Lab 5: working with your project data

Public Health 460

Due: Friday April 16th @ 5:00pm EST

This lab will get you started working with the dataset for your project. One point of this lab is that everyone in your group will bring a different perspective to the data. As usual, you are encouraged to work together on the nuts and bolts of reading in data and getting things to work. However, you are encouraged (although not required) to keep the details of your work private. This is not for academic honesty reasons, but to try to give your group a chance to develop several different perspectives on what graphics to make. If you all work together, you are more likely to end up with a single kind of graphic. If you each work separately, you might find more unique ways of looking at the data, giving you more options for your final product. After the assignments have been submitted, you are encouraged to share your work with your teammates, see what they did, and figure out which (if any) of the approaches might serve a good basis for some of the work on your project.

- You are required to use the main dataset that your group is using. You may use external datasets as well. You must your own graphs using your own code.
- Every group member must turn in their own assignment, as an Rmd and HTML file.
- There are 10 questions for credit.

1) Read in your data using a relative file path (1pt)

2) List the variables that you will focus on for this assignment. Identify the "outcome" variable(s) in the dataset. Write down a few hypotheses about what relationships you might find in the data. (1pt)

3) Generate a "big picture" or "birds eye view" plot that shows as much of the data and variation in the data as possible and addresses one of the hypotheses you wrote down in the previous question. (2pts)

4) In addition to the graph you just made in 3), add a feature to the graph that emphasizes/bolds/highlights a couple of observations (or groups of observations) on the big picture graph. Spend time to make the graph visually pleasing with colors and appropriate visual cues. (4pts)

5) Write down the process you used to select your bold/highlighted observations? (1pts)

6) Take your bold/highlighted observations and create an additional figure, set of figures, or a table that focuses in on those observations and provides more detail about them. Do not lose the the visually pleasing theme you worked on in the earlier questions. (2pts)

7) Do the graphics you made follow your hypothesis about the data? Yes or no? Explain in 1-2 sentences, it is okay if they did not. (1pts)

8) Randomly choose other observations (or groups of observations) to bold/highlight. Redo 4) and 6) using these new observations. (3pts)

9) Does changing the bold/highlighted observations change the way you think about the data after looking at the similar plot in 6)? (1pt)

10) Write a 3-4 sentence conclusion including the following points (4pts)

- What is the main, big-picture take-away from your above analysis?
- What do you learn from focusing on a few of the observations or groups of observations?