

## SOC 4650/5650: PS-05

Christopher Prener, Ph.D.

March 14<sup>th</sup>, 2017

### Directions

Please complete all steps below. All requested deliverables should be uploaded to your GitHub assignment repository by 5:00pm on Thursday, March 23<sup>rd</sup>, 2017. This problem set uses data from the MOBoundary and MOHealth directories in CourseData as well as data you will need to download from American Fact Finder on county-level poverty rates.

The MOHealth directory contains two files. One file has rankings for all 115 counties in Missouri based on their overall population health. The second file has data on the number of low birth weight live births for each county in Missouri. These data are provided by the [County Health Rankings](#) project. Both poverty in childhood and low birth weight have been found to be associated with health outcomes in adulthood including “asthma, hypertension, diabetes, and stroke, heart attack, [and] heart disease”.<sup>1</sup>

The goal of this problem set is to produce three maps that illustrate both of these risk factors as well as an overall assessment of population health:

1. A map of poverty rates in Missouri by county (2010-2015 estimate),
2. a map of low birth weight live births in Missouri by county (for 2016),
3. and a map of health outcomes rankings for all Missouri counties.

Each of these maps will require altering (to varying degrees) tabular data and joining them with spatial data representing Missouri's counties.

### Accessing Missouri County Demographic Data

1. Using American Fact Finder's (<https://factfinder.census.gov>) advanced search tool, select counties as your search geography. Use the dropdown menus to further restrict your search to all Missouri counties, and then click Add to your Selections.

<sup>1</sup> Johnson, R. C., & Schoeni, R. F. (2011). Early-life origins of adult disease: national longitudinal population-based study of the United States. *American journal of public health*, 101(12), 2317-2324.

2. Use the Show Results From dropdown menu to restrict your search results to 2015.
3. Download Table S1701 from the **five year** American Community Survey estimates,<sup>2</sup> which contains data on the number of people living below the federal poverty line.<sup>3</sup> Be sure to download the data in csv format with annotations stored in *separate* files.
4. Move these data into your working directory for this lab.

<sup>2</sup> See [this week's jotter](#) for details on how the ACS is collected.

<sup>3</sup> See [this week's jotter](#) for details on how poverty is defined.

### *Editing the Poverty Data*

5. Using Atom, construct a well-formatted do-file using the headFull snippet. Be sure to edit the appropriate lines in the template that detail the name and purpose of the file.
6. Your do-file should *successfully* accomplish the following tasks. It should include narrative text that explains what each command accomplished.
  - (a) Import the raw data into Stata *selectively* so that the variable labels in row 2 are not imported but the data in rows 3 through 117 are.
  - (b) Drop all variables except for GE0.id2 and the estimate and margin of error variables for: the total population for whom poverty status could be determined and the total population living below the poverty line.
  - (c) Edit the variable names for these five variables so that they are clear, simple, and intuitive.
  - (d) Destring variables as necessary.
  - (e) Edit the format of the county FIPS code variable as necessary.
  - (f) Add dataset-level metadata (a label and dataset notes) describing the data, their source, and the modifications you have made. Also include information on who made the modifications and when they were made.
  - (g) Add variable-level metadata (variable labels and notes) describing each variable, their original names, and how they have been modified.
  - (h) Create a codebook for these data.
  - (i) Export these data both in csv and xlsx formats.
7. Execute the do-file and debug any errors until the code executes without issue.

8. Add the do-file, log-file, and a *tidied* markdown output file to your assignments repository.

### *Mapping the Poverty Data*

9. In a new map document in ArcMap, add the Missouri county boundary data from the M0Boundary directory.
10. Check the coordinate system for your data frame and ensure that it is set to NAD 1983 UTM 15N. If it is not, change the coordinate system to the UTM 15N system.
11. Join the exported poverty data you created in the last section to the county boundaries.<sup>4</sup>
12. Create an *appropriately normalized* thematic choropleth map that shows the number of persons living below the poverty line in each county. Use a color ramp selected from Color Brewer for this map.
13. Add and properly symbolize (as ground) layers representing the surrounding states. Label these polygons as well.
14. Use ArcMap's layout editor to produce a map layout for these data suitable for print dissemination. It should include a title, information about authorship, details about the data and coordinate system, and a legend only for the poverty data.
15. Export the layout as a pdf file at 300dpi. Add this image to your assignment repository.

<sup>4</sup> *Hint:* If you do not see your county FIPS code variable as an option, you may not have it stored in the format that matches the county FIPS code variable in your shapefile. If this is the case, you will need to re-export the data with the variable saved in an alternate data type.

### *Formatting the County Health Data*

16. Using Microsoft Excel, save the two csv files in the M0Health directory as *xlsx* files.

### *Mapping the County Health Data*

17. In a new map document in ArcMap, add the Missouri county boundary data from the M0Boundary directory.
18. Open up the attribute table and create a numeric version of the FIPS variable to facilitate joining the county health data to your shapefile.<sup>5</sup>

<sup>5</sup> *Hint:* As we discussed in Week 8, you will need to create a new numeric variable and then copy the values from the original FIPS variable into your new variable using the field calculator.

19. Add a second copy of the county boundary data from M0Boundary to a new data frame. Ensure that this copy of the county boundaries also has the new numeric version of the FIPS variable you created. If it does not, create this new variable again.
20. Check the coordinate systems for **both** data frames and ensure that they are set to NAD 1983 UTM 15N. If they are not, change the coordinate systems to the UTM 15N system.
21. In the **first** data frame, join the county health outcome ranking data to the counties layer. This join should use the same county FIPS code IDs that we used to join the poverty data above.
22. Create a thematic choropleth map that shows the county health rankings visualized by quintiles (i.e. five equal sized categories).<sup>6</sup> Use a color ramp selected from Color Brewer for this map.
23. In the **second** data frame, join the low birth weight data to the counties layer. This join should use the same county FIPS code IDs that we used to join the poverty data above.
24. Create an *appropriately normalized* thematic choropleth map that shows the number of low birth weight live births in each county visualized by quintiles. Use a color ramp selected from Color Brewer for this map that is different from but compliments the color ramp used for the county health outcome rankings.
25. Use ArcMap's layout editor to produce a single page map layout for **both** data frames suitable for print dissemination. It should include a title, information about authorship, details about the data and coordinate system, subtitles for each of the two maps, and a legend for each map.
26. Export the layout as a pdf file at 300dpi. Add this image to your assignment repository.

<sup>6</sup> *Hint:* use the quantiles method for data classification and ensure that you have five data classes.

### *Assessing Your Maps*

27. Look at all three maps individually and then together. In a *tidy* markdown formatted file (use the headMarkdown snippet in Atom), write a 1 to 2 paragraph assessment of the spatial distributions of poverty, low birth weight, and poor health outcomes (i.e. low ranked counties). In addition to assessing each map, discuss whether there are shared areas of overlap between two or more maps. Finally, in terms of public policy, what do you think we can learn from these maps?