Lab 2: Reproduce an analysis

Public Health 460

Due: 8pm, Friday February 25th, 2022

The deliverable for this lab is a written analysis submitted both as an HTML and Rmd file. The files must be submitted on Moodle by 8pm on Friday, February 25, 2022. Late assignments will not be accepted. Note that the instructor and TAs will not be available for assistance after 5pm on Friday.

Spend some time browsing some data science and statistics blogs (e.g. R-Bloggers, FiveThirtyEight GitHub repo with data and code, Variance Explained, The Economist's Graphic Detail data on GitHub,...) or an academic journal that has a reproducible analysis. You might find some analyses on the Reich Lab blog. You could also browse vignettes from various R packages (try running `browseVignettes()` to see a list of available vignettes that are in packages that you have installed on your machine). Pick an article/entry/vignette that you find interesting. You will need to find an analysis that has made both data and code publicly available. If you leave this until the last minute, you will not find a good article, so spend some time looking for an article that has data early in the week.

As always you may collaborate with fellow students in the class on your choice of article. You can even choose the same article as another classmate and work together on the "Analysis" portion below. Remember, however, that you may not cut and paste code, so while you may work together on the assignment, you must write up your own analysis. And your "Follow-up" analysis as detailed below MUST be different from anyone with whom you choose to work.

The assignment

The deliverable for this lab a "blog post'' (i.e. an HTML file and the accompanying RMarkdown file) of your own. All code should be shown in the post. The file must be submitted on Moodle by 8pm on Friday, February 25, 2022. Late assignments will not be accepted. The assignment must contain your name, and the date. The post must contain the following sections:

- (2 points) Introduction: This section should summarize and provide a link to the original post.
- (8 points) Analysis: The reproduced analysis from the original report. If your chosen article has code then you can use identical code to the original report, but do not copy any expository text. You may have to change some of the code to make it run on your computer.
- (8 points) Follow-up: An addition to the original post that provides some original data analysis and visualizations that you create. This can be a minor modification of an existing analysis (e.g. the same visualization with a different variable). The 8 points for this section is comprised of 3 points for doing something beyond the original post, and 5 points for making it an interesting and more challenging addition to the story. For example, just changing one variable is not a very challenging modification. Be adventuresome here! Try out some new things. Points will be awarded for creativity and demonstrated effort.
- (2 points) Discussion: Summarize your findings in a few sentences. Also, reflect on the activity and identify a few lessons learned and things you would do differently next time.