

# MSIS 2506: R Programming

## Project 2, Fall 2019

### Logistics:

Assigned: Thursday, October 15, 2019

Due: Saturday, November 2, 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>)
- Choose a dataset which no one else has (including the other section) and record it here
- Find 3 interesting facts/patterns about the dataset and present your findings with the use of graphics – at least one table and at least one chart
- Tell a story with the data
- Work in groups of 3 to 4

### Guidelines for judging ‘interesting’:

There are multiple ways things get to be ‘interesting’. Here’s two 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>
- <http://jupyter.org/>
- <https://matplotlib.org/>
- <https://seaborn.pydata.org/>
- <https://pandas.pydata.org/>
- [https://www.mindtools.com/pages/article/newCS\\_85.htm](https://www.mindtools.com/pages/article/newCS_85.htm)
- <https://datavizblog.com/2013/05/26/dataviz-history-charles-minards-flow-map-of-napoleons-russian-campaign-of-1812-part-5/>

## Collaboration:

You will work individually on the assignment, you are allowed and encouraged to use Google extensively

## Submission:

- Name your final file `<your_username>_project1_fall2019` (mine would look like `dvrdojak_project1_fall2019`)
- 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	25%	Interestingness, factfulness, analysis, presentation
Interesting fact 2	25%	Interestingness, factfulness, analysis, presentation
Interesting fact 3	25%	Interestingness, factfulness, analysis, presentation
Use of comments & readability	20%	Documentation of author and dates; Explanation of steps Use of whitespace; Use of new lines; Naming convention of variables; Sequencing of code and outputs

General & Submission	5%	Directions followed correctly
----------------------	----	-------------------------------

**Bonus:**

- Find a set of data online that requires scraping. Scrape it and analyze it instead of data.gov data. If you choose this step, clear it with your instructor first so they can validate your choice of data.
- SUPER BONUS: perform a Kaggle competition and place it on the leaderboard. Again, clear with your instructor.