CS/INFO 3300; INFO 5100
Homework 5
Due 11:59pm Tuesday March 28

Maps and linear models

In this homework you will create two choropleth maps, showing gun ownership and population density for each US state. You may use code from class notes.

1. The file `gunownership.txt` contains a dataset on gun ownership percentages by state along with population and density information, compiled from the Washington Post and Wikipedia. The file `us.json` contains geographic data. Write code to load the provided data file and the US map data asynchronously. (10 pts)

2. In a <p> tag add an `<svg>` element create a map of the United States using the `d3.usaAlbers()` projection. Set the fill for each state based on the gun ownership percentage in the data file. Use a `quantize` scale, which takes a min and max value as its domain and an array of values as its range. The scale divides the input domain into discrete intervals, one for each value in the range array. The resulting scale function maps input values within each interval to the associated output value from that array. In this case, the output array will be a set of 10 colors. Use a constant hue, but vary the lightness and/or saturation. (20 pts)

3. In a second <p> tag add an `<svg>` element create another map colored according to the log of each state's population density. Create a second `quantize` scale, using `d3.max()` to find the largest population density in the dataset. (HINT: pass log values to a `quantize` scale, don't use a `log` scale.) Use the same colors as before, oriented so that lower densities and higher gun ownership percentages have the same color. Below the map describe why the log density is important -- what happens if you don't use it? (20 pts)

4. In this problem you will create a third <p> tag with an `<svg>` element containing a scatterplot that compares log population density on the x-axis to (reported) gun ownership on the y-axis. Write text responses below the `<svg>`.

a. For each state, add a small `circle` and a `text` showing the name of the state. Include axes and labels. (10 pts)

b. Calculate the slope and intercept of the linear regression line using log population density as your input (*x*) variable and gun ownership as your predicted (*y*) variable. Plot that line in blue. (10 pts)

c. Is log density a good predictor of gun ownership rates? Are there states for which log density is less accurate as a predictor of gun ownership? For example, are there states

whose gun ownership is either higher or lower than you would expect based on their density? Describe how this pattern appears on the maps and how it appears on the scatterplot. (10 pts)

d. Calculate the slope and intercept of the linear regression line with the variables reversed (ie predict log density given gun ownership), and plot the regression line in red on the same scatter plot. Remember that you will need to plot the line with respect to the opposite axes. Are the two lines the same? Why or why not, and if not, under what conditions will they be the same? (10 pts)

e. Calculate the Pearson correlation coefficient for gun ownership and log population density (https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient). How does this correlation value relate to the slopes of the two regression models you created in the previous sub-problems? (10 pts)