DSCI 510: Principles of Programming for Data Science

Preliminary Analyses HW4

Tala Tayebi

Github Repo: https://github.com/talajune/DSCI510_FinalProject¶

1. What kind of analyses or visualizations did you do? [5 points]

I am in the process of making a map visualization to show the different distributions of happiness scores across countries. I have also created a scatter plot based on the highest correlation coefficient graph I made with a country's happiness ranking. I am most interested in seeing which features/attributes have the highest correlation coefficient and then doing more visualizations from there. To do this properly, I need to go back into my data collection phase and collect data from all of the 200+ countries on the World Happiness Report and include more features (AKA, don't drop so many columns). I also want to add the "region" column that I previously dropped, just so I can do some analsyis based on regions of the world and their average happiness scores. From here, I think my final visualizations will include a choropleth map showing happiness score distributions, some scatterplots showing the highest correlation between happiness and variables, and also either a) a dot plot with different-sized dots showing different regions of the world and their happiness score relative to certain features or b) a line graph showing the differences for every region. Realizing that I need to make changes to my dataset and data collections set me back a little bit, but I will work on it in the next couple of days.