CS/INFO 3300; INFO 5100
Homework 7
Due 11:59pm Tuesday May 10

Maps and linear models

The file `gunownership.html` contains a dataset on gun ownership percentages by state along with population and density information, compiled from the Washington Post and Wikipedia. The file also contains code based on an example by Mike Bostock that shows a choropleth map of gun ownership.

In this homework you will create a second map that shows population density for each state. Each sub-problem involves edits to this map. You only need to show the final version, but include text answering the included questions about how the map looks after each edit. The example uses 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 is a set of colors. (Use the free-response at the top of the page for text responses.)

1a. Create a new map underneath the sample map, colored according to population density. (HINT: copy and paste the example code and change the displayed variable.) For the `quantize` scale, use `d3.max()` to find the largest population density in the dataset. Is this map useful? Why or why not? (20 pts)

b. Modify the code to display the *log* of each state's density. (HINT: pass log values to a `quantize` scale, don't use a `log` scale.) Is this map more useful? Why or why not? (10 pts)

c. Finally, reverse the color scheme for this plot, so that higher densities are shown in a lighter shade of blue than lower densities. How do the two maps (gun ownership and log density) compare? (10 pts)

2. In this problem you will create a third `svg` element containing a scatterplot that compares log population density on the x-axis to (reported) gun ownership on the y-axis.

a. For each state, add a small `circle` and a `text` showing the name of the state. (HINT: the `d3.values()` function may be useful in converting an object to an array.) Include axes and labels. (10 pts)

b. Calculate the slope and intercept of the linear regression line using log population density as your x variable and gun ownership as your y variable. Plot that line. (HINT: you may borrow code from class notes. You should be able to figure out which.) (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. 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)