THE UNIVERSITY OF SYDNEY

Assignment Coversheet – GROUP ASSIGNMENT

Please fill in your details below. Use one form for each group assignment.

Assignment Details:

Assignment Title	**Hillary Clinton vs Donald Trump**				
Assignment number	**Assignment 2 (Group)**				
Unit of Study Tutor	**Katherine mu**				
Group or Tutorial ID	**CC-08-G06**				
Due Date	**04/11/2024**	Submission Date	**04/11/2024**	Page Count	**20**

Membership

Group Name/Number	**CC-08-G06**				
Family Name	Given Name (s)	Student Number (SID)	Unikey	Contribution + percentage	Signature
Li	Weihao	500021893	weli0358	Task1(Idea + report +evaluation) + 25%	Weihao Li
Wang	Yuhan	510031279	ywan3092	Task1+Task2+Task3(visual + idea + report) + 25%	Yuhan Wang
Fu	Yidan	510027504	yifu6380	Task2(idea + report +format) + 25%	Yidan Fu
Huang	Albert	530215563	qhua0468	Task3(idea + report) + 25%	Albert Huang

Declaration:

1. I understand that all forms of plagiarism and unauthorised collusion are regarded as academic dishonesty by the university, resulting in penalties including failure of the unit of study and possible disciplinary action.
2. I have completed the **Academic Honesty Education Module** on Canvas.
3. I understand that failure to comply with the Academic Dishonesty and Plagiarism in Coursework Policy can lead to the University commencing proceedings against me for potential student misconduct under Chapter 8 of the ***University of Sydney By-Law 1999*** (as amended).
4. This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that it is not my own by acknowledging the source of that part or those parts of the work.
5. The assessment has not been submitted previously for assessment in this or any other unit, or another institution.
6. I acknowledge that the assessor of this assignment may, for the purpose of assessing this assignment may:
   a. Reproduce this assignment and provide a copy to another member of the school; and/or
   b. Use similarity detection software (which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking).
7. I have retained a duplicate copy of the assignment.

Please type in your group number here to acknowledge this declaration:	**CC-08-G06**

How would you visualise your data?

COMP5048/4448 Assignment 2

line 1: Weihao Li
line 2: *500021893*

line 1: Yuhan Wang
line 2: *510031279*

line 1: Yidan Fu
line 2: *510027504*

line 1: Albert Huang
line 2: *530215563*

line 4: CC-08-G06
line 5: Katherine Mu

I. TASK 1

This task first required sifting through the tweets that the two candidates had posted in 2016 to identify the major topics discussed by each candidate and compare how the various topics had changed over time. Then, by analysing the various topics over time, it examined how well the topic was taken up in tweets to determine its effectiveness. This will be useful for subsequent analysis of the social media strategies of different candidates and their impact on the election.

The first step was to identify the main topics discussed by each candidate by calculating and observing the high-frequency words that appeared in the tweets. To identify topics, the text of tweets was analysed to identify the thirty most frequent words used by the two candidates, using bar charts (Figure one).

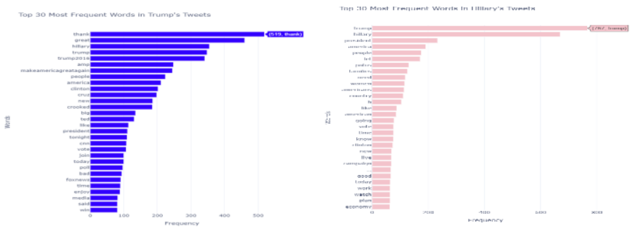


Figure 1

The interactive bar chart above allows for an effective comparison of the frequency of use of vocabulary. By comparing different lengths, it is easy to see the frequency of vocabulary used by different candidates. Using interactive bar charts allows for deeper analysis of the charts' content, either by selecting the desired words with the mouse to hide other words for further comparison and analysis, or by moving the mouse to see specific word frequencies.

The bars show that Trump's tweets use words like 'thank', 'great', 'makeamericagreatagain' and 'trump' and other words of this kind more frequently, which may indicate that part of the purpose of his tweets is to create a positive image in the election campaign and to focus on self-promotion. Hillary's top 30 words include words like 'president', 'families' and 'plan', reflecting her focus on political content and social issues. The focus is on political content and social issues.

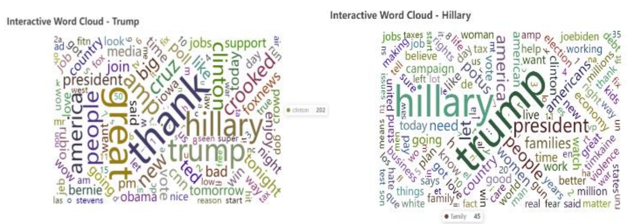


Figure 2

Analysing the tweets of the two candidates in conjunction with bar charts and word clouds (Figure2) provides a richer perspective and information and has significant advantages in relation to the key topics. The bar chart focuses more on the ranking and frequency of the words used by the two candidates, which can help to analyse the different frequency of use of the same word by the two candidates. However, since the candidates use a wide range of vocabulary, the bar chart is not comprehensive enough. On the other hand, the word cloud can provide relative font sizes and colours according to word frequency, which can more intuitively show the main keywords used by different candidates in their tweets. Combining the two visualisations can thus better reveal which words are most important in the candidates' tweets, and interactively understand the specific value of each word to further identify the topics discussed by the candidates.

The available topic groupings for each candidate are first discussed separately, through the high-frequency word cloud:

Trump's tweets can be categorized into several key topics:

1. campaign & slogan. 2. Attacking the opponent

3. National Affairs. 4. Foreign Policy 5. Economy

For Hillary:

1. Campaigns and teams. 2. Mentioned by opponents

3. Politics and social justice. 4. Call and vision

5. State and values

To facilitate the subsequent analysis of sentiment and key moments, it was decided to identify a set of themes (i.e. a set of issues common to the two candidates) by clustering the two candidates' themes using the SOM(Figure 3).

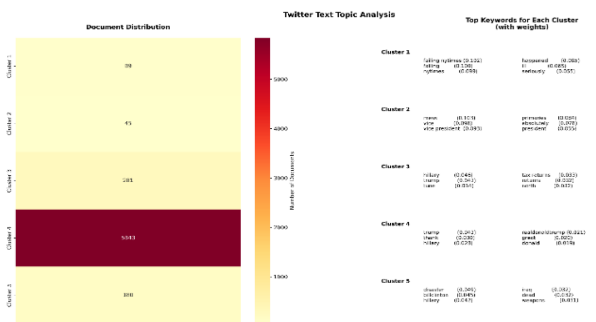


Figure 3

Firstly, the two candidates were observed through high-frequency words and the range of their possible topics was set at five. Then, all high-frequency words were clustered by SOM and combined with the possible topics previously observed by high-frequency words to determine the final topic. The left side of the above visualisation shows the number of documents for each cluster, with the distribution of the number of documents for the different clusters indicated by the colour bar in the middle; the darker the colour, the higher the number of relevant documents, which can help the viewer to see the relative size of each topic (as only the topics are identified in this section, the detailed trends will be explained and analysed later). On the right-hand side, the high-frequency keywords for each cluster and their corresponding weights are listed.

Combining the left and right areas shows that Cluster4 has the highest number of documents at 5843. Combining the corresponding keywords on the right, the highly weighted keywords of Cluster4 contain terms such as ‘trump’, ‘thank’ and Hillary' such terms, which in combination with the previous possible topics leads to the conclusion that the topic is based on vote mobilisation and slogans. A further look at the keywords in the other clusters, for example in Cluster1, where the keywords ‘failing’, ‘nytimes’, seriously' reveals that the theme is about candidates contesting the election result.Cluster2, which includes terms such as “mess”, “vice-president” and other words that refer to the negative evaluation of an event or person.Cluster3, which includes “tax return”, “tune ”, “north” and other words that refer to the discussion of tax plans. Finally, cluster 5, which contains words such as ‘disaster’, ‘billclinton’, ‘hillary’, can refer to socially relevant topics.

The final topics were identified as follows:

1. ‘Trump vs. Hillary Campaigning’
2. ‘Campaign Criticism & Tax Plan’
3. ‘Presidential Discussion & Beliefs’
4. “Thank You & Campaign Slogan”
5. ‘Debates & family issues’

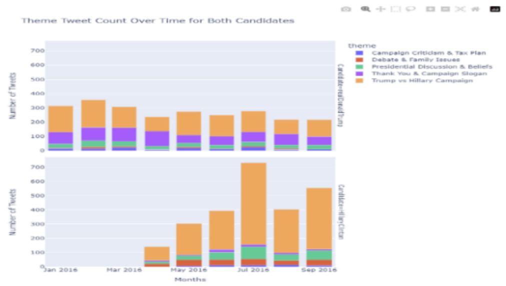


Figure 4

Observe how the topics of each candidate have developed over time by creating a visual stacked bar chart (Figure4). A visual stacked chart can help to show the trend of the candidates' discussion of different topics more intuitively. By displaying the months on the x-axis in this visualisation, it is possible to visualise the change in the discussion topics of the tweets over time, which helps the viewer to quickly observe the development of the candidates on the topics and thus deduce possible events such as party nominations and debates [1].

The number of tweets is displayed on the Y-axis, enabling the user to quantify the level of discussion on each topic at any given time. The interactive visualisation also allows the user to take a closer look at the volume of each topic in a given month by hovering over a specific topic with the mouse. Furthermore, the viewer can select sections of interest for detailed viewing by clicking them in a For the visual variables, each topic is assigned different colours and the topics represented by each colour are indicated by a legend in the upper right corner, allowing viewers to quickly distinguish between different topics and see the changing trends of different topics.

As for the main topics for each candidate, the election 'Campaign Trump vs. Hillary' is the dominant topic for both candidates. For Trump, the topic dominated all topics (tweets) from January to September 2016, with the exception of April, a relatively even overall trend. Next is Thanks & Campaign Slogan (Purple), which is the second most popular topic discussed by Trump, and it can be seen that from January, the number of posts Trump makes about & with this topic steadily increases until it peaks in April, possibly due to the addition of Hillary, which has shifted the focus of his tweets. From there, we see that the number gradually decreases in a wave-like trend from the peak in April. For 'Campaign Criticism & Tax Plan', which represents Trump's attacks on his opponents as well as his tax plan, we see that the trend for this topic fluctuates with the trend of the total number of tweets. For the other topics, the distribution is sparser. This suggests that Trump is focusing more on himself to build his image in people's minds, while also denigrating his opponents a little.

For Hillary, the number of tweets on the topic 'Trump vs. Hillary Campaign' has gradually increased since Hillary started tweeting, especially in July and September 2016, when the number of tweets on this topic increased significantly. This could indicate increased attention on the campaign during this period and imply that events about the campaign were taking place during this period. Secondly, 'Presidential Discussion & Beliefs' is the second most popular topic Hillary focuses on, following a similar trend to the campaign topic. It can be concluded that when she posts a large number of tweets on the campaign topic, she is simultaneously promoting herself and thus increasing her campaign advantage by demonstrating her willingness and ability to do so. In addition, Hillary also focuses on the topic of ‘Debate & Family Issues’, a topic that is trending more steadily.

Figure 5

The number of retweets and likes for each candidate's posted threads were added up to get a total engagement score, and then the average was calculated. Since the chart needs to include information from both candidates, we consider an reversed bar chart (Figure5) for presentation.

Setting the topic as the x-axis allows viewer to compare the engagement between topics. Setting the average number of interactions on the y-axis allows viewers to better compare how much attention the two candidates are paying to tweets on the same topic through engagement. Different colours have been assigned to the different topics to make comparisons between topics clearer. Bar charts can effectively show the comparison between values, especially when quantifying this type of data with multiple categories.

The bar contains three labels, namely average retweets, average likes and the number of tweets for the topic. The chart also features an interactive design that displays further information on engagement when the viewer moves the cursor over the topic. This helps the viewer to make a quick judgement and get more detailed information when looking at different topics and candidates. Viewers also have the option to hide topics or anything that is of interest to the viewer, can also keep the topics they are interested in using the legend in the upper right corner, with the filter function working for both candidates at the same time.

Looking at the various themes, Hillary posted a total of 51 tweets on ‘Campaign Criticism & Tax Plan’, with an average engagement of approximately 8216. Trump posted a total of 121 tweets on this theme, with an average engagement of approximately 26k. It can be seen that not only did Trump post a higher number of tweets on this topic than Hillary, but his tweets also had a much higher level of engagement than Hillary's. This suggests that Trump's unique campaign style attracted more attention than Hillary's attacks on his opponent's policies or tax plans [2]. For ‘Debate & Family Issues’, Hillary had 219 tweets, while Trump had 53 tweets, the average engagement for both was 9519 and 26k respectively. Hillary posted far more tweets than Trump on this topic, but Trump's average engagement is still higher than Hillary's, which shows that Trump is more able to attract public attention and interact when it comes to these types of topics.

Furthermore, the topic with the largest difference in engagement between the two candidates was ‘Presidential Discussion & Beliefs.’ Hillary posted a total of 292 tweets with an average engagement of about 10k. Trump posted a total of 270 tweets on the subject, with an average engagement of about 29k. Trump used a more ‘authentic’ voice to present himself, which allowed the public to engage more with the campaign and thus gain more attention [2]. For the other two topics, it was about the same as before.

Overall, Trump's average engagement on each topic was much higher than Hillary's. This suggests that Trump built his brand on social media through his unique campaign style, which has led to much more online attention for Trump's candidacy than for Clinton's [2].

II. TASK 2

A. Processing of data

In the data preparation stage, the tweets were first cleaned, including deleting stop words, removing noisy information, restoring stemming, and removing punctuation to improve the accuracy of sentiment and tone analysis. The removal of deactivated words (e.g., "the" and "is") allows the model to focus more on meaningful words, while noisy information (e.g., hyperlinks, emoticons) avoids interfering with sentiment analysis. Through stem reduction and punctuation removal, the normality of the data is further improved, laying a good foundation for sentiment classification and tone analysis.

Next, the NLP technique was used to construct the sentiment lexicon and perform sentiment analysis. Sentiment scores are computed using the VADER and SentiWordNet lexicons, where 0 denotes neutral sentiment, negative denotes negative sentiment, positive denotes positive sentiment, and the larger the absolute value, the stronger the sentiment [3]. The VADER lexicon is particularly effective for analyzing short social media texts [4], while SentiWordNet covers a broader vocabulary base [5]. In the context of election campaigns, we adapt the lexicon, e.g., by assigning substantial negative weights to words such as "fake" or "crooked," to capture the emotions expressed in candidates' tweets more accurately.

Once the sentiment classification is complete, the first step is to identify general patterns by observing the tweets. Then, the identification and classification of the tone of the tweets using the regularization method allows us to further understand the candidates' expression strategies from a tone perspective. This combined analysis provides a solid foundation for the multidimensional mood, tone, and theme interpretation.

B. Emotional categorization and thematic speculation

In the Sentiment Classification and Theme Speculation section, we further analyzed the two candidates' emotional characteristics on different topics and their emotional changes at critical stages of the campaign process by plotting the average sentiment distribution (Figure 6) and the change of sentiment over time (Figures 7 and 8). These charts clearly show the emotions the candidates tend to express on different topics and reveal how they adjust their emotional expressions at essential moments to realize their campaign strategies better [6].

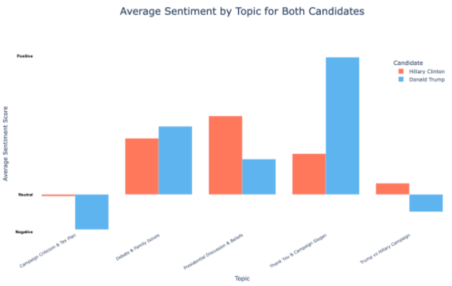


Figure 6

In Figure 6, we can see the differences in the emotions of the two candidates on various topics. Red represents Hillary, and blue represents Trump, the horizontal coordinates are the five high-frequency themes, and the vertical axis indicates the type of emotion. Users can hover to see the detailed scores. For example, in the theme of "Campaign Criticism and Tax Plan," Trump's sentiment score is negative (-0.1), which indicates that he criticizes Hillary's policies more, while Hillary's sentiment is close to neutral (-0.01) on this theme, showing a more restrained attitude, Similarly, in the theme of "thanks and slogans," Trump shows stronger positive emotions (0.41), indicating that he boosted his support by thanking his supporters and promoting his slogans, while Hillary is more conservative, with a score of only 0.12. These differences in emotions reflect the different emotional tactics of the two candidates in their campaigns: Trump's emotional displays were more intense, with greater emotional fluctuations, especially when criticizing his opponents and boosting his own image, while Hillary's emotions were more neutral overall, as she maintained restraint and presented an image of stability and reliability that appealed to voters who were not interested in overly intense emotional displays [7].

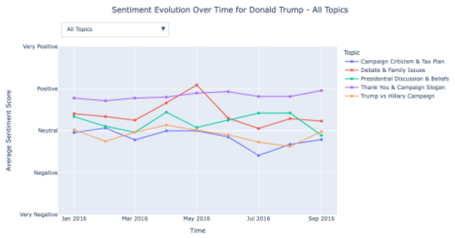


Figure 7

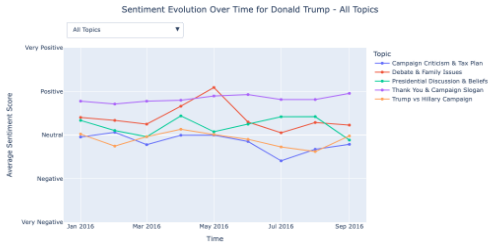


Figure 8

Figures 7 and 8 illustrate the dynamics of the two candidates' moods during the campaign, especially before and after critical events. The vertical axis represents the sentiment score, the horizontal axis represents time, and the different colored lines represent different themes. The interaction box allows the user to select a specific theme to view the change in sentiment and hover to view specific sentiment scores. For example, during the mid-campaign period from May to June, sentiment for the "Debate and Family Issues" theme rose significantly, indicating that candidates were using this issue to present a positive image to attract voters. At the height of the campaign (August-September), Figure 7 shows Trump's rising sentiment on the topic of "Thank yous and campaign slogans," while at the same time, his sentiment on "Campaign criticisms and campaign slogans" is rising. While maintaining negative sentiment on "Campaign Criticism and Tax Plan," this combination shows him consolidating support while appealing to voters who were skeptical of Hillary. In contrast, Hillary showed poise and confidence by maintaining a steady positive sentiment on "Debates and Family Issues" and "Presidential Discussions and Faith." [8]

Overall, the moods of the two candidates were relatively stable in the early part of the campaign, with mood swings increasing as the campaign progressed, especially during major events such as debates and policy announcements. Trump's mood changes were more pronounced on several themes, actively guiding voters' emotions through emotional mobilization, while Hillary maintained a steady positive mood, conveying an image of reliability.

C. Tone of voice strategy

Figure 9

In order to more visually analyze the tone strategies of the two candidates during the campaign period, we drew a tone distribution chart (Figure 9) to show the change in their tone at different stages. The horizontal axis of the chart indicates the time. In contrast, the vertical axis shows the proportion of each type of tone, with different colors representing the five-tone categories, and users can hover to see the specific tone in each period.

The overall trend shows that a "supportive tone" occupies a significant proportion of both candidates' tone strategies. Hillary's supportive tone showed a high degree of consistency throughout the campaign, with the explicit intention of strengthening the voter base and conveying the image of a united campaign through positive expressions. On the other hand, Trump's supportive tone was dominant in the early part of the campaign but gradually declined, suggesting that he reinforced his stance more through confrontation later and relied less on supportive expressions [9].

The difference between the two candidates is particularly striking regarding "offensive tone." Trump's offensive tone remained high throughout the campaign. It increased in the later stages, reflecting his campaign style that emphasized criticizing his opponents and challenging the system in order to appeal to voters who were dissatisfied with the current system. On the other hand, Hillary was relatively restrained in using an offensive tone, with a low and stable percentage, showing her more cautious strategy of criticizing, intending to portray the image of a trustworthy candidate through a rational and positive tone [7].

In addition, other types of tone, such as 'confident,' 'hesitant,' and 'neutral,' were relatively small in proportion between the two candidates, and the differences were minimal, suggesting that they were mainly focused on communicating precise positions rather than emphasizing vague or overconfident expressions. These lower percentages further reveal differences in their strategies for campaign language, with Hillary's remaining rational and supportive and Trump drawing attention to himself through a higher percentage of aggressive tone, creating two distinct campaign styles.

D. *Correlation analysis of mood and theme*

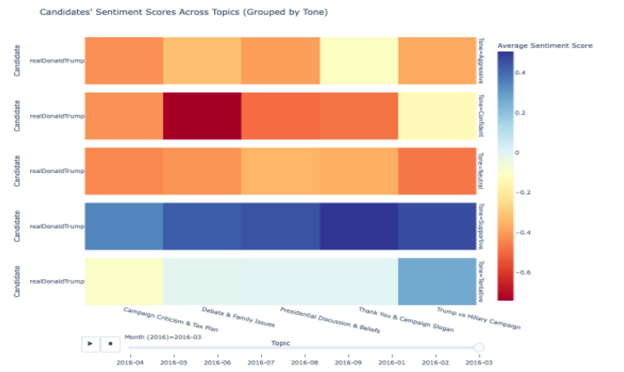


Figure 10

Ultimately, by mapping out the integration of emotion, topic, and tone, we can explore their emotional tendencies and tone strategies across different topics in greater depth. This integrated display allows us to visualize the changes in candidates' expressions and emotional strategies on each key topic. The horizontal axis of Figure 10 shows the five high-frequency campaign topics, the vertical axis represents the candidates and their tone classifications, and the color intensity indicates the positive and negative sentiment scores. The darker the color, the stronger the sentiment. Users can select a specific period through the timeline below the graph to carefully observe the changes in sentiment and tone over time and also hover over the graph to view detailed data information, such as the themes, candidates, and tone categories at a specific point in time, as well as the corresponding average sentiment scores.

Figure 10 shows significant differences between the candidates' sentiment and tone choices on specific topics. For example, on the theme of "Campaign Criticism and Tax Plan," Trump's sentiment scores were significantly negative, accompanied by a high percentage of aggressive tone, reflecting his intense style of criticizing his opponents and reinforcing his positions. This strategy allows him to use intense emotional expression to direct public attention. On the other hand, Hillary mainly used a supportive or neutral tone and was more emotionally neutral, showing a more restrained attitude and attempting to portray a rational and trustworthy candidate image through moderate expressions. This choice of mood and tone not only reflects the different campaign styles of the two candidates but also reveals how they precisely adjusted their expression strategies on specific topics to suit different groups of voters [7].

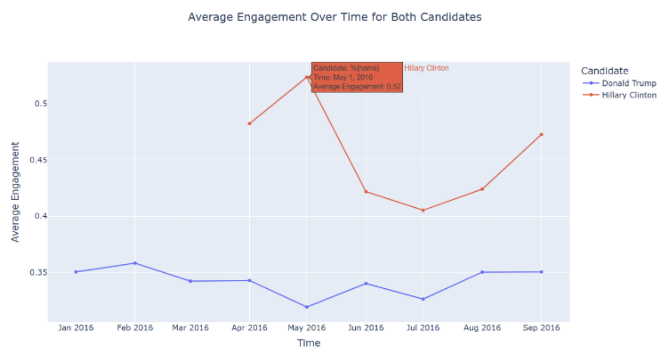
Overall, therefore, the two candidates' mood and tone strategies on crucial campaign issues show their characteristics: Hillary tends to consolidate her voter base and maintain the continuity and consistency of the campaign process through a stable and supportive tone and positive emotions, while Trump shows more significant mood swings across multiple topics, and is more aggressive especially when it comes to his opponents' criticisms. This significant difference not only reveals the campaign styles of the two candidates but also shows how they flexibly adjusted their strategies on different topics through their moods and tones to steer public opinion and voter sentiment at critical junctures effectively [10].

III. Task 3

In this task, we use the result from previous tasks as key attributes to help us create visualizations needed. The first objective of the task is to identify a critical moment in a visualization where a candidate could have tweeted differently. The expected outcome of the objective is to bring insights into the analysis of completion of Hillary Clinton and Donald Trump in the Twitter (now X) platform. User should be able to clearly identify the critical moment where the battle between the two 2016 US president candidates took a turn as they both tweeted something that lead to this moment.

To define our own terms of critical moment, there are few things that we considered. First, as any moment, it should be a temporal concept and therefore, the analytic of critical moment should be a temporal analysis based on time attributes or data created in previous tasks or provided in the source data. Second, to consider a moment critical, it means that this moment should have great effect on something, and in our consideration, due to the uniqueness and singularity indicated by this moment, the maximum effect of something is the key indicator of a critical moment. Combine this thought with the background and real word meaning of the data given, a critical moment should mean a point of time when a biggest effect is happening which impact the most on the battle between Hilary Clinton and Donald Trump on Twitter. Third, followed by the above design of critical moment to bound the term more to a reality meaning, the "biggest effect" in clear definition is the largest difference of some measurement of effect of tweets by the two candidates.

After the key concepts as well as the core value of the visualization is defined, we carry on creating the visualization. Below is the interactive visualization created for the critical moment. Later, we will justify our design of the visualization and analyse the visualization to get the critical moment.



Overall, the visualization is designed to be a plot graph as it could clearly show trends and most importantly, show differences of values for each segment of time. And considered the simplicity of the elements involved in this visualization, it should be the optimal practise of temporal analysis as required.

To further explain our visualization, axes arrangement is the first vital aspect. Derived from the concept of critical moment, two attributes are essential and needed: time and measurement of effect. By using the works built in previous tasks, time attribute we used here is presented by periods of months in month-year format. And the measurement is the average engagement of tweets within each period. In this way, the critical moment can represent and reflect on two candidates greatest and most important tweets engagement difference that affect their campaigns later. An orthogonal construction is adopted based on the number of attributes and linearly independent and uncorrelated between them. Naturally, Time is set to be the x-axis and Average Engagement set to be the y-axis with their names as labels. The scale of x-axis is the maximum range of periods given with ticks for each month to provide more details on time itself for better analysis. For y-axis, the range is determined by the highest and lowest value of Average Engagement possible. According to that range and value distribution, it is set to be between 0.3 and 0.55 with 0.05 as scale of the ticks for visual comfort. The overall axes arrangement enhances data readability and helps user interpret differences of values in the data more accurately.

For visual variables, colour is one main subject. The background of the visualization is using a light theme of colour combination (white and light blue) to give a good contrast to information with values and show details of values with nice visibility. Then, the two plots of each candidate are visualized in blue and orange colours respectively for good differentiation. The overall use of colour is ensured to be colour-blind safe for accessibility. Other than colour, legend is shown in the top right corner with text information to indicate the belongings of each plot. A text is also shown at the top of the diagram to express the direct meaning of the visualization. The visual variables designed here enhances the comprehensibility of data, highlighting the trends and differences of values, making user easier to decide and understand choice of critical moment.

One interaction in this visualization is set to be highlighting. As all key information can be shown within this visualization, there is no need to interaction like filter, graph manipulation here. Only more detailed information on each node of the plot is needed as it could provide exact value of average engagement. The plot can provide visually perceptual information on difference for user to quickly identify the critical moment while the highlighting action on each node can give user a more evidence-supported idea of the scale of difference, contribute to the profonde agreement of why this moment is critical in the analytic processes. User is also allowed to zoom in and out as well as navigate around the visualization for better exploration.

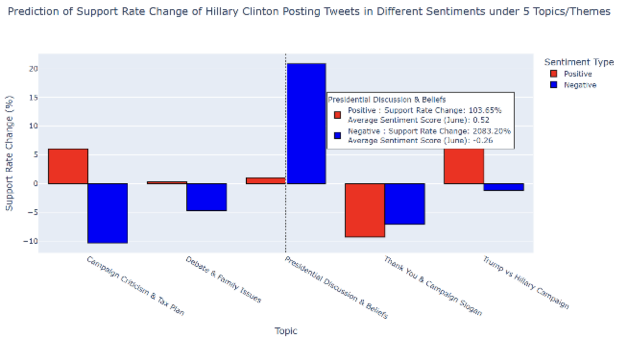
The critical moment derived from the visualization is May 2016 where the difference of tweets engagement between Hilary Cliton and Donald Trump is maximum. In this moment, Hilary and Donald Trump both posted tweets that lead to the downfall of Hilary Cliton and uprising of Donald Trump. The critical moment represents the most significant reverse of this duel and provides background of the alternative tweet discussed in the following section.

The second objective of this task is to suggest an alternative tweets and its prediction of change of response from the opponent or the public using visualization. The expected outcome of the visualization is that the user should be able to derive an alternative tweet that might change the course of this fight between Hilary Cliton and Donald Trump on Twitter by using the visualization. Furthermore, the visualization should show clear prediction of effect by Hilary Cliton / Donald Trump after posting that presumed tweet. The visualization would provide an in-depth analysis on the factors behind the trend, and give interesting insights into this famous, world-affecting US election duel.

One key concept here is the alternative tweet. Concretization of this concept is what we focus here. The first thing to decide is the serving object of the alternative tweet. To keep the visualization grounded to reality meaning without spreading focus too widely, one candidate was select as the serving object of the alternative tweet. By viewing changes of average engagement of critical moment, it is obvious that Hilary Cliton gained a greater change than Donald Trump. To maximize the effect of alternative tweet and gain a visually more distinctive graph, Hilary Cliton was selected as our serving object here. The content of the alternative tweet is the next focus. What to tweet is the question. Constrained by the size of the dataset and approach taken on this task, generate an alternative tweet with actual text contents with Machine learning or AI is undesired and a deviation from the purpose. Therefore, with the power of visualization and the information gained from previous tasks, the way we adopted is to proposed directions of the alternative tweet. More specifically, providing suggestion of sentiment type for each topic / theme. What needs to be mentioned is that since the different of tone engagement in task B existed with small differences of sentiment score, we picked sentiment type (positive and negative) over tone and emotional engagement for overall visual significancy.

Another key concept to be considered is the measurement. In our solution, supporting rate change is the new concept created and used as a measure of effect of alternative tweet. Support rate change derived from engagement score in previous tasks, it is the combined calculation of two processed engagement score (for each candidate) with proper scaling and transformed in percentage. It indicates the current change of percentage of support a candidate gained. With this measurement, the effect of alternative tweet can be quantified.

In conclusion, our alternative tweet is a suggestion of what sentiment and topic / theme should have been taken for Hillary Cliton in her critical moment to gain a positive outcome of support rate change. The visualization shows all possible outcome of every combination of sentiment type and themes to discover the best one as suggestion of alternative tweet. Visualization is shown below, and justification of our choices of features will be explicitly elaborated afterwards.



The visualization is a stacked bar chart. Bar chart is selected to showcase quantitative data support rate change and present data in a clear and simple way. And to display all possible outcome for prediction, both types of sentiment is considered, and therefore it is a stacked bar chart.

In detail, topic is the x-axis and support rate change are the y-axis since vertical layout is more accepted in general cases. For x-axis, all five different topics (themes) are displayed and arranged. This allows direct comparison between best possible outcomes for each topic. And for each topic, it is grouped with its corresponding sentiment types. This enables user to pick sentiment types under a topic. Each topic is labelled alongside with x-axis, the text is arranged in tilt to ensure readability. For y-axis, the range is decided by the minimum and maximum value of support rate change with -10 to 20 % marked and 5% ticks deployed for quick assessment of change degree. The value of support rate change is determined by the data of average support rate change of the candidate using the same sentiment and topic in all time periods, with proper normalization procedure taken to minimize the effect of different quantities of every type of tweets. Orthogonal construction is adopted here for usage simplicity and content focus under two axes design.

In the case of visual variables, the main colour is red and blue with light colour used in background. This use of colour ensures the contrast is easily spotted, distinct different type of sentiment from each other. For position, each pair of positive and negative bars is placed adjacent to each other contribute to a direct comparison. Size of the bar is arranged with proper value so that the overall view of the visualization stays decent and clear. Text title is shown expressing the core characteristic of the visualization. And lastly, legend is placed on the top right corner for user readability and differentiation between two types of sentiment bar.

Interaction is set to be highlight as user will gain detailed information on sentiment score, support rate change for each type of sentiment when user hover on corresponding part. This feature enhances analytical insight by drawing attention to the most impactful topic without detracting from the overall chart structure. Since the visualization is considered straightforward and convey information effectively, there is no need for further selection and manipulation interactions. However, user is still allowed to navigate freely through the visualization as part of the interaction.

From the visualization, it can be derived that presidential discussion & beliefs is the overall most impactful topic as it has the highest support rate change while using both sentiment types can lead to a positive respond. Furthermore, negative sentiment used here is indicating a growth of support rate by 20%. In conclusion, an alternative tweet using negative sentiment on the topic about presidential discussion & beliefs would be the most desired choice for Hilary Cliton. The real-world meaning of Hilary Cliton posting such tweet in May 2016 that attacks on, disagree on Donald Trump's discussion and beliefs of presidency could lead to significant winning over public favour and maintain advantage.

IV. Task 4

NPL (Post-SOM) diagrams in Task 1 is the example we selected to showcasing our evaluation processes.

Method and justification: Use of questionnaires as an evaluation method.

The use of questionnaires as an evaluation method helps us to get feedback on visualisations and interactions from the audience's point of view. The ability to collect qualitative (i.e. quantitative) data from user feedback will help us to evaluate the user experience more intuitively in terms of interactivity, information delivery and visual design. In addition, different users have different feelings about visualisation, so using a questionnaire for evaluation can cover all types of user opinions and collect feedback from different perspectives.

Questionnaires provide a structured way of data collection where users rate the visualisations based on predetermined questions, thus capturing direct feedback from users on such questions with a defined scope, which is more suitable for quantitative analysis.

Design and structure of the questionnaire: A variety of question types were designed to ensure a comprehensive and diverse evaluation.

1) Open-ended: 'Do you think that the diagrams directly reflect the distribution of the various themes?', 'If so, how do you think the themes can be roughly categorised?'

2) Scale: 'Please rate the readability of the diagram (on a scale of 1-5)'

3) Closed: 'Do you think the information in the tables and diagrams can help you find a topic?' a) Very Easy b) Easy c) Average d) Difficult e) Very Difficult

Improved visualisation:

Initial selection: In the initial stage, we used an NLP approach for topic selection of tweets to identify the main topics by automatically analysing the tweets. However, we found that the method did not distinguish some topics very clearly, resulting in roughly the same keywords appearing in two different topics with only minor differences, and that the method was unable to handle noisy data.

Shortcomings: In addition, users were asked to provide feedback using a questionnaire after the presentation of the visualisation. This showed that users were satisfied with the colour coding and interactive features, but most users found it difficult to distinguish directly between topics and the information in the diagrams was somewhat misleading.

Final solution and improvements:

Considering the limitations of NLP methods, we considered several approaches and finally opted for the SOM technique to group themes using a clustering algorithm able to find a clearer distribution of themes in a multidimensional space.

Feedback on the questionnaire and improvement:

Based on user feedback, it was found that users had made some comments on the colour coding, so the colour coding was fine-tuned and supported by the addition of colour bars. In addition, users suggested that the information in the diagrams should be more appropriate and therefore the interactive features could be removed to ensure attention.

References

[1] BBC News. (2016, January 25). US election 2016: All you need to know. BBC News. https://www.bbc.com/news/world-us-canada-35356941

[2] Zompetti, J. (2019, January 1). Rhetorical incivility in the twittersphere: A comparative thematic analysis of clinton and trump's tweets... Unknown. https://www.researchgate.net/publication/338189571_Rhetorical_Incivility_in_the_Twittersphere_A_Comparative_Thematic_Analysis_of_Clinton_and_Trump's_Tweets_During_and_After_the_2016_Presidential_Election

[3] Z. Drus and H. Khalid, "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review," Procedia Computer Science, vol. 161, pp. 707-714, Jan. 2020. Available: https://doi.org/10.1016/j.procs.2019.11.174

[4] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in Proceedings of the International Conference on Web and Social Media (ICWSM), vol. 8, no. 1, pp. 216-225, May 2014. Available: https://doi.org/10.1609/icwsm.v8i1.14550

[5] A. Esuli and F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," in Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC), Genoa, Italy, May 2006.

[6] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in Proceedings of the International Conference on Web and Social Media (ICWSM), vol. 8, no. 1, pp. 216-225, May 2014. Available: https://doi.org/10.1609/icwsm.v8i1.14550.

[7] M. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment," in Proceedings of the Fourth International Conference on Weblogs and Social Media, Washington, DC, USA, 2010.

[8] J. Grimmer and B. M. Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," Political Analysis, vol. 21, no. 3, pp. 267-297, Jan. 2013. Available: https://doi.org/10.1093/pan/mps028

[9] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008. doi: https://doi.org/10.1561/1500000011

[10] Z. Drus and H. Khalid, "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review," Procedia Computer Science, vol. 161, pp. 707-714, Jan. 2020. Available: https://doi.org/10.1016/j.procs.2019.11.174.

How Would you Visualise your data?

COMP5048/4448 Assignment 2

line 1: Weihao Li
line 2: *500021893*
line 4: CC-08-G06
line 5: Katherine Mu

I. Task 1

Active participation in the preparation of the project. At the beginning of the preparation phase of the project, he discusses with the group members how the various tasks should be distributed and expresses his own opinion, responds positively to the team members' communications and organizes the flow of the meeting. Before the group set a date for the meeting, I understood and explored the information of the assigned part in advance, so that I could express my own opinions and insights during the meeting, and I could quickly respond to the doubts of the group members, patiently listen to the opinions of my teammates and respond to them when necessary, which ensured the efficiency of the meeting. I was responsible for the first task and evaluation part of the project, and since the first task was the key point of the whole project, I tried to confirm the theme as soon as possible. After confirming the theme, I received an invitation to a meeting from my teammates, and during the meeting I received feedback from my teammates about changes that needed to be made to the theme, which helped to clear up some of my ambiguities. After the meeting, the theme was first changed and the first task was completed as much as possible to make it easier for the following team members to proceed

The pre-processing of the dataset was done by the group members who filtered the attributes of this dataset and eliminated the unwanted attributes. For task 1, I further pre-processed the dataset by filtering the attribute 'text' using the spacy library to remove the disabled words in the tweets and further filtered it by looking at the word frequency to remove special words like 'http ", "-' etc. According to the assignment, the group member was supposed to do some of the programming tasks. However, I felt that his task was more involved, so I took the initiative and wrote the code required for my task myself and informed the group member about it.

Also, before writing about the theme area, I read many summaries of reports containing information about Trump's and Hillary's campaigns, which included a lot of relevant content about the issues and tweets, such as, 'Rhetorical Incivility in the Twittersphere: A Comparative Thematic Analysis of Clinton and Trump's Tweets During and After the 2016 Presidential Election'[1] , which was able to deepen my understanding of assignment one. When I finished my section, I adjusted the formatting to meet the requirements of assignment 2 and added the cover sheets to the document to make it easier for the group to complete. When working on task 1, inform the group that part 1 is a little too long and let them know that I will shorten it if necessary to ensure that the overall length of the report does not exceed the requirements.

II. Task 2

A. Determining/Designing Visualization Methods

The first step was to understand the task and the dataset in its entirety. After understanding what the dataset contained, it was determined that the attributes that needed to be used for Task 1 were 'handle', 'text', retweet_count' and "favourite_count". As the requirement of Task 1 is to identify the candidate's theme, it was considered that frequent words could be extracted from the candidate's tweets and displayed visually using an interactive word cloud and frequency graph. The word cloud can show the audience the most frequent tags and keywords in the two candidates' tweets at a glance, and the word frequency is reflected in the size of the word, which can attract the audience's attention. The different high-frequency words are then clustered by combining them with the SOM to help identify theme.

In addition, it was considered important to show the evolution of the themes over time as well as the level of engagement with the different topics. It was decided to represent this through a bar chart divided into upper and lower sections to differentiate between the candidates while helping the audience to better understand how the issues publicised by the different candidates have changed over time. An inverted bar chart showing engagement on different themes helps the viewer to observe and compare the engagement on the themes of the different candidates in the first place.

B. Determining/Designing Visualization

1. Wordcloud(for main word):

For the word cloud of tweet keywords, the 'text' attribute is first extracted from the data set, and the word cloud is created by summarising all words in the tweets into a flat list, which is then subjected to a word frequency calculation to determine the most frequent words.

2. Frequency bar graphs [2]:

Arrangement of axes: The x-axis represents the frequency of the individual keywords, the y-axis the keyword itself. This arrangement allows the user to visually understand the frequency of each word and makes it easier to compare different word frequencies.

Visual variable selection [3]: A bar chart with different colors was used for the two candidates. Since there are fewer variables in the bar charts and they are only used for frequency comparison and viewing, there is no multi-colored differentiation in the same chart to avoid too many colors that could mislead the viewer.

3. Segmented bar charts (Theme Tweet Count Over Time for Both Candidates) [2]:

Arrangement of axes: the x-axis represents the month, allowing the viewer to clearly observe the evolution of each theme over time, and the y-axis represents the number of tweets posted on the theme, helping the user to quantify the level of discussion on each theme in a given time period.

Choice of visual variables [3]: A faceted bar chart distinguishes the two candidates and allows a more intuitive comparison of the evolution of the number of tweets posted by the two candidates on the different themes in the different months. This avoids the confusion of overlapping information and improves readability and contrast. Different colours are assigned to each theme and the legend in the top right corner shows the theme represented by the respective colour, so that the viewer can quickly distinguish between the different themes and assess the trend of the different themes.

4. Inverted bar charts (Average Engagement per Theme for Each Candidate) [2]

Arrangement of axes: The x-axis is used to represent the different themes and provide a visual comparison of engagement between themes. The y-axis is set to the average number of interactions and is designed to help viewers quickly quantify engagement by comparing the difference in engagement between two candidates on the same theme.

Choice of visual variables [3]: The use of a bar chart that tilts up and down to show the data for the different candidates separately ensures that viewers can easily compare the engagement of the different candidates on the same theme. Different colours have been assigned to different themes to facilitate comparison between themes. The use of the y-axis as a quantitative criterion allows the viewer to quickly determine the level of engagement for different themes based on the height of the columns. In addition, labels are added to the columns to make the engagement information even clearer, which helps the viewer to make an initial judgement when looking at different themes and candidates and helps them to get more detailed information.

C. Determining/Designing user interaction

To enable viewers to get more effective information from the graphs, I added features such as hover interaction and tooltips to the visualisation graphs. When the viewer hovers over the relevant data, information about the corresponding data is displayed, such as the number of occurrences of the word in the interactive word cloud, the number of tweets or the topic they belong to in the bar chart, and the detailed values of each item in the engagement chart. In addition, there are box and filter functions, and in items 4 and 5, viewers can exclude topics they do not want to see by clicking on the filter in the top right corner. It is also possible to use the mouse to highlight the data you want to watch. [4]

D. Determining/Designing the evaluation process

Set an evaluation objective: Ensure that the visualisation effectively communicates thematic trends over time and trends in engagement. Ensure that colour variables are used appropriately, labels and axes are designed appropriately and interactions produce more detailed information [5].

Start with a comprehensive self-assessment by evaluating and reviewing each visualisation:

Labels and axes: all visualisations in the diagram are appropriately arranged and contain clear labels and titles that provide useful information to the viewer.

Colour coding: The use of different colours to represent topics is effective for differentiation.

Interactivity: Interactive features are present and detailed information is displayed when hovering over the columns. They do not interfere with the other visualisations.

For further evaluation, the project requirements were presented to a friend to operate the visualisation charts alone, which resulted in the following feedback [5]:

Messaging: 1. Ability to quickly identify and compare high frequency words through word clouds and frequency bar charts; 2. Ability to clearly understand the percentage of different topics in each month through compartmentalised bar charts, and how it changes over time with topics; 3. Ability to see at a glance how engaged the public is with different candidates' tweets on different topics;

Interactivity: able to provide appropriate information

Layout: a good visual layout and design that makes it easy to recognise at a glance what is being said and communicated.

E. Conducting and analysing the evaluation

Viewer feedback on the colours and layout was generally positive, as they felt that the different colours made it easy to distinguish between topics and that this faceted presentation reduced visual distractions, making comparisons easier.

Improvements and changes: In the initial stages, it was found that the spacing between the reversed bar charts was too small, which could make it difficult for viewers to distinguish between the two candidates, so it was increased. More labels were also added to give users more useful information before interacting.

REFERENCES

[1] Zompetti, J. (2019, January 1). Rhetorical incivility in the twittersphere: A comparative thematic analysis of clinton and trump’s tweets... Unknown. https://www.researchgate.net/publication/338189571_Rhetorical_Incivility_in_the_Twittersphere_A_Comparative_Thematic_Analysis_of_Clinton_and_Trump’s_Tweets_During_and_After_the_2016_Presidential_Election

[2] M. Takatsuka, "COMP5048 Visual Analytics, Week 08: Design and Representation," School of Computer science, University of Sydney, 2024.[Lecture Notes]

[3] M. Takatsuka, "COMP5048 Visual Analytics, Week 02: Color," School of Computer science, University of Sydney, 2024.[Lecture Notes]

[4] M. Takatsuka, "COMP5048 Visual Analytics, Week 04: Fundamentals of Human Computer Interaction(HCI)," School of Computer science, University of Sydney, 2024.[Lecture Notes]

[5] M. Takatsuka, "COMP5048 Visual Analytics, Week 06: Evaluation," School of Computer science, University of Sydney, 2024.[Lecture Notes]

How would you visualise your data?

COMP5048/4448 Assignment 2

line 1: Yuhan Wang
line 2: *510031279*
line 4: CC-08-G06
line 5: Katherine Mu

I. Task 1

A. Project preparation and Idea Development

In the initial stages of the project, I took a leadership role in conceptualizing and planning our approach for all three tasks. Before our first team meeting, I thoroughly reviewed the assignment requirements and analyzed the dataset to identify key opportunities for visualization.[1]

1. Thematic Analysis and Engagement (Task 1):

I proposed a comprehensive approach for identifying and visualizing the different themes of discussion between the candidates. Recognizing that understanding the main topics emphasized by each candidate was crucial, I suggested a multi-faceted method:

Clustering with Self-Organizing Maps (SOM):

Understanding that high-frequency words alone might not reveal underlying themes, I proposed using SOM clustering to group related words.() This method would help us uncover hidden patterns and more nuanced themes within the candidates' tweets.[2]

Defining Common Themes: Based on the clustering results, I led the effort to interpret the clusters and define five key themes common to both candidates. I ensured these themes were relevant to the context of the 2016 U.S. Presidential Election and reflective of significant public discourse topics.[7]

By designing this approach, I laid a solid foundation for our thematic analysis, enabling us to delve deeper into the candidates' communication strategies and providing valuable insights for the subsequent tasks.

2. Sentiment Analysis (Task 2):

Building upon the thematic analysis, I devised a strategy to track and visualize the evolution of each candidate's themes and sentiments throughout the campaign. My approach included:

Sentiment Scoring with Advanced Lexicons: I suggested applying sentiment analysis techniques using the VADER and SentiWordNet lexicons to assign sentiment scores to each tweet. This would quantify the positive, negative, or neutral tone of their messages, providing a measurable way to assess emotional content.

Time-Series Visualization of Sentiment Trends: I proposed creating interactive line graphs to plot sentiment scores over time for each theme and candidate. This visualization would allow us to observe trends, fluctuations, and shifts in sentiment corresponding to key campaign events or strategies.[3]

Linking Themes to Sentiment: To provide a comprehensive view, I recommended integrating thematic and sentiment analyses by linking themes directly to their associated sentiments. This involved designing visualizations where users could explore how sentiment varied across different themes over time, offering insights into the candidates' strategic messaging adjustments.

Tone and Emotional Engagement Categorization: I introduced the idea of categorizing tweets based on tone (e.g., supportive, aggressive, neutral) and incorporating this into our visualizations. By analyzing the tone alongside sentiment, we could demonstrate the emotional engagement levels of the candidates and how they influenced public perception.

My strategy allowed us to uncover deeper insights into how each candidate adapted their messaging and emotional appeals throughout the campaign, highlighting their communication tactics.

Critical Moment Analysis (Task 3):

I conceptualized an innovative method for identifying critical moments in the campaign where a candidate could have tweeted differently. My strategy involved:

Defining Critical Moments Through Data Analysis:

Engagement Metrics Identification: I suggested analyzing engagement metrics (likes, retweets) over time to identify peaks and troughs, which could indicate moments of significant public interest or disengagement.

Sentiment and Tone Shifts: I proposed tracking sudden changes in sentiment and tone to pinpoint moments when public perception shifted significantly, potentially affecting the campaign's momentum.

Alternative Tweet Simulation and Impact Projection:

Creating Predictive Models: I introduced the idea of simulating alternative tweets by modifying sentiment intensity or changing the thematic focus. Using predictive analytics, we could estimate how these alternative messages might have influenced public engagement.[4]

Visualization of Potential Outcomes: I designed visualizations to compare actual engagement metrics with projected ones based on the alternative tweets. This would visually illustrate how different messaging strategies could have altered responses from the public or the opponent.

Strategic Recommendations:

Actionable Insights: Based on our analyses, I suggested we provide strategic recommendations on how the candidates could have adjusted their messaging for better outcomes, backed by data-driven evidence from our simulations.

By designing this comprehensive method, I enabled our team to explore the dynamic interplay between messaging strategies and public response, offering valuable lessons on the potential impact of different communication approaches.

B. Dataset preprocessing

The data preprocessing was completed by me and my team members. The team members were responsible for preliminary screening of the original data set and removing unnecessary attributes. Then, to meet the requirements of the first task, I used the spaCy library to clean the text attributes in the tweets, remove stop words, and further remove low-frequency words and special symbols (such as "http" and "-

"). These processing steps laid a solid foundation for subsequent topic modeling and sentiment analysis.[5]

II. TASK 2

I participated in all visualization tasks in all three parts, so below I will describe what I did in the visualization under the three topics respectively.

A. *Topic analysis and interactive visualization*

Determine/design visualization methods

First, I identified five possible discussion topics in the tweets of the two candidates through SOM clustering and high-frequency word analysis. Based on these topics, I decided to use interactive stacked bar charts to show the interaction of different discussion topics. Stacked bar charts can effectively show the overall interaction level of each topic and the specific contributions of different interaction types (such as forwarding and likes), helping the audience to intuitively understand the popularity and discussion heat of each topic.[2][6]

And to show the engagement of each topic, I use the ratio of likes to reposts as a reference value for engagement. This design can better reflect the popularity of each topic among the audience and the depth of user interaction, thus providing a reliable basis for subsequent data analysis.

Determine/Design visualization components (coordinate axes, visual variables, etc.)

Coordinate axis arrangement: I chose to use the x-axis to represent different topics and the y-axis to represent the level of interaction (reposts and likes), so that different topics can be easily compared.

Color scheme: Set different colors for different types of interaction (such as reposts and likes) so that the audience can easily distinguish them. The choice of color ensures the clarity of data and enhances the efficiency of information transmission.[7]

User interaction design

Hover interaction: Users can hover over different parts of the bar chart to view the specific number of interactions (such as the number of reposts and likes for a specific topic). This interaction can help users understand the detailed interaction information of each topic more deeply.[8]

Evaluation and Iteration

Based on the heuristic evaluation and user feedback, I made several improvements to the stacked bar chart, including adjusting the color scheme to improve contrast and adding hover tips to better understand the interactions for each topic. These improvements made the final visualization more readable and user-friendly.[9]

Evaluation Analysis

During the evaluation, I found that the initial version of the bar chart did not have clear color contrast, making it difficult for users to distinguish different types of interactions. By adjusting the color scheme and adding more detailed hover tips, the user experience was improved and readability was increased.

B. *Sentiment Analysis Visualization*

Determine/Design Visualization Method

For the sentiment analysis part, I analyzed the time series nature of the data and the need to compare the changes in sentiment between candidates. And by observing the patterns in the tweets, I summarized the classification method of tone. I finally chose to use an interactive line chart to show the changes in the candidates' sentiment throughout the campaign. Line charts are best suited to display time series data, especially when the sentiment fluctuates over time, and can effectively convey the evolution of sentiment during the campaign.[3][10]

Identify/design visualization components (axes, visual variables, etc.)

Axis arrangement: I chose to use the x-axis for time (months) and the y-axis for the range of sentiment scores, from positive (1) to negative (-1). This setup allows for a clear view of how sentiment changes over time.

Color scheme: Choose appropriate color coding for different sentiment types, for example, green for positive sentiment, red for negative sentiment, and gray for neutral sentiment. This allows users to quickly understand the overall trend of sentiment and the contrast between candidates.[7]

User interaction design

Hover interaction: Users can hover over a point in time to view the specific sentiment score at that moment, which allows users to gain insight into the state of sentiment at a particular point in time.

Highlighting: Users can click on a line to highlight it, so they can focus on the sentiment changes of a specific candidate and compare the sentiment trends of different candidates.[8]

Evaluate and iterate

In the early version of the sentiment line chart, users reported that the lines of sentiment categories were difficult to distinguish in some cases, especially when the color contrast was not high. Based on this feedback, I adjusted the color contrast and increased the thickness of the lines to improve visibility. In addition, some users suggested adding a trend line to show the trend of overall sentiment changes, so as to observe the emotional fluctuations of candidates more intuitively. Therefore, I added an average trend line of sentiment changes to the graph. [9]

Evaluation and Analysis

After several iterations, the final line chart was highly recognized by users. By adjusting colors, adding trend lines, and improving interactive functions, users said that they could better understand the emotional changes of different candidates during the campaign. These improvements significantly improved the ease of use of visualization and the accuracy of information communication, making it more suitable for campaign sentiment analysis scenarios.

C. Critical Moment Analysis Visualization

Determine/Design Visualization Methods

In the critical moment analysis, I found the months between May and June when Hillary's support rate dropped significantly. In order to show the changes in sentiment under different topics during this period, I designed a line chart to show the changes in average participation over time. The line chart can clearly show the trend changes during the campaign, especially in certain key time periods, helping us identify the time points when support rates rise or fall. And in order to achieve predictions, I used existing data to predict possible futures, which greatly improved the interpretability and accuracy.[4][11]

Identify/design visualization components (axes, visual variables, etc.)

Axis arrangement: I chose to use the x-axis to represent time periods (e.g., months) and the y-axis to represent changes in support. In this way, viewers can easily understand how each topic performs at different times and how support fluctuates.

Color coding: Different colors (e.g., red and blue) were chosen for positive and negative sentiment changes to quickly distinguish between positive and negative sentiment changes. This color coding not only makes the information expression clearer, but also allows users to more intuitively understand the emotional impact of different topics.[7]

User interaction design

Average sentiment score: When the user hovers the cursor over the graph line, I designed an interactive prompt that displays the average sentiment score, allowing users to see the overall change in positive or negative sentiment for the topic from May to June. This information helps users understand the impact of specific sentiment changes on support.

Highlight interaction: By adding a highlight interaction function to the graph line, when the user selects a specific topic, the sentiment change line of the topic will be highlighted, which allows users to focus on the analysis of a certain topic without being distracted by the data of other topics. [8]

Evaluation and Iteration

Based on the evaluation and user feedback, I found that users need to understand the changes in sentiment scores in more detail. Therefore, I added the function of displaying the average sentiment score when the mouse hovers. At the same time, in order to improve the readability of the chart, I also adjusted the line color and the layout of the chart to make the differences between the topics more obvious.[9]

Evaluation and Analysis

Through multiple iterations, the visualization of key moment analysis can finally effectively help users understand the changes in support rates across different topics and time periods, and help identify whether campaign strategies are effective under specific emotions. User feedback shows that such improvements greatly improve the understanding of the data and the possibility of further analysis.

Relationship with Learning Outcomes

My contributions to visualization design, interaction implementation, and evaluation reflect my comprehensive learning and application in this unit. Specifically:

From the courses in Week 8 and Week 9, I learned how to choose the appropriate visualization type to represent multivariate data and time series data, which was applied in sentiment and topic analysis.

The concept of decision support visualization in Week 9 helped me design the visualization of key moments and provide optimization suggestions for candidates.

The user interaction design content in the week 10 and week 11 labs influenced the way I added hover effects, highlights, and filtering functions to my visualizations to improve the user experience and ease of data exploration.

Finally, the evaluation skills learned in week 6 enabled me to effectively evaluate and iterate our visualizations to ensure that they are both informative and meet the needs of users.

REFERENCES

[1] Mu, K. (2023a). Week 01 - Introduction. COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[2] Mu, K. (2023b). Week 09 - Visualization Design and Representation. COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[3] Mu, K. (2023c). Week 08 - Exploratory Data Analysis. COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[4] Mu, K. (2023d). Week 09 - Visualization Design and Representation. COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[5] Mu, K. (2023e). Week 03 - Data Types / Semiology of Graphics (1). COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[6] Mu, K. (2023b). Week 09 - Visualization Design and Representation. [Same as Reference 2]

[7] Mu, K. (2023f). Week 02 - Human Visual System / Color. COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[8] Mu, K. (2023g). Week 05 - Fundamentals of HCI. COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[9] Mu, K. (2023h). Week 06 - Evaluation Methods. COMP5048/4448 Data Visualization Course, University of Sydney, Semester 2, 2023.

[10] Mu, K. (2023c). Week 08 - Exploratory Data Analysis. [Same as Reference 3]

[11] Mu, K. (2023d). Week 09 - Visualization Design and Representation.

How would you visualise your data?

COMP5048/4448 Assignment 2

line 1: Yidan Fu
line 2: *510027504*
line 4: CC-08-G06
line 5: Katherine Mu

I. Task 1

In the early stage of the project, I had several in-depth discussions with the junior members to clarify the task flow and assign their respective responsibilities. Through these exchanges, we initially planned the general framework of data analysis, which laid the foundation for the subsequent implementation. In the division of tasks, I undertook the cleaning and analysis of the data in the very early stage, and was also responsible for the preparation of the text for the sentiment analysis section, as well as the final organization and validation of the report.

A. *Data Cleaning and Preprocessing*

I first carried out the initial data cleaning and simple preprocessing work, mainly including data inspection, removing irrelevant attributes to find out the missing values appearing in the data, and text standardization. First, the data was initially examined to remove attributes that were not needed in this task (e.g., in_reply_to_status_id, in_reply_to_user_id, etc.) to reduce possible interference and retain core information. Secondly, the data is processed for missing values, since the data is text data, the conventional filling method cannot be used, so for the attributes that have fewer missing values, I directly use the rows where the missing values are located to delete them. For attributes with a missing rate greater than 80, I choose to delete the attribute directly to ensure the overall cleanliness of the data. Finally, in order to ensure the unity of the text format, I removed the deactivated words from the tweets and deleted the noisy information in the data to ensure the cleanliness and readability of the text, and also unified the word forms in the tweets to minimize the interference caused by different word forms. Further processing of the data was left to another team member who was able to do so [1].

B. *Orientation of Sentiment Analysis*

After the data cleaning, I focused on the direction of sentiment analysis, trying to reveal the emotional expression strategies of the two candidates in the campaign. In order to more accurately understand the emotional expressions of the two candidates on social media, I reviewed relevant literature and thoroughly researched the political styles and emotional characteristics of the two candidates as well as various key political points in time during the 2016 campaign. Through these studies, I gained an in-depth understanding of the various time points and emotional expressions, which provided a strong background support for the sentiment analysis.

C. *Organization of the report*

After all team members had completed their respective responsible sections, I carried out a systematic organizing work for the report. First, I carefully reviewed each member's content to ensure the accuracy of the information conveyed, and checked the logical flow of each section one by one so that the structure of the content conformed to the overall framework of the report. Next, I standardized the format and layout of the report to enhance the overall professionalism and readability. At the same time, in order to maintain the consistency of the report's style, I made appropriate adjustments to the expression of each part, so as to make the tone and wording of the text more harmonious and to avoid any inconsistency in tone among the parts. In addition, in the process of integration, I focused on the articulation of the contents of each section, so that the various parts of the report supported each other and the transition was natural. Through this series of meticulous organization, I finally ensured the consistency of the report in terms of structure, logic and presentation, laying a solid foundation for the overall presentation of the project.

II. Task 2

In the part of visualizing and displaying the sentiment analysis part, I gradually constructed a comprehensive graph that effectively displays the candidates' sentiment and tone strategies, from determining the methodology, designing the structure, optimizing the interaction, to conducting the evaluation and analysis. As for the average sentiment distribution graph, the sentiment over time graph, and the tone distribution graph, they were done by me in consultation with another member of the group, but I did not give much operational help.

A. *Identify/design visualization methods*

In order to clearly present the mood changes of the two candidates in different themes, tones and key time points, I chose heatmap as the core presentation method. This choice takes into account the characteristics of the data type and the need for intuitive expression of emotions. Sentiment data is structured in two main forms: Quantitative Data and Categorical Data, which gives heatmaps a unique advantage in visual communication [2]. In the case of Quantitative Data, this is represented in sentiment analysis as sentiment scores, ranging from negative (representing negative sentiment), zero (neutral sentiment), and positive (positive sentiment). In heatmaps, this numerical information is encoded through color shades and hue changes, with the darker the color the stronger the sentiment, helping users to quickly capture the fluctuations and magnitude of changes in sentiment. Meanwhile, by using warm colors (red, orange, etc.) and cold colors (blue, cyan, etc.) for the distinction between positive and negative emotions, the trend of emotions is clearer and users can intuitively judge the extreme degree and direction of emotions. For categorical data, the main purpose is to categorize the tone of voice, such as “supportive” and “aggressive”, as well as the theme of the classification. This combination of numerical and categorical data makes heatmaps the best choice for presenting sentiment changes, not only for simplicity and clarity, but also for visualizing trends and extremes in sentiment.

B. Determining/Designing Visualization

In the design of this heatmap, I set the horizontal axis to five high-frequency topics during the election campaign, such as “campaign criticism and tax plan”, “debates and family issues”, etc., to help users clarify the distribution of emotions in different topics. The vertical axis represents tone categories, including “supportive” and “aggressive”, which can visualize the differences between the two candidates in each tone category. To enhance the visual effect, a color gradient is used for the visual variables: warm colors (e.g., red to yellow) represent positive emotions, while cool colors (e.g., blue to light blue) represent negative emotions. Emotional intensity is more pronounced through color shades, with dark red areas representing peaks of positive emotion, often reflecting high levels of candidate support or confidence in a topic, and dark blue indicating peaks of negative emotion, often corresponding to more critical or aggressive expressions. This choice of color scheme is based not only on the needs of the data features, but also takes into account natural human identification preferences and ways of perceiving color, such as sensitivity to lightness and saturation [3]. In terms of color perception, the contrast of darker shades helps to highlight the range of emotional fluctuations and emotional patterns at important points, thus making the direction and intensity of emotions readily apparent, and helping the user to observe the differences in the emotional strategies of the two candidates along the timeline at different campaign stages and themes.

C. Determining/Designing user interaction

In the user interaction design, I paid particular attention to improving the operability of the data visualization and the exploitability of the information. To achieve this goal, I introduced two core interaction functions that enable users to explore the changing patterns of the sentiment data more flexibly and interpret the candidates' sentiment strategies at a deeper level of detail.

First, I designed a time slider function that allows users to select a specific period for observation based on their needs. By sliding the timeline, users can dynamically adjust the visualization content and conveniently compare and analyze the changes in emotions at different time points. This design not only facilitates users to observe the overall development trend of emotions from a macro perspective but also effectively assists users in analyzing the fluctuation of emotions before and after significant events, thus revealing the impact of critical nodes on the expression of emotions. For example, users can use the slider to compare the changes in sentiment during the debate with those before and after the policy announcement, thus visualizing how the two candidates adjusted their emotional strategies during different campaign stages. Through this interactive feature, the time series of emotional changes can be vividly displayed, helping users to understand the dynamic trend of emotional evolution more intuitively [4].

Secondly, adding the hover function further enhances the depth of the user experience. When users hover over a specific area on the chart, they can instantly get detailed information about the area, including the sentiment score, tone category, and corresponding theme information. This nuanced design allows users to observe overall trends and drill down into the details of individual mood swings when needed. Especially when analyzing a candidate's specific attitudes and tone choices on a given topic, the detailed information provided by the hover function makes it easy for users to capture the subtle connections between sentiment and tone. For example, users can hover over Trump's emotional intensity on "Campaign Criticism and the Tax Plan," learn about his aggressive expressions over a specific period, and further analyze how emotion and tone interact. This interaction was designed to provide a rich hierarchy of information to satisfy users' need to analyze sentiment trends from a global perspective and support their nuanced interpretations at the level of specific tones and themes [4].

Overall, the setup of these two interactions makes the entire chart intuitive and easy to understand regarding visual presentation while being highly actionable. Users not only have easy access to the overall pulse of sentiment changes but can also explore the subtle changes in mood and tone in-depth to develop a multi-layered understanding. This design aims to enhance the employability of the data, enabling users to gain a more comprehensive and three-dimensional interpretation of the candidates' emotional expressions and strategic adjustments during the campaign process, thus gaining a deeper understanding of the data [5].

D. Determining/Designing user interaction

After the charts were designed, I set up a detailed evaluation process to ensure that the charts met expectations in terms of messaging, interactivity, and user experience. The evaluation process included a multi-dimensional examination of clarity, information accuracy, and ease of interaction. Specifically, I used a questionnaire to quantitatively analyze user feedback to ensure the reliability and consistency of the feedback. The questionnaire covers indicators such as the ease of use of slider and hover functions, the legibility of chart colors, etc. to comprehensively assess the user experience in actual operation. In addition, I also conducted experiments to observe users' performance on different platforms to gather intuitive feedback on the charts. Based on these data, I optimized the color contrast and interaction details to further improve the readability and user experience of the chart. This series of evaluation steps guaranteed the effectiveness and usefulness of the chart design [6].

REFERENCES

[1] M. Takatsuka, "COMP5048 Visual Analytics, Week 07: Exploratory Data Analysis, EDA(I)(II)," School of Computer science, University of Sydney, 2024.[Lecture Notes]

[2] M. Takatsuka, "COMP5048 Visual Analytics, Week 02: Data Types," School of Computer science, University of Sydney, 2024.[Lecture Notes]

[3] M. Takatsuka, "COMP5048 Visual Analytics, Week 02: Color," School of Computer science, University of Sydney, 2024.[Lecture Notes]

[4] Plotly Technologies Inc., "Plotly Python Open Source Graphing Library," Plotly. Available: https://plotly.com/python/. Accessed: Nov. 4, 2024.

[5] Plotly Technologies Inc., "Python API Reference," Plotly. Available: https://plotly.com/python-api-reference/. Accessed: Nov. 4, 2024.

[6] M. Takatsuka, "COMP5048 Visual Analytics, Week 06: Evaluation," School of Computer science, University of Sydney, 2024.[Lecture Notes]

Individual Report

COMP5048/4448 Assignment 2

Albert Huang
530215563

CC-08-G06
Katherine Mu

III. Group Conrtibution

A. Data Processing

The initial data processing was done by me. Before the formation of project group and my first meeting with my group members, I put effort in advanced to examine through data and applied preprocessing technique to prepare data.

To preprocess the data in the correct way, I make sure myself understand the dataset. I read through the source of the dataset on kaggle to gain a comprehensive view of the dataset: background, usage, feature, related works and analysis… After that, I have a clear picture of the dataset is about then carry on into data exploration.

I mainly use python as the tool to explore and preprocess the dataset. I verified the meta data of the dataset with python coding skills to ensure the there is no error or missing value produced by wrong version of the dataset or undetectable downloading issues. I went on to look for any outstanding mistakes in the dataset like duplicates, outliers, data type error that might jeopardize the integrity and accuracy of the dataset and lead to bias or deviated outcome later. In the process of exploring data, I summarized every need of assignment and built a relation network graph based on the need and attribute shown in the dataset. From that work, I was able to identify useable and informative attributes and those attributes useless in this assignment.

Data cleaning was carried out according to the work I done above. I wrote python code to eliminate useless attributes, left the dataset with only 10 essential attributes: id, handle, text, is_retweet, original_author, time in_reply_to_screen_name, in_reply_to_status_id, in_reply_to_user_id, is_quote_status, and retweet_count.

My works on data processing accelerate the process on project initiation, reduce computational resource need for later tasks. And I believe my role as early participant before formal group assembling create good working moral and solid foundation of our project.

B. Document Management

Our group collaborates in a high individual functioning way. No official shared online storage of documents was created. However, using my experience in project management, I kept a copy of every document related to this project and created a Github repository for a better version control on codes and reports. While requesting all documents being sent to me for task C completion, I kept those documents organized in local and remote repository in case of unforeseen risks that might interrupt and damage the integrity of contents in our documents.

C. Programming

I did not involve in direct programming that adopted in the final version of the project except data processing. However, I was actively involved in the designing visualization part of the project (detailed explanation in visualization-related tasks contribution section) that create the specific requirements coding is tailored to.

Despite the adopted version, I also have multiple unadopted attempts of creating codes for SOM and spatialization and a multiple layer graph building system (as mentioned in lecture), however, most of the works there are unfinished, not meet ideal standard or abandoned due to constrains of the project (time, resources).

D. Deriving Visualization Applications

As mentioned in above section, there is attempt for a self-made visualization application that would allow SOM and relationship of tweet author and tweets' supporting rate to be shown in an organized overlapping way. This approach was thought and designed to be a good tool for prediction on task C as it might show some degree of truth trajectory. Unfortunately, the attempt failed by lack of time and resources. However, there is quite amount of effort in this attempt with fragment and research on related area conducted.

E. Literature Survey

Our group function in a highly independent but interconnected way that each person was responsible for conducting their own literature survey according to the needs of their roles and responsibilities.

My part of literature survey mainly focuses on visualization applications and prediction using interactive visualization. There are:

General visualization application theory literature review [1, 2, 3, 4], practise of visualization application review [5, 6, 7], a few Github tutorials on building visualization application [8, 9, 10], and visualization-prediction-related literatures [11, 12, 13].

Although not all literatures review has a strong connection to the project, they still deepen my understanding of visualization analytics in a comprehensive and insightful way, providing broader context and enriching my knowledge of the field.

F. Project Management

There is no official project management approach like waterfall, agile adopted for this project. However, I considered myself in the position of project manager on some occasions that keep the team's direction on a correct path. Through out the process of project building, I continuously monitor the progress and content created for the project. And I updated group with new needs and requirements that would benefit our work at best as the project progresses. I made decisions on changes and modifications on project, in multiple areas like visualization building, redesigning, request gathering, auditing, report alterations… I considered myself actively participant in every possible project management aspect using my IT management expertise and experience on serval project leading. Key turning points of the project like the in early stage build and design evaluation system for our visualization, adopting SOM as visualization, spotting tone and emotional engagement requirements, task C decision-making… My heavily involvement and idea creation in these moments help the team make the best decision in a constrained scenario.

Besides that, I also arranged risk management for the project with precaution plan to mitigate risks like unsatisfied visualization quality by attempting different approaches of the same task.

IV. VISUALIZATION-RALTED TASKS CONTRIBUTION

The part of visualization I made significant contributions to is task C, where a visualization of critical moment and a visualization of predictions of alternative tweets are produced.

A. *Determining/Designing a Visualization Method*

For the first objective in task C, I chose line graph as the visualization method. It is decided in this way because the critical moment analysis would be a temporal analysis that involve in time attributes. There is expected to be a continuity, and a trend of average engagement needs to be presented, therefore, line graph would be the suitable choice here. (Lecture Week 9 Design and Representation)

For the second objective in task C, there is a need for comparison. In my design, the alternative tweet is derived from the visualization by comparing effects on support rate changes on different combinations of sentiment types and topics. It is obvious that comparison is the key thing here. Hence the KPI is also created here: the best combination with highest positive support rate change. (Lecture Week 5 KPI Deriving Methods). Naturally, bar chart was selected here to satisfy the core user activity: comparison of data. More specially, stacked bar chart was used to show the two sentiment types for every topic. (Lecture Week 9 Design and Representation) This method can also reduce the amount of comparison as all combinations are shown in one graph, user can quickly find out the best match among all possibles.

What also needs to be mentioned is that an alternative method for second objective was tried. I designed the method as a SOM with relationship graph of two candidates and their tweets' supporting rate overlapping on top. This was inspired by the last lecture content of "arms around the world" visualization. It is expected to shown clusters as some form of alternative tweets with different combination of features (topics, sentiments…) and the overlapped relation graph to shown desired support rate change tendency. (Lecture Week 9 Relation Data and Lecture Week 11 SOM) An alternative tweet is expected to be found following the visualization. However, due to the time and resources constraints, it was not adopted in the final version of the project.

Overall, in determining and designing visualization method, I carefully thought of the goal of each visualization, tailored their needs by proposing different visualization approaches.

B. *Determining/Designing Visualization*

The whole design process of the two visualization here, I am the main person responsible as well as the decision maker. Our group's coder working as assistant to provide coding and opinion support.

For the critical moment visualization, when designing the axes arrangement, the first to considered is the what are the key attribute / data. By analyzing the requirement and semantics meaning of the critical moment (Detailed explanation of this can be found in Group Report part), time data and average engagement is selected as the two attributes. Next to consider is which construct to be used. In this visualization, orthogonal construction is selected since there is no dependency between the two attribute and there is no correlation relationship between. (Lecture Week 3 SOG) It would reduce the complexity of the visualization and improve maintainability. The position of the x and y axis are set according to the general preference of visualization which in a temporal analysis, x-axis is preferred to be time axis to better present the trend and increase readability.

For the alternative tweet visualization, design of axis arrangement took a similar approach where the key attributes used were first decided by conducting task and data analysis (more details in Group Report). In here the key attributes are topics, sentiment types and support rate change. With method had been determined to be a stacked bar chart, orthogonal construction was also being propose here due to the characterises of attributes as well as guidelines from unit material. (Lecture Week 3 SOG). For better visibility and general user preference, the x-axis is deigned to be topics, support rate change as y-axis while two types of sentiment shown for each topic and topics are hheterogeneous dimension. (Lecture Week 3 SOG)

The main visual variable used in both visualizations are colour and legends (text variables). In my design, the colour selection representation the belonging of a attribute: Hilary Cliton or Donald Trump. Colour here represent a selective organization, (Lecture Week 3 SOG) and for better accessibility, the two colours are set to be blue and orange for colour-blind-safe. Legends was used to improve readability, telling user which candidate a colour is associated with.

I have produced symbolic representations of SOG for each objective while designing the visualization. The actual visualizations are derived from their corresponding symbolic representations of SOG and choosing appropriate methods fitting this representation. (Lecture Week 3 SOG) This designing approach helps me enhance the clarity and understanding of my design concepts. It promotes the consistency of my works and increases my overall design efficiency.

C. *Determining/Designing User Interaction*

User Interactions of the two visualizations are highly similar or the same in the sense of interaction types.

The design of interactions starts with understanding the need of interactions for the two visualizations. It can conclude that both visualizations express sufficient information that user needed for finding, agreeing and completing the KPIs and KGIs. Therefore, the need for selection and manipulation interactions are not strong here. (Lecture Week 4 HCI) On the contrary, another type of interaction, exploration and navigation interaction is desired as this type of interaction help user better understand and walk through our visually represented space. (Lecture Week 4 HCI) Highlight is the interaction that both visualizations have by default in our design. It could provide user with more detailed information and directed values, that enable deeper analytic processes.

Other than highlighting, I also modified the visualizations to allow user zoom in and out of visualizations and freely moving around in the visualize space. This is aimed to improve user's ability of focusing on details, gaining better control, growing better contextual awareness and most importantly, enhancing user engagement.

D. Determining/Designing the Evaluation Process

The two visualization I in charge here both adopted Questionnaire as the evaluation method. The technique and inspiration were gained from lecture slide week 6 Evaluation.

Evaluation is selected for the following reasons:

1. Efficient Collection: Questionnaires allow quantitative techniques to be taken, with a well-designed score / result calculation sheet, the feedback can be quickly gathered and stored in an organized way. Enables further analytic processes.
2. Standardization with Consistency: Questionnaires can have a same set of standard questions that gather information feedback of the general aspect of the visualization like scale of visibility, readability etc. These standard questions can be reused for efficiency.
3. Reduce Bias: Differ from internal evaluation, questionnaires answers by person outside the group can guarantee there is no bias. Subjectivity is ensured. Moreover, group can gain evaluation feedback from real users.
4. Flexibility: Questionnaires are flexible and easy to modify the structure and content to better suit the changing evaluation requirements as the project progresses.

I decided to adopt multiple types of question type to ensure the vast dimension of coverage in the features captured as feedback for our visualization. And each visualization will gain a tailored open-end question that ask user a specific question according to the visualization's KGI and KPIs.

E. Conducting and Analysing the Evaluation

The evaluation was conducted by me as the form of questionnaires. Each visualization in task C has its own questionnaire. I invited my friends to interact with the visualizations and taken the questionnaires. For each visualization, the result of its questionnaire was stored by questions. The highest and lowest scores will be eliminated, and average score will be pointed as the suggest outcome of that question.

As the result of evaluation, improvements have been made. More specifically, for critical moment visualization, scale of the overall visualization had been adjusted for better view and for prediction of alternative tweet, color should be made as color-blind friendly for accessibility.

References

[12] C. Upson, T. A. Faulhaber, D. Kamins, D. Laidlaw, D. Schlegel, J. Vroom, and A. VaASn Dam, "The application visualization system: A computational environment for scientific visualization," *IEEE Computer Graphics and Applications*, vol. 9, no. 4, pp. 30–42, 1989. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[13] D. Carr, "Guidelines for designing information visualization applications," in *Proceedings of ECUE'99*, Dec. 1–3, 1999.

[14] M. Sadiku, A. E. Shadare, S. M. Musa, C. M. Akujuobi, and R. Perry, "Data visualization," *International Journal of Engineering Research and Advanced Technology (IJERAT)*, vol. 2, no. 12, pp. 11–16, 2016. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[15] M. O. Ward, G. Grinstein, and D. Keim, *Interactive Data Visualization: Foundations, Techniques, and Applications*. Natick, MA: AK Peters/CRC Press, 2010; M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.

[16] S. R. Panta, R. Wang, J. Fries, R. Kalyanam, N. Speer, M. Banich, and V. D. Calhoun, "A tool for interactive data visualization: application to over 10,000 brain imaging and phantom MRI data sets," *Frontiers in Neuroinformatics*, vol. 10, p. 9, 2016.

[17] J. L. Rubio Tamayo, M. Barro Hernández, and H. Gómez Gómez, "Digital data visualization with interactive and virtual reality tools: Review of current state of the art and proposal of a model," *Digital Data Visualization with Interactive and Virtual Reality Tools*, vol. 2018.

[18] J. Chishtie, I. A. Bielska, A. Barrera, J. S. Marchand, M. Imran, S. F. A. Tirmizi, and S. Jaglal, "Interactive visualization applications in population health and health services research: systematic scoping review," *Journal of Medical Internet Research*, vol. 24, no. 2, p. e27534, 2022.

[19] R. Verma, *Data Analysis and Visualization Web App*, GitHub repository, https://github.com/rituraj009/Data-Analysis-Visualisation-Web-App. Accessed: Oct. 31, 2024.

[20] McKinsey & Company, *Vizro: A Data Visualization Library*, GitHub repository, https://github.com/mckinsey/vizro. Accessed: Oct. 31, 2024.

[21] A. Prakash, *Interactive Data Visualization Application*, GitHub repository, https://github.com/arjunprakash027/interactive-data-visualization-application. Accessed: Oct. 31, 2024.

[22] L. M. Zintgraf, T. S. Cohen, T. Adel, and M. Welling, "Visualizing deep neural network decisions: Prediction difference analysis," *arXiv preprint arXiv:1702.04595*, 2017.

[23] I. Roesch and T. Günther, "Visualization of neural network predictions for weather forecasting," in *Computer Graphics Forum*, vol. 38, no. 1, pp. 209–220, Feb. 2019.

[24] S. Chiang, R. Moss, A. P. Black, M. Jackson, C. Moss, J. Bidwell, and T. Loddenkemper, "Evaluation and recommendations for effective data visualization for seizure forecasting algorithms," *JAMIA Open*, vol. 4, no. 1, p. ooab009, 2021.