MSIS 2506: R Programming

Project 2, Fall 2019

Logistics:

Assigned: Tuesday, November 5, 2019

Due: Day of Final Exam, Fall 2019 (11:59PM)

Objective:

Perform a data analysis on a data.gov dataset using R.

The requirements for this project are:

- Select a dataset from data.gov (http://catalog.data.gov/dataset)
- Your data must have both categorical variables and numerical variables
- Find 3 interesting facts/patterns about the dataset ad present your findings with the use of graphics – at least two tables and at least three charts/graphs
- Tell a story with the data
- Work individually
- Use the dplyr library to compare and contrast at least three categories

Guidelines for judging 'interesting':

There are multiple ways things get to be 'interesting'. Here's tow of the best heuristics we know:

- This fact/pattern is so interesting you would make your discoveries a topic of conversation.
- This fact pattern is crucial to understanding the topic: e.g. For Gerrymandering, that would be something like 'There are *x* many districts, that are up for debate every *y* years, and *z* are the decision-makers. If *a* many districts shift red/blue, then the odds of the election swaying is way *b%* higher." By the way, this would be one of the three sections – not all three in one.

Storytelling & Visualizations:

- Present the data in a manner which draws people in and keeps them engaged
- Be concise, clear, concrete, correct, coherent, complete and courteous (7 C's of communication)
- Use comments for code
- Pictures are worth a thousand words. Use them to distill complicated data into an easily graspable chart or table.

Resources:

- https://catalog.data.gov/dataset
- https://www.kaggle.com

Collaboration:

You will work individually on the assignment, but you are allowed and encouraged to use Google extensively, as well as any online resources (cite your work!), class notes, and R documentation.

Submission:

- Name your final file <your_username>_project3_fall2019 (mine would look like dvrdoljak_project1_fall2019)
- Submit an R Notebook with both your analysis and your code. Plot the charts/graphs inline, and make sure to explain your analysis.
- Make sure it runs completely and correctly on your computer
- Submit it via Camino
- (We will run your program on our computer to test your answers)

Section	Grade	Criteria
Interesting fact 1	20%	Interesting, factful, good analysis
Interesting fact 2	20%	Interesting, factful, good analysis
Interesting fact 3	20%	Interesting, factful, good analysis
Use of comments, readability, & General Submission	15%	Directions followed correctly
R Notebook	25%	Correct use of R Notebook; markdown correctly used; Graphs plotted inline; Good and Correct Presentation of Results in an R Notebook