "Classify" and "Apply" again to see the change.
- Experiment with different ways to classify by changing the classification "Mode" and the number of classes (the latter is to the right of the mode).
- Change the tab to "Histogram" and then hit the "Load Values" button. This allows you to classify data manually by moving the break points around.
- Keep hitting apply and noting how things change, until you've found a classification scheme that you think works for the data and the message you want to convey.

Color

Color is an underappreciated component of making maps visually compelling. Navigate to colorbrewer2.org These color palettes were designed to maximize the capacity to differentiate between classes (some of them are also colorblind and/or print-friendly). Note that human capacity for differentiating among classes is not great, so 3 to 6 classes will be recommended for most applications. If you would like more flexibility, another helpful resource for creating color palettes is the [Chroma.js Color Palette Helper]().

Task: Change the color scheme on your map

- Change the color ramp and number of categories to fit your message.

No Floating Polygons

It can be confusing to people less familiar with the topic to display your data with no context at all. Our next step is to add in some context, using a basemap pulled from the web. This will give a map user more points of reference to understand the information you are trying to present.

Task: Add a web-based basemap

- We will need to add a plugin to our QGIS installation.
    - From Plugins□Manage and Install Plugins search and add "QuickMapServices."
- Once installed, note what the icon looks like. You should be able to find it on the toolbar toward the top of the interface. Alternately, you can go to Web□QuickMapServices. Regardless of how you open QuickMapServices, pick a layer to put under your map.
    - *Note: this may take a few moments to load.*

Layout

The final step in our process is to create a new layout for printing our map or saving it to file. Section 4.5 of *Introduction to GIS Using Open Source Software* has a lot of useful information on this.

Start with Project□New Print Layout to open a new view where we can compose our map. Give your layout a name.

Task: Make a great looking map and export it to turn in!

- *Note: you can use the "Add Item" menu at the top or click on icons on the left of the interface to add the map, legend, and text.*
- Add a new map item. You will have to drag a box to show what portion of the paper you want this map to represent.
- Add a legend item
- Add text indicating the Source of your data.
- Save your map as a .pdf file by going to Layout Export as PDF. Please title your PDF as follows: "FirstnameLastname-map.pdf". Then, turn it in here.

Congratulations! You've completed the Geodata Bootcamp mapping tutorial!