# *SOC 4650/5650: Lab-A - Health Insurance Rates by County*

*Christopher Prener, Ph.D.*

*Spring 2020 (Remote)*

## *Directions*

Using data accessed from the `lecture-10` repository, `USHealth`, and the `US_BOUNDARY_Counties` shapefile, create the maps below related to health insurance rates by county for all fifty states. Your entire project folder system, including data and RMarkdown output, should be uploaded to GitHub by **Monday, April 6th** 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. Please ensure **all** required elements are present. You will need to add the `US_BOUNDARY_Counties` shapefile to your project. You will also need the `US_HEALTH_noIns.csv` and `stateCapitals.csv` files.

## *Part 1: Data Preparation*

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

1. 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 `mapview` before exporting the data.
   (b) The resulting data should be exported as a shapefile with the `NAD 1983` geographic coordinate system applied. The shapefile should be saved to a subfolder of `data/` named `cleanData/` in your lab's folder hierarchy.
   (c) Import both the county boundary data and the health insurance data, and complete a table join to combine both data sets.

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

(e) Subset your data to remove observations that are not located in the contiguous United States. You will need to remove Alaska, Hawaii, and all overseas territories from your state capitals and insurance data frames.

(f) The resulting data should be exported as a shapefile with the `NAD 1983` geographic coordinate system applied. The shapefile should be saved to a subfolder of `data/` named `cleanData/` in your lab's folder hierarchy.

## *Part 2: Mapping Health Insurance Data for the Contiguous United States*

The goal of this section is to produce a stand-alone map of the contiguous United States (i.e. the "lower 48" states) that shows the percent of individuals lacking health insurance by county using either `ggplot2` or `tmap` (your choice!) in `R`.

2. Select a projected coordinate system for this map that is appropriate for mapping the contiguous United States (i.e. the "lower 48") - either the Albers or Lambert projected coordinate systems. Apply this to both the state capital and the county data.

3. Create a thematic choropleth map for that shows variation in the number of individuals without health insurance.

4. Overlay the county boundaries (symbolized with a hollow fill) to make it easier to identify counties that have not seen large decreases in the uninsured population since the introduction of the Affordable Care Act.

5. Add state capitals on top of these two layers as point data.

6. Export the map image as a `pdf` at `300dpi`,