

# MSIS 2607 - Winter 2020 - Project 1

## Logistics:

Assigned: Thursday, January 23, 2020

Due: Thursday, February 6, 2020

## Objective:

Perform a data analysis on a data.gov dataset using Jupyter Notebooks.

## The requirements for this project are:

- Select a dataset from data.gov (<https://catalog.data.gov/dataset>)
- Choose a dataset which no one else has (do your best to avoid using someone else's dataset).
- Use Jupyter Notebooks and submit the .ipynb file
- 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

## 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 go to a party and say: "Guess what I found out about xyz!"
- 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 to red/blue, then the odds of the election swaying one way is  $b\%$  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, and Jupyter elements for storytelling
- 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-napoleon-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\_winter2020.ipynb (mine would look like dvrdojak\_project1\_winter2020.ipynb).
- 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)

## Grading Rubric:

Section	Grade	Criteria
Interesting Fact 1	20%	Interestingness, factfulness, analysis, presentation, data preparation
Interesting Fact 2	20%	Interestingness, factfulness, analysis, presentation, data preparation
Interesting Fact 3	20%	Interestingness, factfulness, analysis, presentation, data preparation
Use of comments & Readability	20%	Documentation of author & dates; Explanation of steps Use of whitespace; Use of new lines;  Naming convention of variables; Sequencing of code and outputs
General & Submission	20%	Directions followed correctly , code is correct and error free