Lab 2: Reproduce an analysis!

Overview

For this lab, you will work with a partner at your table to reproduce a published data analysis (blog, academic paper, etc...). If your table has an odd number of people, you may work in groups of three.

Spend some time browsing some data science and statistics blogs (e.g. Stats Blogs, R-Bloggers, FiveThirtyEight, Variance Explained, ...) or an academic journal that has a reproducible analysis. Pick an article/entry that you both find interesting. You will need to find an analysis that has made both data and code publicly available. You should fill out this Google Form with info about your selected article by 5pm Thursday, September 20.

The assignment

The deliverable for this lab a "blog post" (i.e. an html page, knitted with RMarkdown) of your own. All code should be shown in the post. The post must contain the following sections:

- (5 points) Introduction: This section should summarize and provide a link to the original post.
- (25 points) Analysis: The reproduced analysis from the original report. 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.
- (15 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 15 points for this section is comprised of 5 points for doing something beyond the original post, and 10 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.
- (5 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.

Handing in files for this project

To hand in the draft of your reproducible analysis, one member of your team will need to use RMarkdown to create an HTML document. Note that when you create/"knit" the file in RStudio, you will see a little "publish" button on the top right corner of the file. Click this button. You will be prompted to create an "RPubs" account (please do so) and then the file can be shared via a public URL. Once the file has been uploaded, verify the link and send the link to the instructors/TAs via the Google Form. Please note, if you would rather not use your full names on the documents, please just use your first names and your table number. You must send the link by 5pm on Friday, September 28.