Design Proposal

TP3 updates are on the next page

I.Project Description

The name of my term project is “Road to 270!”. This app allows the user to visualize the results of this year’s presidential election, compare the results to previous years, understand the public opinion on social media and visualize data on presidential speeches.

The project will have three parts in total. Since this is a Data Visualization project, all parts will involve some type of plots.

1) Map Visualization: In this part, the user would be able to see the color-coded map of the country and when clicking each state, the details on each state for this year and previous years is visualized.
2) Twitter Sentiment Analysis: Using the twitter handle and the hashtag that the user provides, the user can visualize the sentiment analysis results, which would show the public opinion of the specific candidate on a specific topic of interest(economy, COVID..etc)
3) Speech Language Analysis: The user can upload any presidential speeches and can see the graphical representation of the analysis on that particular text.

II.Competitive Analysis

My project is similar to the election visualizers on most online news platforms. Some similarities between my project and these visualizers is that the data on the election is tailored specifically for each visualization so that the user can access specific trends, associations and comparisons. However, one big difference is that the news platforms are focused on pure visualization of the logistical data from the election(e.g. Number of votes, voter turnout rates, comparison with polls) but my project involves social media, speeches made during presidential campaigns to represent the public opinion as well as giving an in-depth insight into what the candidate’s focus was.

III.Structural Plan

The structure of the entire project is based on Object Oriented Programming for animations, especially involving the usage of the App class, and the subclassing of ModalApp and Mode. The main ModalApp class branches out into separate modes. In this case, there are three modes in total, the Map mode, the Twitter Mode and the Speech Mode. In the Map Class, data on the 2020 election and data on previous elections(1976-2016) are imported from csv files that I have produced. The csv files will be imported from the data folder, which is within the same directory as the main file. The Map Class will have a subclass called VisualizeState so that each state’s data could be visualized under a single outline. Most of the data imported from its superclass would be utilized here. For Twitter mode, data will be scraped from Twitter and based on that, plots will be generated. Lastly, in the Speech Mode, the user will upload a .txt file of a speech by a presidential candidate or choose a file already in the directory ‘speechfiles’, which are preloaded files.

IV.Algorithmic Plan

The most algorithmically complex part of this project is the sentiment analysis of tweets and speeches. I will be using many types of the Naive-Bayes algorithm to conduct this sentiment analysis, and use voting strategy to optimize the algorithm. Some of the classifiers that will be involved are the MNB classifier, the LogisticRegression classifier, the BernoulliNB classifier and many more. The results from these classification will be verified through the use of confidence percentages.

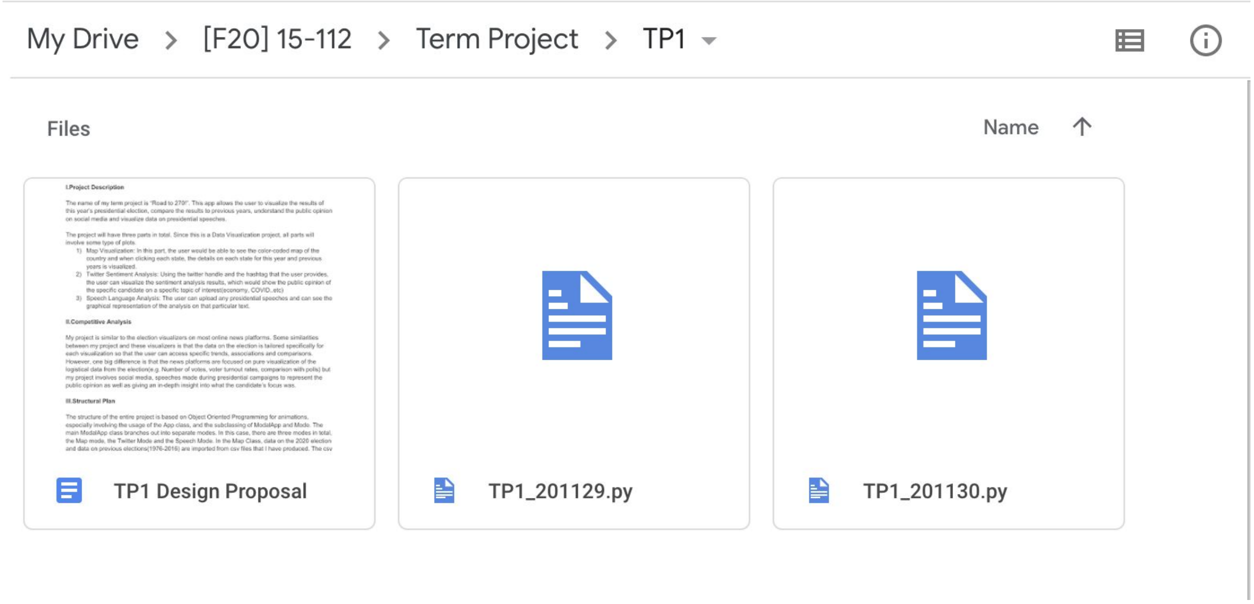
Another algorithmically complex part of this project is producing algorithms for each specific visualization. Since I am not using any modules for graphing, I am creating my own visualization techniques within 112 graphics. Processing the data to extract the information that I need and implementing that dataset to various plots is the most difficult part.

V. Timeline Plan

11/23 - TP0
11/24 - Making a Splash Screen, creating the map visualization
11/30 - Complete the Map Mode, including the plots in the subclass, TP1
12/1 - 12/5 - Complete Twitter Mode, Complete Map Mode
12/5 - TP2(hope to reach MVP here)
12/5-12/9 - add additional features, finish debugging, improve on user experience

VI. Version Control Plan

The image below is my Google Drive folder for the TP and I am saving daily backup files in this folder.



I.Project Description

The name of my term project is "Road to 270!". This app allows the user to visualize the results of this year's presidential election, compare the results to previous years, understand the public opinion on social media and visualize data on presidential speeches.

The project will have three parts in total. Since this is a Data Visualization project, all parts will involve some type of plots.

1) Map Visualization: In this part, the user would be able to see the color-coded map of the country and when clicking each state, the details on each state for this year and previous years is visualized.
2) Twitter Sentiment Analysis: Using the twitter handle and the hashtag that the user provides, the user can visualize the sentiment analysis results, which would show the public opinion of the specific candidate on a specific topic of interest(economy, COVID..etc)
3) Speech Language Analysis: The user can upload any presidential speeches and can see the graphical representation of the analysis on that particular text.

II.Competitive Analysis

My project is similar to the election visualizers on most online news platforms. Some similarities between my project and these visualizers is that the data on the election is tailored specifically for each visualization so that the user can access specific trends, associations and comparisons. However, one big difference is that the news platforms are focused on pure visualization of the logistical data from the election(e.g. Number of votes, voter turnout rates, comparison with polls) but my project involves social media, speeches made during presidential campaigns to represent the public opinion as well as giving an in-depth insight into what the candidate's focus was.

III.Structural Plan

The structure of the entire project is based on Object Oriented Programming for animations, especially involving the usage of the App class, and the subclassing of ModalApp and Mode. The main ModalApp class branches out into separate modes. In this case, there are three modes in total, the Map mode, the Twitter Mode and the Speech Mode. In the Map Class, data on the 2020 election and data on previous elections(1976-2016) are imported from csv files that I have produced. The csv

VII. Module List

- Pandas - for loading csv files
- Tweepy - for scraping data off twitter
- nltk - to analyse text

"TP2 Updates"

I.Project Description

I originally had a speech mode as one of the modes that I wanted to implement but I have decided to remove that so that I can do a more in-depth approach to the two modes that I currently have. I might decide to reimplement it if I have extra time at the end.

IV.Algorithmic Plan

For now, I have not decided to use the naive-bayes algorithm so that I can focus more on perfecting the visualization instead of trying to learn a difficult algorithm. Therefore, I left it till after MVP so that I can take more time to understand it and decide upon whether I want to use it in my project or not.

"TP3 Updates"

III.Structural Plan

For Twitter mode, data will be scraped from Twitter and based on that, I have decided to generate treemaps rather than wordclouds.

IV.Algorithmic Plan

For Twitter mode, the most algorithmically complex part was using recursive backtracking to place each rectangle that represents the frequency of the top 10 words. In order to check whether placing the rectangle is a legal move is the most complex part because it has to be done with multiple for loops that can track multiple points at the same time. If the size of the rectangles have multiple factors and hence can have multiple orientations(e.g. A rectangle of size 6 can be 1x6 2x3 3x3 or 6x1), this algorithm will allow exploring of all the possible orientations so that all pieces can fit in. The proportion of each rectangle will be in relation to the total frequency of all words combined so that regardless of the size of the total frequency, the rectangles can all fit into the given canvas.