CSC-148 Fall 2020

**Final Project**

**Project Overview**

This project will give you more experience processing and analyzing real-world data using Python. You will get to know and practice more useful packages. We would like to ask you to process the *U.S. President 1976–2016* dataset provided by MIT Election Data and Science Lab in 2017, and use the knowledge you have learned in this class to produce some insights. The final write-up and code of the project is due **Friday, December 4** before midnight, even though you should be able to report your results already during class on December 4th.

**Dataset**

We will use the *csv* file downloaded from MIT Election Data and Science Lab, and you will need to use the *panda* library in Python to read and convert it to a dataframe. You are also welcome to use other methods to process the data, as long as you are able to generate the desirable results. The dataset can be accessed through the link below:

<https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/42MVDX/MFU99O&version=5.0>

You can download the **1976-2016-president.csv** file directly from Blackboard.

**Teamwork**

The final project can be accomplished individually or with a team of max of 2 students. Each team member ‘s contribution will be peer reviewed. Team project has higher expectations than individual work.

**Required Tasks**

You need to complete the following tasks and translate the data into meaningful insights which help you to increase your understanding of the data. This is a practice of increasing awareness and sharpening your perspective, so that you will be able to use these analytical skills to a wide range of real-world data-driven challenges from a variety of industries, such as marketing, education, public sector services and beyond.

1. Find the total number of votes of the country for the election from 2000 to 2016.

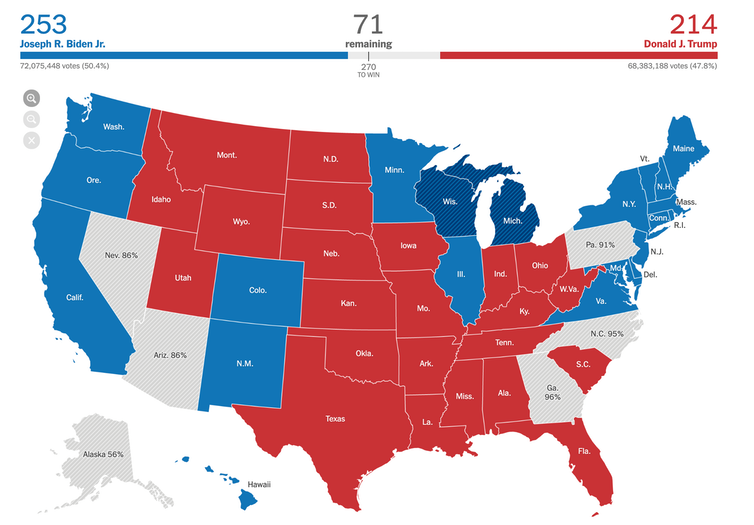
2. Compare the number of votes and percentages for Republican to those for Democrat across states you are interested in from 2000 to 2016.

3. Find the Top 5 states that have the highest supportive rate for Republican and Democrat in 2012 and 2016 elections.

4. Pick a state (e.g. CA) and track its votes over time for both parties between 1976-2016.

5. Visualize the above statistics, using histogram, pie chart, line chart and other appropriate plots. Be creative of choosing the right graph.

6. Plot the results of an election year of your choice on a US map (see matplotlib/basemap) with different colors representing R and D. Here is an example:



7. *Open Questions:*

Besides the aforementioned tasks, please find **TWO** other insights that you think they extract valuable and actionable information, and try to interpret these insights. For example, you can track changes of party, state, and find out which state flips the most.

**Coding styles:**

1. Your code must be well-organized, easy to read, and have as few global variables as possible. Use functions and returns to pass results. That means most of the code should be in *def* and *class*.
2. Import all the modules on top of your program.
3. You should have a *main( )* function that controls the user interface and sends different arguments to call different functions. Question2.py from Lab 9 sets a good example.
4. You should write specifications for each function. Please follow the link below for style guides in names, comments, etc.

<https://www.cs.cornell.edu/courses/cs1110/2018fa/materials/style.php>

1. Please use appropriate comments and docstrings to annotate your codes.
2. Extra bonus (3-5pts) might be considered if you submit your well-documented and correct code to Github after the project deadline. <https://guides.github.com/activities/hello-world/>

**Submission of the project**

1. All files should be saved in a folder named Project\_Lastname (for example, Project\_Smith), this folder should be **zipped** and submitted on Blackboard.
2. Project\_Lastname folder contains the following:
3. All necessary .py files that help you to generate the results and insights.
4. **1976-2016-president.csv** file
5. **2-page PDF report t**hat includes all necessary plots, and your interpretation of the insights. The report should be named **Report\_Lastname.pdf** (for example, Report\_Smith.pdf)
6. Each student is expected to turn in their own code/PDF even if they worked in a team.
7. Please make sure if you work in a team that clarify the names of the team members.
8. You can also submit powerpoint PPT slide for your final presentation.

**Presentation**

**Every team/individual will present the results/figures of the project in our lab class on December 4th. Each team has 5 minutes to present. You can make a short powerpoint presentation of your main visualizations and findings.**

**Toolkit:**

1. Importing csv using *pandas* <https://pandas.pydata.org/pandas-docs/version/0.23.1/generated/pandas.DataFrame.from_csv.html>
2. Plot histograms:

<https://data36.com/plot-histogram-python-pandas/>

1. Plot Bar charts, Pie Charts, and Scatter plots with Matplotlib and Numpy

Import Matplotlib.pyplot as plt

Import numpy as np

<https://matplotlib.org/>

1. Plot Data on US map

Matplotlib

https://matplotlib.org/basemap/

<https://stackoverflow.com/questions/39742305/how-to-use-basemap-python-to-plot-us-with-50-states>

<https://towardsdatascience.com/easy-steps-to-plot-geographic-data-on-a-map-python-11217859a2db>

Here are some examples

<https://medium.com/data-visualization-weekly/20-electoral-maps-visualizing-u-s-election-results-a1bec436d87f>

**Grading schemes:**

1. Weights of the components

|  |  |
| --- | --- |
| **Name** | **Percentage** |
| Completion of Required Tasks and Open Questions. | 60% |
| Use the 2-page Report to show your analysis of the insights, and your understanding of the data. | 20% |
| Coding Styles, Organizations, Clarity, Comments, User Interface. | 10% |
| 5 minute Presentation | 10% |

1. Break down of Completion of Required Tasks and Open Questions. The total points will then be put into percentage for final grade.

|  |  |
| --- | --- |
| **Task** | **Points** |
| Required Task 1 | 10 |
| Required Task 2 | 10 |
| Required Task 3 | 10 |
| Required Task 4 | 10 |
| Required Task 5 | 10 |
| Required Task 6 | 10 |
| Required Task 7 | 10 |
| Open Question 1 | 15 |
| Open Question 2 | 15 |

Project Design Credits: Chenxi Liao and Bei Xiao