CSC 1930.01: Data Science
Spring 2018 – Assignment 4
Dr. Ben Mitchell

Instructions

This assignment is focused on visualization, and has two parts. The first part involves studying some visualizations made by others and writing a short response; the second part is to make your own vizualization of the same data.

This is a paired assignment; the same rules apply as for previous paired assignments. There is no requirement that you work with the same partner as on the previous assignment; in fact, I encourage you to switch around who you work with, since this will be good practice for being a data scientist who needs to work with a diverse group of collaborators.

Part A

Go to the MakeoverMonday homepage (http://www.makeovermonday.co.uk/) and read what the site is about; then check out this entry: http://www.makeovermonday.co.uk/week-2-2018/. Also be sure to read the original article (https://yougov.co.uk/news/2017/11/29/personality-more-important-looks-across-globe/) Read the text, then focus on the examples in the "favorites" area at the bottom of the page. Select two examples (based on whatever criteria you like), and respond to the following prompts about each of them.

- Who made it?

- What's the title?

- What's the overall story being told?

- Describe the type of visualization being used, and how it helps to tell that story.

Part B

The second part involves making your own visualization of the same data. Since this is likely your first attempt at doing this type of task, start simple and then try making progressively more complex visualizations. You'll want to begin by loading the data set (available from the course Resources page) into a Pandas dataframe using the `read_excel` method, as shown in the starter code.

As the MakeoverMonday page makes clear, the most important thing in work like this is to make sure you are "telling a story" with your presentation of the data. This means both the actual graphs or plots and the surrounding text (in your case, this should be Markdown); I'm not just looking for a bunch of graphs, I want graphs plus a small amount of text that convey some meaningful message. A big part of the job of data science is to help others interpret data; this assignment is fundamentally about communication more than it is programming and statistics (though of course some programming and statistics are going to be used in the process, and as always communication requires you to have done some interpretation and judgement to decide what you want to say).

That said, the visualization of the data should be the primary focus; I'm not looking for an essay with accompanying graphics, so much as graphics with a bit of accompanying (or integrated) text. Again, use the examples from the MakeoverMonday sight as a guidepost here.

The exact design of your visualization is up to you, but there's a Jupyter Notebook posted for this assignment that has a step by step walkthrough of a simple example to get you started. For more details on visualization with pandas, check out Chapter 9 of the McKinney text. Also, check out the documentation for matplotlib, seaborn, and pandas visualization, including the example galleries:

https://pandas.pydata.org/pandas-docs/stable/visualization.html

https://matplotlib.org/2.1.1/tutorials/introductory/sample_plots.html

https://seaborn.pydata.org/examples/index.html