SOC 4650/5650: PS-06 - Stroke Rates by County

Christopher Prener, Ph.D.

Spring 2018

Directions

Using data accessed from the `lecture-10` repository, `USHealth`, and the `USBoundary.gdb` geodatabase, create the map below illustrating stroke rates by county for all fifty states. Your entire project folder system, including data and notebook output, should be uploaded to GitHub by **Monday, April 2ⁿᵈ** at 4:15pm.

Part 1: Analysis Development (Review from Lectures 01 and 02)

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. Make sure to include all relevant directories, a well formatted notebook, and a 'README' that traces the changes you make to all of your data.

Part 2: Data Preparation

The goal of this section is to produce two shapefiles from the raw data provided. These shapefiles should illustrate stroke rates by county in the United States as well as the point location of state capitals.

1. Add the following data to a subfolder of `data/` named `rawData/` in your lab's folder hierarchy:

   (a) In `ArcCatalog`, open the `USBoundary.gdb` geodatabase in the course data release and export the U.S. county boundary data to a shapefile.

   (b) Copy-and-paste the file `US_HEALTH_noIns.csv` from the `US_HEALTH` folder in the course data release.

   (c) Copy-and-paste the `stateCapitals.csv` data from the `lecture-10` repository.

2. Using R, complete the following steps:

   (a) Import the `stateCapitals.csv` data and project it using the included `x,y` coordinate data. Check your projection using `leaflet` before exporting the data. The resulting data should be exported as a shapefile with the `NAD 1983` geographic coordinate system applied.[1]

   (b) Import both the county boundary data and the stroke data, and complete a table join to combine both data sets.

   (c) There are some values of `-1` in the stroke data. Those are "missing" counties that the CDC does not provide stroke rate estimates for. To remove them, subset your observations so that you only have observations remaining where the variable `strokeRate` is greater than or equal to `0`.

   (d) The resulting data should be exported as a shapefile with the `NAD 1983` geographic coordinate system applied.

[1] Feel free to recycle code from this week's lab or replication file to easily re-create these data!

Part 3: Mapping Stroke Rates for the Contiguous United States

The goal of this section is to produce a data frame showing the stroke rate by county in the contiguous United States (i.e. the "lower 48" states).

3. In a new map document, add both the county health insurance estimate and state capital shapefiles created in the previous section. Also add the state boundary data from `USBoundary.gdb` to your map.

4. Select a projected coordinate system for this map that is appropriate for mapping the contiguous United States (i.e. the "lower 48").

5. Create a thematic choropleth map for that shows variation in the number of individuals without health insurance. Make sure to use a Color Brewer palette as well as Jenks Natural Breaks with 5 data classes for your symbology.

6. Overlay the state boundaries (symbolized with a hollow fill) to make it easier to identify states that have particularly high rates of stroke mortality. Also overlay the county boundaries on your data so that they are above the choropleth data but below the state boundaries. These are important because there are a number of counties that are missing stroke mortality estimates. Symbolize these with a hollow fill, and make sure that the hue for the outline

matches the hue used for country boundaries on the choropleth
layer.

7. Add and appropriately symbolize ground layers using the data
available in `USBoundary.gdb`.

Part 4: Mapping Stroke Rates for Alaska

The goal of this section is to produce a data frame showing Alaska's
stroke rate by borough (the equivalent of counties in Alaska).

8. In a new data frame, copy the data from the previous section
and change the extent of the map so that it shows only Alaska.
Notice the counties that are white - these are the counties that had
missing data that we managed in our query in the first section.

9. Change the projected coordinate system of this second data frame
so that it is appropriate for mapping Alaska .

Part 5: Mapping Stroke Rates for Hawaii

The goal of this section is to produce a data frame showing Hawaii's
stroke rate by county.

10. In a new data frame, copy the data from the previous section and
change the extent of the map so that it shows only Hawaii.

11. Change the projected coordinate system of this third data frame
so that it is appropriate for mapping Hawaii.

Part 6: Creating and Evaluating the Map Layout

The goal of this section is to produce a well designed map layout that
includes inset map for both Alaska and Hawaii. Make sure they are
clearly delineated from the main map - the contiguous United States
(and surrounding ground layers). It is typical for these insets to be
placed over Mexico.

12. Create your map layout, and be sure to provide all of the relevant
data about your map layout (title, authorship, projection systems,
etc.) as well.

13. Export your complete map layout as a `pdf` file at `300dpi`.

14. In a new section of your `README.md` file, evaluate in a paragraph the spatial distribution of stroke rates. What areas of the country appear to have higher rates of stroke mortality?