2020 US Election Tweets Analysis Visualization Project

NAME: YUE ZHANG STUDENT ID: 30976316

LAB NO: 02 TUTOR NAME: ANANYA PANDEY

Contents

1. Introduction	2
2. Design	2
Methodology:	2
2.1. Sheet 1: brainstorming	2
2.2. Sheet 2: Alternative design sheet	2
2.3. Sheet 3: Alternative design sheet	3
2.4. Sheet 4: Alternative design sheet	4
2.5. Sheet 5: Final design sheet	5
3. Implementation	6
3.1 General layout and navigation.	6
3.2 Library used in this project	6
3.2 Tab 1: General Distribution	6
3.3 Tab 2: Word Inside Tweets	7
3.3 Tab 3: Tweets VS Votes	8
4. User Guide	9
4.1 Open page	9
4.2 Introduction page	9
4.3 General Distribution page	10
4.4 Word Inside Tweets page	12
4.5 Tweets VS Votes page	13
4.6 Conclusion page	15
5. Conclusion	16
6. Bibliography	16
7. Appendix	17

1. Introduction

The main topic of this narrative visualisation is the tweets analysis of 2020 US Election related tweets of the two candidates, Trump and Biden. The 2020 US Election is a very large event in US and has numerous related tweets on Twitter. In addition, the two candidates analysed in this visualisation project are two very popular president candidates at that time.

Therefore, based on the background, the intended audience including:

- (1) Election candidates: they may be interested on how tweets affect votes of them.
- (2) Trump or Biden's supporters: they may want to know how people talk about the two candidates or what are people's sentiment when talking about them.
- (3) Other people using Twitter that curious about 2020 Election: they may want to know how the tweets distribute of the two candidates.

Regarding the intended audience listed above, this narrative visualisation is mainly focusing on three parts:

- (1) Tweets distribution of these two candidates in terms of time and other aspects (country, source).
- (2) What is the word or sentiment inside each tweet?
- (3) Comparison and corelation of tweets and votes for these two people in each state in US.

In this project, the original data sources used are listed below:

- a. US Election 2020 Tweets. Oct 15th 2020 Nov 8th 2020, 1.72M Tweets (two dataset inside)
- b. US Election 2020 Race to Presidential Election 2020 by County

2. Design

Methodology:

To design for this project, the five sheet design methodology is used to implement the design process.

2.1. Sheet 1: brainstorming

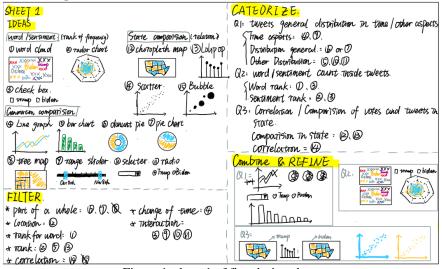


Figure 1. sheet 1 of five design sheet

In the brainstorming process, first various kinds of ideas to solve the three questions are listed. Then they are filtered by different use and categorized into different questions for this project. Finally, they are combined and refined together for the visualisation in different questions.

2.2. Sheet 2: Alternative design sheet

In this sheet, the layout is divided into three tabs, each tab represents a question mentioned in sheet 1.

For tab 1, I intend to show general distribution by a donut pie chart since it is easier to view who occupies what percent of the total distribution. Also, I choose line graph to show the distribution in different time, and it is control by a range slider, which means that users can choose the time period they want to view, and the line graph will change corresponding to user choice from range slider. The distribution in other aspects (country, source, hours and weekdays) for total, Trump related tweets and Biden related tweets are shown using bar chart, and user can use selector to switch between aspects. The advantages are various kinds of charts and interaction supplied by controllers. However, this design enough interactions, and there is many repeated charts (bar chart and pie chart).

For tab 2, word cloud is chosen for showing the frequent words rank in tweets, and for sentiments, I use a radar chart to view the sentiment distribution of these two candidates. The chart choice for this part is good, as the word cloud can attract users, and it is good to illustrate word ranks, but for the check box that controls the view of each candidate, it is useless as the radar chart can show the distribution of the two people at the same time.

For tab 3, to demonstrate the comparison of tweets and votes, two choropleth maps are used, and the colour in each states represents which candidate has more tweets or votes in this state. For example, if one state in the tweets map has orange colour, it means that in this state, Trump has more tweets. Also, when user click on each state, a tooltip will show to list the details number (tweets/votes for Trump and tweets/votes for Biden). Furthermore, the two scatter plots give the information about the corelation between tweets and votes for the two people. The map is a good choice to show location distribution, but the tooltips may contain too many information and it is hard for user to read all of them. Also, there is no interaction in corelation part.

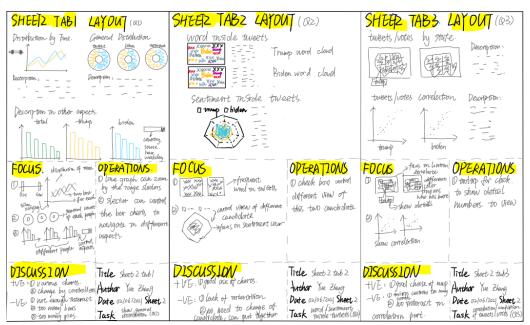


Figure 2. sheet 2 of five design sheet

2.3. Sheet 3: Alternative design sheet

Based on the advantages and disadvantages analysis in sheet 2, some changes are made as the alternative design.

For tab 1, to reduce the repeat of bar chart, a selector is added to let users choose to view distribution of total, Trump and Biden. However, there is still not enough interaction, and the selector is repeated in one graph. In addition, although it is hard to show distribution of hours and work days by line graph, they are more likely belongs to the time distribution instead of other distribution.

For tab 2, an alternative design of sentiment analysis is listed in this sheet, which consist of a lollipop chart to show each sentiment count and rank with a selector. This design can let user to switch between distribution in total or for the two people. The disadvantage for this design is that the selector is frequently used in tab 1 and the lollipop seems less attractive than radar chart for this part.

For tab 3, instead of having tooltip to show the detail numbers, I add two pie charts for the click event on each map. If user click any state on the map, the two pie charts will show the detail numbers, which is easier to compare these two candidates' tweets count difference in a particular state. The use of graphs is good but there is still lack of interaction in corelation part.



Figure 3. sheet 3 of five design sheet

2.4. Sheet 4: Alternative design sheet

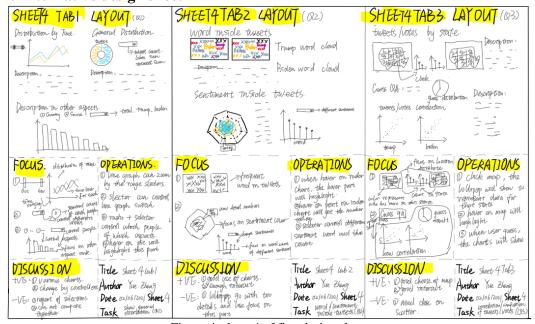


Figure 4. sheet 4 of five design sheet

In this sheet, some common interactions are added in each tabs, such as highlight or tooltips when hover.

For tab 1, the aspect of hours and weekdays are deleted to simplify the distribution analysis since source and country are more popular analysis aspects, and the control of the view of the aspects are changed to radio input, which reduce the repeat of selector. In addition, the pie charts are combined with a selector to switch between count of tweets, sum of likes and sum of retweets. Besides, link interaction is added, which is that if the line graph changes by the range slider, the pie chart will also change according to the count in the given time period.

For tab 2, the radar chart and lollipop graph are used together to show the sentiment analysis, and user can select from the selector to view the words rank for a particular sentiment. However, this design may go into too many details, and only sentiment distribution may be enough to show and focus.

For tab 3, the pie charts in the click event are replaced with lollipop graphs, which can better show the comparison of tweets or votes count in each state. In addition, to increase the interaction, one guess game is added in corelation part, the user need to first guess is there correlation between tweets and votes. After they click on their guess, the pie chart will show about how many people choose yes and how many people choose no, and the scatter plots will also show after guessing.

2.5. Sheet 5: Final design sheet

Based on the analysis for sheet 2,3 and 4, the final design is combined in sheet 5. Each tab contains two parts according to the detail information.

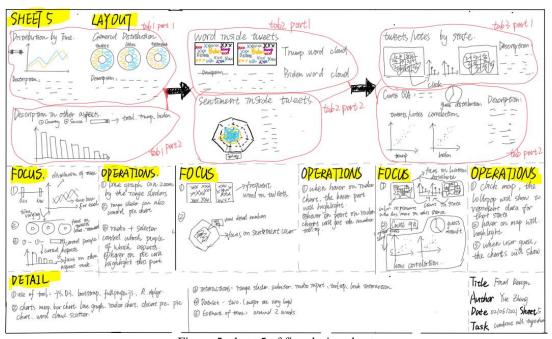


Figure 5. sheet 5 of five design sheet

Tab 1 part 1: (1) One line graph to view the distribution in different date, hover on the line will show detail information by tooltips. (2) three pie charts to show distribution in different people in given time period (default is the whole period), hover on the pie chart will make the hover part highlighted. (3) One range slider controls the time period, the pie charts and the line graph will both change when the time range in range slider changed by users.

Tab 1 part 2: (1) One bar chart to view distribution in a given aspect. (2) Selector controls how details to view (total, Trump only and Biden only), radio controls which aspects to view (country or source).

Tab 2 part 1: (1) Two word clouds to show and compare the frequent word ranks for these two candidates.

Tab 2 part 2: (1) One radar chart to view distribution of sentiment for the two candidates, and hover in radar chart will lead to the highlight of the selected parts, and the tooltip of the detail number for the selected sentiment.

Tab 3 part 1: (1) Two choropleth maps to illustrate who has more votes/tweets in each state in the US. (2) When state is clicked, two lollipop graphs will show to represent the detailed comparison of votes/tweets of the two candidates in this state. (3) Hover on map will highlight the selected state, and click on map will show the name of the state on the map.

Tab 3 part 2: (1) One interactive guess question to attract users. (2) After user guess, the charts will show. (3) One pie chart of the users guess result. (4) Two scatter plots to show the corelation, and the initial show of the scatter plots will have animation.

Details of the design (level of difficulty): (1) Use of different graphs. (2) Very large original dataset (the major one is more than 400 MB and has more than two million rows) (3) Multiple dataset (two original dataset are used together), an additional geojson file is used to draw the map for d3. (4) Use of js.d3 programming. (5) Link interaction, animation and other interaction for charts.

In terms of the communications, the three tabs focus on each question for this project, and no irrelevant information is presented in each section. Also, to increase the interaction for communication, different interaction approaches are added in each tab. Furthermore, the narrative parts are near to the location of each charts to help user understand and explore about the data and charts.

As for the human perceptual system, adequate charts are chosen for data to represent information. For example, line chart is chosen to represent change with time (date distribution in tab 1), map is chosen to represent locational distribution, world cloud is chosen to represent the word rank to catch human eyes, scatter is chosen for showing the correlation, pie chart is chosen to show the part of a whole and other charts, like bar chart, lollipop and radar chart is more suitable to demonstrate the rank.

3. Implementation

3.1 General layout and navigation.

I have used bootstrap, fullpage.js and simple slider to help form the general html layout. The navigation between different tabs is control by user scroll event or the navigation list on the top left of the screen. The order of scroll event is: introduction page, tab 1, tab 2, tab 3 and conclusion page, which is shown in below figure.

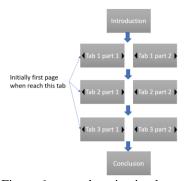


Figure 6. general navigation layout

3.2 Library used in this project

Html layout: bootstrap 3.3.7, fullpage 2.8.9, d3-simple-slider (for the layout of range slider)

Data visualisation: d3 v4

Data wrangling: the data used in this project is already processed in the data exploration project using r, addional preprocessing (count and sum to make it suitable for use in d3) is added using the dplyr library in r studio.



Figure 7. library used in this project

3.2 Tab 1: General Distribution

The three pie charts are implemented by pie function in d3 library. For the line graph, first the dataset is split by nest function and then the graph is implemented by path in the svg using d3 library, and the hover All the hover event is set by customising the mouseover, and mouseout event.

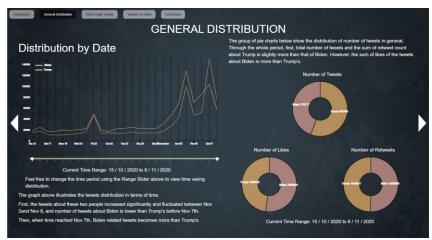


Figure 8. tab 1 part 1 implementation

The bar chart is implemented using the rect in svg using d3 library, the interaction of controllers is implemented by adding the change of bar chart inside the onchange listener of selsctor and radio input.



Figure 9. tab 1 part 2 implementation

3.3 Tab 2: Word Inside Tweets

The word clouds are built using text and by the layout.cloud function in d3 library. The location of each text, different colours and font sizes are set inside my own function.

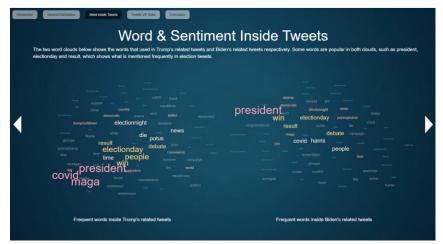


Figure 10. tab 2 part 1 implementation

The radar chart is implemented by changing the layout of a pie chart, the area in the radar chart is built using path in svg, the point is implemented using circle. The tooltip hover interaction is implemented by my own function that defined in the mouseover and mouseout event.



Figure 11. tab 2 part 2 implementation

3.3 Tab 3: Tweets VS Votes

The two choropleth maps are built using the geoPath function in d3 library with AlbersUsa projection. For the maps, a geojson file of states geo information is used. The lollipop graph is initialized and changed using merge function and line and circle to implement. The click and hover event is implemented by setting the on mouseover, on mouseleave and on click event. There is link interaction for the maps, which is that is one map is hovered or clicked, the same effect will show in both the two maps. For example, if one state is hovered in tweets map, the highlighting will show on both the two maps. This interaction is also achieved inside the on mouseover and mouseleave event.

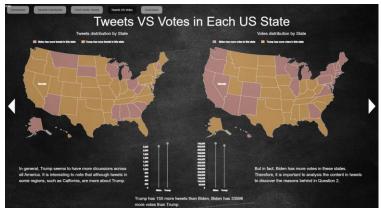


Figure 12. tab 2 part 1 implementation

The charts that hide and show is controlled inside the onchange event of the yes or no input. The pie chart live data (users' response) is get and stored using the localStorage.getItem and localStorage.setItem function. The scatter is implemented by dots in svg using d3 library, and the animation of showing the scatter is using the delay function when creating the plot.

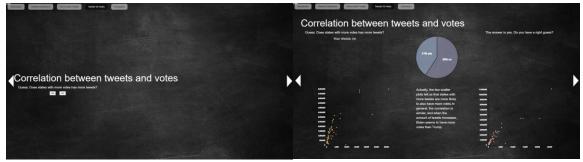


Figure 13. tab 2 part 2 implementation

4. User Guide

4.1 Open page

The page runs in chrome, the html file is called visulisation_project.html, and it is opened using the live server in vs code.

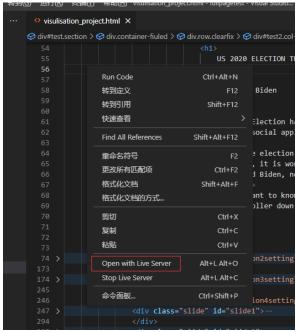


Figure 14. open with live server

4.2 Introduction page

After opening the page, you will jump into the introduction page, and have a briefly overview of what the project is, and which aspect will be shown in this project. Then you can scroll down to the next page, or click the navigation on the top left to select aspect to view.



Figure 15. introduction page

4.3 General Distribution page

When scrolling down on the Introduction page, or clicking 'General Distribution' on the navigation, the General Distribution page will show, it consists of two parts, including distribution by date and other distribution, and you can control the arrow to switch between them.

The first page you will see is the general distribution by date.

General charts:

- (1) A line graph shows the tweets distribution in terms of time
- (2) Three pie charts that represent the distribution of number of tweets, number of likes and number of retweets respectively in general in the given time range (the default is the whole period).

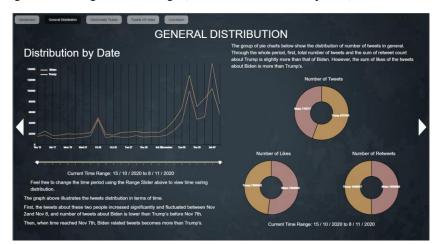


Figure 16. first page of general distribution page

Interaction:

(1) You can hover on the line graph, and it will show the tooltip contains the value information of both the two candidates for the point you focus on.



Figure 17. hover on line graph

(2) You can hover on the pie chart, and the parts you hover on will be highlighted.

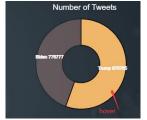


Figure 18. hover on pie chart

(3) You can change the range slider to view data in a particular time range, and the pie charts and line graph will change according to your choice.

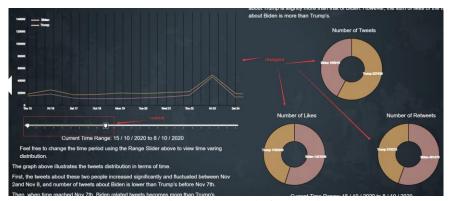


Figure 19. control of range slider

The second page you will see is the general distribution in other aspects

General charts:

(1) A bar chart of the distribution in a selected aspect.

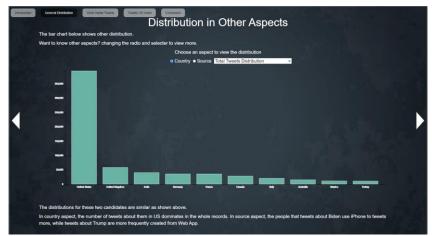


Figure 20. second page of general distribution page

Interaction:

(1) You can use the radio to control which aspects to view (country or source, the default is country)

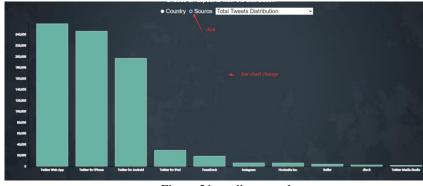


Figure 21. radio control

(2) You can also use the selector to control how details you want to view, including total distribution, Trump related tweets distribution and Biden related tweets distribution.

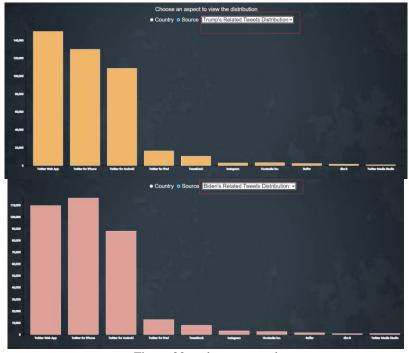


Figure 22. selector control

4.4 Word Inside Tweets page

When scrolling down on the General Distribution page, or clicking 'Word Inside Tweets' on the navigation, the Word Inside Tweets page will show, it consists of two parts, including word inside tweets and sentiment inside tweet.

The first page you will see is the general distribution by date.

General charts:

(1) Two word clouds of the frequent words inside Trump's related tweets and Biden's related tweets.

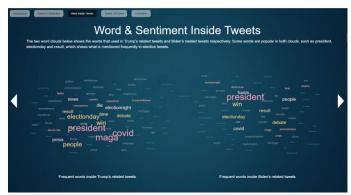


Figure 23. first page of word inside tweets page

The second page you will see is sentiment analysis.

General charts:

(1) A radar chart shows the distribution of ten different sentiment.

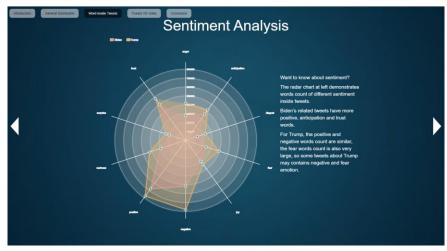


Figure 24. second page of word inside tweets page

Interaction:

(1) You can hover on the area, and the part you hover on will be highlighted.

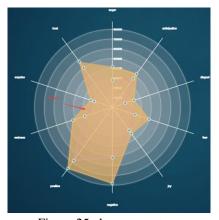


Figure 25. hover event on area

(2) You can also hover on the point, and the tooltip will show to let you know the exact number for this sentiment.

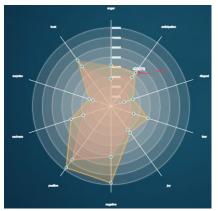


Figure 26. hover event on point

4.5 Tweets VS Votes page

When scrolling down on the Word Inside Tweets page, or clicking 'Tweets VS Votes' on the navigation, the Tweets VS Vote page will show, it consists of two parts, including tweets vs votes in each US state and correlation between tweets and votes.

The first page you will see is the tweets vs votes in each US state

General charts:

- (1) Two maps to show who has more tweets/votes in this state.
- (2) (Only show when click) Two lollipop graph of the exact count in the state that clicked by user.

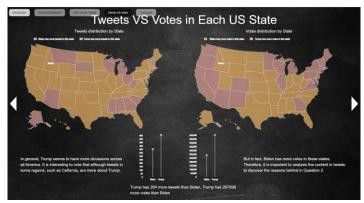


Figure 27. first page of tweets vs votes page

Interaction:

(1) You can hover on each state, and the state you hover will be highlighted in both maps.



Figure 28. hover event on state

(2) You can also click on each state, and the state you click will show the name on the map, and the corresponding lollipop graphs that represent the numbers in this state.

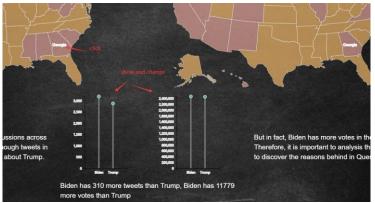


Figure 29. click event on state

The second page you will see is the correlation between tweets and votes.

General charts:

- (1) A guess question that you can click for your answer.
- (2) (Only show after guess) One pie chart of the users guess response.

(3) (Only show after guess) Two scatter plot of the corelation

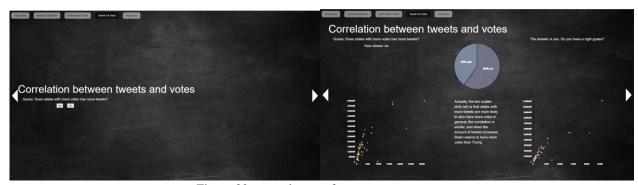


Figure 30. second page of tweets vs votes page

Interaction:

(1) You can guess the correlation, and after guess the other charts will show.



Figure 31. guess question

(2) The initialization of the scatter plot shows the animation.



Figure 32. animation

4.6 Conclusion page

Finally, you scrolling to the final page, which is conclusion page, it contains the review of all the questions we have explored and the original data source information.

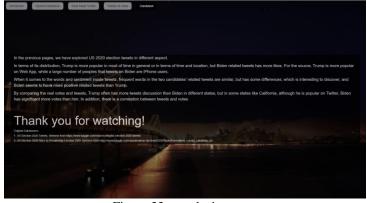


Figure 33. conclusion page

5. Conclusion

In conclusion, in this project, there are three main tabs that focus on three different questions that the intended audience may be interested to and related to the exploration project, which are:

- (1) What is the distribution in terms of date, country and source?
- (2) What is the frequent word and the sentiment count inside the tweets?
- (3) What is the comparison and correlation between tweets and votes in each US state?

To communicate those questions, the design process of five design sheet is included, and the implementation is the same as the final design decisions in design sheet five. In design process, different visualisation charts and communication methods are filtered and chosen based on the communication propose and the human perceptual system. In implementation process, the charts is implemented using javascript d3, some common shapes and svg elements to create different charts are used.

In user guide, all the user instructions from start the project to all the details in each tabs are shown to let user understand how to use this visualisation project. Also, narrative explanation of each charts and focus is provided in an appropriate way.

In conclusion, the visualisation project runs well and includes all the things that need to communicate with the users.

6. Bibliography

Mark Otto, a., 2021. Bootstrap. [online] Getbootstrap.com. Available at: https://getbootstrap.com/

cdnjs. 2021. fullPage.js - Libraries - cdnjs - The #1 free and open source CDN built to make life easier for developers. [online] Available at: https://cdnjs.com/libraries/fullPage.js/2.8.9

US Election 2020. (2021). Retrieved 7 June 2021, from https://www.kaggle.com/unanimad/us-election-2020?select=president county candidate.csv

US Election 2020 Tweets. (2021). Retrieved 7 June 2021, from https://www.kaggle.com/manchunhui/us-election-2020-tweets

7. Appendix

