Getting and Using Census Data: Black Voting Age Population in the South

Overview

In this unit we will go out into the wild and collect data that can tell us about the black voting age population in the southern United States. To get from start to finish we need to know a little bit about how the U.S. Census is organized a little bit more about how to access that data and we need some basic data cleaning skills. The core GIS technique that we will teach today is quite simple; just joining tabular and spatial data together and putting it on a map. Tomorrow we will think a little bit more carefully about how to actually make that map so that it communicates information well.

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. In our example today we will download information from the national historic GIS project 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. The Census Bureau itself is in a period of transition where it is migrating away from American Factfinder, which it has been its primary method for distributing data, to a redesigned data.gov portal that looks to be extremely difficult to use. 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
- Set Filters:
  - Geographic levels = County
  - Years = 2010
  - Topics =Age, Race
  - Datasets = Decennial Census, 2010_SF1a
- Select Source Tables
- Do not select GIS files (but take a look at what is there)
- 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, but it is fairly painless.

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. We will be downloading data on the population over 18 and then selecting out only those persons who name black as their only or at least one of their races. Note that part of the reason we are going through this exercise is to familiarize you with how the Census' race variables actually work. Note that 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: GEOID, STATEA,COUNTYA,Total (H7S001), Black

Alone(H7S046), H7S011,H7S016-19,H7S027-30,H7S037-42,H7S048-53,H7S058-61,H7S064-67,H7S069,H7S071

- Create a new variable "BlackVAP" by summing all of the population variables together (except H7S001).
    - Convert the new variable from formula to fixed values by selecting the column, copying, and then Paste→Values
- Delete population variables other than total and your new variable
- Create a new variable BVAPPct by dividing BlackVAP by Total (H7S001).
- Create a new variable "FIPS" equal to STATEA*1000 + COUNTYA
- Delete 2nd row of 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 like periods.
- Critical Formatting Step--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

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 tabular data layer 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.