

SOC 4650/5650: Lab-07 - Weather Data Organization

Christopher Prener, Ph.D.

Spring 2019

Directions

Using data from the data/lab-07/ folder available in the lecture-08 repository, create several maps using RStudio. Your entire project folder system should be uploaded to GitHub by Monday, March 11th at 4:15pm.

Analysis Development

The goal of this section is to create a self contained project directory with all of the data, code, map documents, results, and documentation a project needs.

- a. **Clone** the lecture-07 repository if you have not already done so.
- b. Create a project folder system with all of the necessary components, and drag the lab data from lecture-07/data/lab-07/ into your RStudio Project's data/ subdirectory. Also add an ArcGIS Pro project to this same directory, and move the default geodatabase into your data/ subdirectory.
- c. Create a README.md text file. Add a quick description of your project and outline the key directories and files that are included.
- d. Create a well-formatted RMarkdown document for your data cleaning efforts.
- e. in RStudio, load the three pieces of raw data - MO_BOUNDARY_Counties.shp, MO_DEMOS_CountyPop.shp, and MO_DEMOS_CountyDisability.csv.

Part 1: Data Wrangling

1. Begin by creating a pipeline in RStudio that takes the county disability data and:
 - (a) Creates a new variable that adds the total values for under50_dis, btwn50_99_dis, and btwn100_150_dis - this is the total number of people living with the county who have a disability and live between 0% and 150% of the poverty line.

- (b) Removes all columns except the GE0ID and the new variable you calculated.
- 2. Next, remove the NAME column from the county population data and convert it from a sf object to a regular data frame.
- 3. Then, join the county population and disability data together.
- 4. After the data are joined, create a new variable that contains the number of people per county who have a disability and live between 0% and 150% of the poverty line divided by the total number of people living in the county - the ratio of people we're particularly concerned about in in-climate weather to the total population of the county.
- 5. At this point, there should be four columns in the data set in this order - GE0ID, the new variable you created with the county of people in poverty who have a disability, the total population, and the ratio of those two variables. Re-order your columns if necessary.
- 6. Remove all columns from the county shapefile data except the unique identification variable that matches the unique identification variable in the demographic data set you've created.
- 7. Join the county geometry to the demographic data set.
- 8. Calculate the square kilometers per county by getting the area of county from the geometry column and converting it from square meters to square kilometers.
- 9. Write the data as a shapefile to your data/ folder.

Part 2: Geodatabase Creation

Add your new shapefile and as well as the metro weather event shapefiles for tornado and hail events to your ArcGIS Project's geodatabase in the data/ subdirectory.

Analysis Development Follow-up

Don't forget to knit your document when you are done! Also be sure to go back and update your README.md file with any changes to your project's organization or contents.