**Assignment 3**

Melissa Hunfalvay

Email: [Melissa.Hunfalvay@gmail.com](mailto:Melissa.Hunfalvay@gmail.com)

Data 650 9040

Summer 2022

Professor Ozan Ozcan

Date: July 19th, 2022

**Table of Contents**

Introduction………………………………………………………………………………………..3

Purpose……………………………………………………………………………………5

Methodology………………………………………………………………………………………7

Data Loading………………………………………………………………………………7

Data Verification…………………………………………………………………………..7

Data Exploration…………………………………………………………………………..9

Experimental Design & Statistical Methodology………………………………………..21

Data Pre-Processing Steps…………………………………………………………24

Data Analysis…………………………………………………………………………………….30

Word Cloud…………………………………………………………………………30

Trend Analysis……………………………………………………………………..32

Apriori Rules Analysis……………………………………………………………..33

Discussion………………………………………………………………………………………..42

Limitations……………………………………………………………………………….43

References………………………………………………………………………………………..47

Appendix…………………………………………………………………………………………48

Appendix A: Additional Analysis……………………………………………………….49

Appendix B: Number of tweets per topic area by candidate…………………………….74

**Introduction**

*Clearly explain the topic choice and the purpose of the study. Explain what data you are analyzing.*

The dataset used for this analysis was from Crowdfunder and was called “FIRST\_GOP\_DEBATE” (Crowdfunder, 2015). The data included Twitter tweets about the early August 2015, GOP debate in Ohio.

The dataset has 13,871 rows (tweets) and 21 columns (Figure 1). The columns and descriptions can be seen in Table 1.

Text

Description automatically generated

Table

Description automatically generated

Figure 1: Number of rows in the First\_GOP\_Debate dataset

Table 1: Column Names and Descriptions

|  |  |  |
| --- | --- | --- |
|  | **Column Name** | **Column Description** |
| A | Id | Row id |
| B | CANDIDATE | Candidate mentioned |
| C | CANDIDATE\_CONFIDENCE | Confidence of the candidate mentioned |
| D | RELEVANT\_YN | "no" means that the tweet was meant to be part of the dataset but was not available when contributors went to judge it |
| E | RELEVANT\_YN\_CONFIDENCE | confidence in the existence/non-existence of the tweet |
| F | SENTIMENT | Tweet Sentiment |
| G | SENTIMENT\_CONFIDENCE | Confidence of the sentiment |
| H | SUBJECT\_MATTER | Tweet subject |
| I | SUBJECT\_MATTER\_CONFIDENCE | Confidence of the subject matter |
| J | CANDIDATE\_GOLD | whether the candidate was included in the gold standard for the model |
| K | NAME | the user who tweeted |
| L | RELEVANT\_YN\_GOLD | whether the tweet yn value is golden |
| M | RETWEET\_COUNT | number of times the user has retweeted |
| N | SENTIMENT\_GOLD | if the profile is golden, what is the sentiment |
| O | SUBJECT\_MATTER\_GOLD | whether the subject matter was included in the gold standard for the model |
| P | TEXT | the text of the tweet |
| Q | TWEET\_COORD | if the user in column K has location turned on, the coordinates as a string with the format "[*latitude*, *longitude*]" |
| R | TWEET\_CREATED | When tweet was created |
| S | TWEET\_ID | Tweet identification number |
| T | TWEET\_LOCATION | User’s country, city, state |
| U | USER\_TIMEZONE | User time zone |

Contributors to the dataset were asked to conduct sentiment analysis and data categorization. First, they were required to determine if the tweet was relevant to the debate, then which candidate was mentioned, what subject was mentioned and the sentiment for each tweet.

This FIRST\_GOP\_DEBATE dataset provides a large dataset from which to experiment using different methods of big data text analysis. The data has two “big data” fundamental characteristics; high variety and high value for the election (IBM, 2022).

Sentiment analysis of the twitter data from the FIRST\_GOP\_DEBATE dataset was conducted to determine if positive, negative, or neutral tweets. Sentiment analysis is an application of natural language processing (NLP) that identifies expressions that reflect opinion and attitudes (positive, negative, neutral) towards entities (in this dataset, the candidates; Cambria, Das, Bandyopadhyay, Feraco, 2017). These classifications can then be used to determine if any interesting patterns emerge in the text about the candidates or topics of discussion.

Many people were surprised by the outcome of the 2016 elections (Flores, 2016). How did Mr. Trump, a seemingly long shot for the presidency, win?

Mr. Trump was an unconventional candidate. After the first debate, a Washington Post article titled “Winner and losers from the first Republican presidential debate” stated that Mr. Trump was: a) “the hardest candidate to judge” and; b) “he may be beyond normal political predictions” ([Cillizza,](https://www.washingtonpost.com/news/the-fix/wp/2015/08/06/winners-and-losers-from-the-first-republican-presidential-debate/) 2015).

Can we learn anything from this early debate that could help explain the unexpected outcome? Could future political candidates and their advisors learn anything from these early debates that could help them run successful campaigns in the future?

Given these questions, the *purpose* of this analysis is to find associations within the FIRST\_GOP\_DEBATE dataset that may help explain the unexpected success, of an unconventional, and difficult to predict, Mr. Trump. More specifically, to answer the following questions from the dataset:

1. What types of sentiments resonated with the audience? Did Mr. Trump use those sentiments?
2. Which candidate garnered the most attention from the debate?
3. Does subject matter confidence relate to candidate confidence?

To better understand the data, and to answer the specific questions, three types of analysis were conducted:

1. *Word cloud:* which are visualizations of word frequencies within the dataset. The visual word cloud provides insights of the commonly occurring terms within the data. Colors can be used to make words “pop” and provide another element of visual learning (Alida, 2012).
2. *Trend analysis:* shows the number of tweets leading up to, on the day of and after the debate. The tweets are categorized by sentiment (positive, negative, and neutral). A ‘timeline’ such as this can help in two unique ways: first, to understand the sentiment over time and, second to understand the sentiment during the debate and how that sentiment analysis compares to tweets before and after. This essentially provides insight into the degree of sustainability of sentiments which may influence the level of commitment to a sentiment and possibly influence or even help to predict general feelings about future GOP debates and candidates.
3. *Apriori Rules:* This type of analysis is association rule learning, whereby associations, or sets of frequent items, between variables are explored to try and find insights and hidden relationships in large datasets (Han, Kamber, Pei, 2011).

Other methods were also viewed to determine if they helped provide insights into the main questions for the analysis (see Appendix A), these include:

1. K-Means Cluster analysis: which is designed to group similar terms together while also creating clusters of differing terms in separate groups.
2. Dendrogram: which is a tree showing the taxonomy of relationships
3. Network of terms analysis: is used to analyze the relationship among terms to include the recurrence of the terms.

These methods were not considered to either; a) provide any new information or insights beyond the three methods chosen, and/or, b) provide information pertinent to answer the specific questions for the analysis. Therefore, these types of analysis were not used to address the analytical questions.

**Method**

1. **Data Loading**

*Use DB2 on Cloud web console to load the data. At the define step, edit the data types.*

*When the load completes, you will be redirected to the status page. Take a screenshot of that page and discuss the load status in the paper.*

Figure 2 shows the status page and successful upload of the FIRST\_GOP\_DEBATE data. The figure shows number of rows read (13,871), rows that were loaded (13,871), and the number of rows that were rejected (0). It also shows any errors in loading (no errors) and the target schema where the data is stored. Finally, it shows the start and end times for loading the data. The source file is shown which states where the data originated. A log of the load status can be downloaded for more information.

Graphical user interface, application, Teams

Description automatically generated

Download the Log

Source File

Figure 2: Status Page

1. **Data Verification**

*Discuss the programmatic approaches you used to verify if the data was loaded as expected, including checking the row counts and table metadata.*

The first approach used to verify the data was loaded as expected was to use the “View Table” to visually inspect the uploaded data (Figure 3). Although not a programmatic approach, it is always important to visually confirm expectations of how the file should appear.

Graphical user interface, application

Description automatically generated

Figure 3: View Table button used to visually inspect the uploaded data file

Per the instructions, various columns needed adjustments to the number of characters, for example, candidate gold needed to change to 22 characters. Figure 4 shows this manual change was accepted and is displayed in the successful uploaded file.

Table

Description automatically generated

Figure 4: Changes to the Candidate Gold Number of Characters (n = 22).

The first programmatic way to inspect the data file is to check for the number of rows (see Figure 1).

The second programmatic way to inspect the data file is to check the file name, number of columns, the creator, and the change time (ctime). The data file is shown to have 21 columns, with the correct ctime, creator and file name (see Figure 5).

Text

Description automatically generated

Figure 5: Programmatically shows the file name, number of columns, creator and change time (ctime).

The third programmatic way to check the upload and to begin to explore the data is to examine some of the meta data. This includes the variable types (COLTYPE; Figure 6), if variables have no data (NULLs) and to again validate the length of each variable (LENGTH). Figure 6 shows there are missing values in the variables throughout the dataset.

Text

Description automatically generated with low confidence

Table

Description automatically generated

Figure 6: Metadata for the FIRST\_GOP\_DEBATE data set

1. **Data Exploration**
2. *Write a query that returns how many times each candidate is mentioned in column B. Include the query in the SQL script. Include a screenshot of the query output and output discussion in the paper.*

Figure 7 and 8 shows the query and return for how many times each candidate is mentioned in the tweets. It is important to note that in 7491 tweets no candidate was mentioned. The candidate with the highest number of tweets was Mr. Trump. This simply shows that Mr. Trump was “spoken about” more often than other candidates. It does not however, state whether people liked or disliked what he was saying or even what he was talking about. It does, however, give an indication as to how much attention Mr. Trump was getting early in the debate season.

The next candidate with the highest number of tweets was Mr. Cruz who was mentioned with a frequency of only 22% (n = 637) of the number of tweets compared with Mr. Trump. The candidate mentioned the least number of times was Mr. Kasich (n = 242).

Graphical user interface, text, application

Description automatically generated

[Graphical user interface

Description automatically generated with low confidence](https://cognitiveclass.ai/courses/what-is-big-data/)

Figure 7: SQL query and output for the number of times a candidate is mentioned in the tweets

Chart, pie chart

Description automatically generated

Figure 8: Pie chart for the number of times a candidate is mentioned in the tweets

1. *Write a query that returns the following 5 columns:*

*Column 1 – Candidate name*

*Column 2 – Number of negative mentions for a candidate in column 1*

*Column 3 – Number of positive mentions for a candidate in column 1*

*Column 4 – Number of neutral mentions for a candidate in column 1*

*Column 5 – Total number of mentions for a candidate in column 1*

When breaking out the tweets based on sentiment, the debate itself seemed to draw a significant number of negative sentiments (Figure 9). Specifically, negative sentiments not associated with any candidate were the most frequent (n = 4724). The next most frequent sentiment was neutral and not associated with any candidate (n = 2087). Total number of tweets (negative, positive, and neutral) not associated with any candidate were most frequent (n = 7491).

Table

Description automatically generated with medium confidence

Figure 9: Candidate total, positive, negative and neutral tweet mentions

When examining tweets related to candidates, again, Mr. Trump garnered the most attention with a total of 2813 tweets. Negative (n = 1758, 62%), positive (n = 609, 22%) and neutral (n = 446, 16%) tweets for Mr. Trump show that a predominance of negative sentiments. The next closest candidate was Mr. Bush with 704 dedicated tweets. Mr. Bush also had a majority of negative tweets (n = 598; 84%).

Only two candidates had more positive than negative tweets ( Mr. Rubio and Mr. Cruz). Interestingly, when examining who won the debate neither Mr. Bush nor Mr. Rubio were sighted as ‘winners’ by reports in the Washington Post ([Cillizza,](https://www.washingtonpost.com/news/the-fix/wp/2015/08/06/winners-and-losers-from-the-first-republican-presidential-debate/) 2015). In fact, Mr. Bush was seen as the ‘biggest loser’ of the debate by Politico (Neffingger, 2015).

Figure 10 shows the total number of negative, positive and neutral tweets (irrespective of candidate) across all the tweets from the debate. The total negative tweets were more than four times more frequent than positive tweets and almost three times more frequent than neutral tweets. This is consistent with figure 8 and shows a general dissatistfcation with the debate. These results also show that most watchers of the debate were not happy with at least some portion of the discussion.

Chart, bar chart

Description automatically generated

Figure 10: Total number of negative, positive and neutral tweets

Given the results of figures 9 and 10 it is clear that negative sentiments outweighed positive or neutral sentiments. This was true when tweets were associated with specific candidates and when tweets were not linked to a candidate. Such results show dissatisfcation with the debate in general as well as the candidates. Nevertheless, this data helps answer the first question of the analysis showing that the most frequent sentiments were negative and Mr. Trump also had the most negative number of tweets.

1. *Advanced SQL queries: Write four (4) additional different meaningful queries. Use the subqueries, and aggregate functions. For each query, include the SQL statement in the SQL script. Include an output (or partial output if an output does not fit on a single page) screenshot, and* ***a paragraph*** *on query purpose and returned output.*

Given that there were so many tweets (n = 7,491) without mention of specific candidates (see Figure 7) a natural question is “what topics were the candidates talking about?” Figure 11 lists the debate topics based on tweets. Results reveal several “hot button” issues at the time including abortion, gun control, health care and immigration. At first sight this may indicate an interesting “on topic” debate.

A picture containing text

Description automatically generated

Graphical user interface, application

Description automatically generated

Figure 11: Debate topics in the FIRST\_GOP\_DEBATE tweet data set

The next query (Figure 12) investigates the number of tweets for each of the subject matter areas in figure 11. The vast majority of tweets were in none of the topic areas (n = 8148; 59%). This is curious as the topic areas (as mentioned in the prior query) are “hot button” issues at the time. So why are almost 60% of the tweets not on these issues?

Text

Description automatically generated with medium confidence

Text

Description automatically generated with low confidence

Figure 12: Code and results for number of tweets per topic area in the FIRST\_GOP\_DEBATE data set

Perhaps it is because of the next highest number of tweets, were also not on the topic areas. The second most frequent tweets were directed toward the debate moderaters (FOX News or Moderators; n = 2900; 21%). To include tweets such as “Why doesn’t Chris Wallace ask the politicians about their finances and where their money comes from?” and “Fox is cherry picking the candidates.”

When combining all other topics (religion, foreign policy, women’s issues, rocial issues, abortion, jobs and ecomony, immigration, LGBT issues, Healthcare and Gun control) only 18% (n = 4,297) of the total number of tweets were counted! This is a very interesting as it may help validate that the moderators were not controlling the debate, as they became part of the most talked about topic.

For those voters watching the debate to learn about candidate views on important topics, they would have been dissapointed. Although such topics were brought up, there were few tweets on them. This could also indicate that the candidates were in agreement on issues such as gun control, hence the topic did not take much time to discuss and therefore not many tweets were written about them. It could also mean that Mr. Trump dominanted the discussion and was not talking “on topic.” The next query examined this further.

Figure 13 shows the number of tweets per topic area by candidate. Results for each candidate can be seen in Appendix B.

Text

Description automatically generated

Figure 13: SQL query generating tweets by topic area for each candidate

Figure 14 shows results by subject matter for Mr. Trump. The majority of tweets related to Mr. Trump were not on any of the topics of the debate (n = 1604, 57%). Next most frequent tweets referenced Mr. Trump and FOX News or Moderators (n = 849, 30%). Collectively, therefore, tweets referencing Mr. Trump were not about the topic areas of the debate 87% of the time!

Tweets referencing Mr. Trump and a topic of the debate were most frequent on Womens Issues that were not abortion related (n = 116) and least frequently on LGBT issues (n = 1).

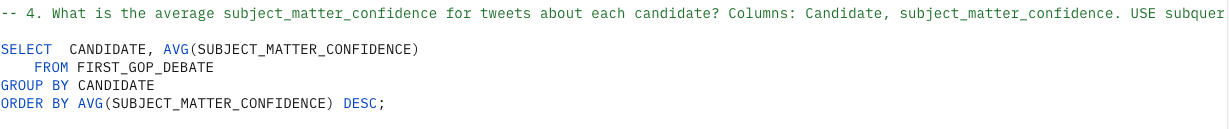
Application, table

Description automatically generated

Figure 14: Candidate Donald Trump, total tweets per subject matter and by ercentage

Subject\_matter\_confidence is a score generated by the observers regarding how confident they perceived the candidate on the subject matter in the tweet. In trying to understand Mr. Trumps unexpected win, a sub-question relates to the percived confidence that the candidate portrayed and as demonstrated in the tweets. When examining the confidence of the subject matter on a scale of 0 (no confident) and 1 (completely confident) the column “subject\_matter\_confidence” was examined (Figure 15).

Mr. Bush was seen as the most confident followed closely by Mr Cruz. Then, Mr. Kasich and Mr. Trump. The least confident of the subject matter was Mr. Huckabee. According to the Washington Post, of these top four confident candidates, only Mr. Kasich was seen as a ‘winner’ of the debate, and only for the first hour. This would seem to indicate that confidence may not play a role in being perceived as a winner of this debate.



Table

Description automatically generated

Figure 15: Average subject matter confidence for each candidate

To explore this topic further a query was designed to determine if those candidates who were seen as most confident lead to greater retweets. Apparently not! According to Figure 16, the number of retweets that were not about a candidate were at least three times more frequent than retweets about any specific candidate. Furthermore, although Mr. Trump was only fourth on the list of subject matter confidence, he got more than double the amount of retweets. This indicates either the amount of interest in Mr. Trump and/or Mr. Trumps ability to garner attention during the debate. Could it be that any attention is, in fact, good attention?

Graphical user interface, text

Description automatically generated with medium confidence

Application

Description automatically generated with low confidence

Figure 16: Retweets per candidate

Furthermore, tweets don’t need to be positive to be retweeted. In fact, most of the retweets were of negative sentiments (66%; n = 419,335; Figure 17).

Graphical user interface, application

Description automatically generated

Figure 17: Number of retweets by sentiment

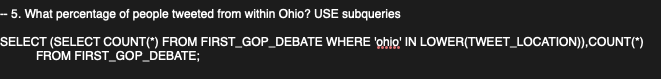
The most retweeted negative tweets were seemingly ridiculous claims (see Figure 18). Indicating perhaps a desire for attention rather than true debate.

Text

Description automatically generated

Figure 18: Negative retweets over 100 times

As this was the first debate, one may argue that not many people were paying attention outside of Ohio, where the debate was held. If so, perhaps the results of this analysis could not be generalized across the nation. A final query was run to determine the number of tweets that were generated from inside of Ohio (Figure 19). Results revealed that only 32 tweets were generated from within Ohio (Figure 19). This is 0.2 percent of all tweets. Such a result indicates that the debate garnered attention nationwide and perhaps set the stage for future debates and a momentum gethering run for Mr. Trump.



Background pattern

Description automatically generated with low confidence

Figure 19: SQL Query and resulting output of number of tweets generated in Ohio

1. **Experimental Design & Statistical Methodology**

Three methodologies were chosen to analyze the data. The rational for these methodologies (as previously discussed in the introduction) are that they are best suited to the purpose of the analysis which is to help explain the unexpected success of Mr. Trump.

Word clouds were used to determine word frequencies. The statistical package used for this analysis was “wordcloud” (R. Documentation, n.d.). Initial statistical parameters for the word cloud were set to a minimum frequency of a word (n = 35, Figure 20), maximum number of words to be plotted was set to 100 (R. Documentation, n.d.). This allowed less frequent words to be dropped. Words were set to a 90-degree rotation for ease of reading (rot.per). The words were scaled to a vector length of 0.9 with words that are most frequent having different colors (color=dark2).A picture containing website

Description automatically generated

Figure 20: Statistical parameters for word cloud analysis

Trend analysis was used to examine the frequency of sentiments before, during, and after the debate. The statistical package installed for this purpose was ggplot (R. Documentation.b, n.d.).This package is used to plot sentiments (positive, neutral, negative) over time (dates occuring before, during and after the debate, Figure 21). The geom\_line connects the sentiment across dates on the x.axis. The y axis is set to ‘count’ or frequency of the sentiment on a particular date. The plot is grouped by sentiment (i.e. positive, neutral, negative).



Figure 21: Statistical parameters for trend analysis

The *Apriori rules algorithm* was used to learn associations between frequently occuring items. The variables used to generate rules are seen in figure 22. The statistical package used for this analysis is ‘arules’ (Cran-R, n.d.).



Figure 22: Statistical input variables for Apriori Rules generation

The Apriori principle is based on how item sets are generated. Apriori algorithm is based on the frequency (or count) of the item set and the principle is “If an itemset is frequent, then all of its subsets are frequent (Han, Kamber, Pei, 2011).” This is important because it improves the performance and efficiency of the model by reducing the number of transactions needed to form item sets. The algorithm intuits that if the itemset is frequent then it does not need to go and determine if all similar (subset) items are also frequent.

The Apriori algorithm is based on an association rule, that is, an implication of an expression (Han, Kamber, Pei, 2011). Together, if the frequency of variables occur often, then they form an itemset, for example, trump and negative, *may*, logically form an item set.

The logic behind the formula is in the form of an antecedent (right hand side: rhs) and consequence (left hand side: lhs). The antecedent is the “before” part of the equation, essentially stating the “if” this. The consequence is the “after,” or other half of the statement and says, “then the probability of that.” For example, *if* Trump is mentioned in the tweets, *then* there is some probability that the tweet is also negative. We could assume that Trump is the antecedent, could adversely affect tweet sentiment. However, we cannot infer any causality or prediction from a Apriori model. We can only say that a group of variables frequently occur together.

When put together, the antecedent and consequence form the Apriori algorithm and an association rule. The association uncovers relationships, represented by the frequency of item sets occurring together. The relationships can be expressed as such:

Consequent (rhs)

Antecedent (lhs)

*X -> Y*; where X and Y are item sets.

Associations are then generated as rules. The strength of the association is measured in terms of support and confidence and form part of the key input parameters for the Apriori model.

Support is how often the rule is applied to a specific data set. Whereas confidence determines how frequently items in Y appear in transactions with X (Han, Kamber, Pei, 2011).

Diagram

Description automatically generated

U = union, σ = count of the number of transactions where this union exists.

Length is also a key input parameter. Length of the rule is the number of items in the left-hand side (lhs) plus those on the right-hand side (rhs). Parameters were set to a minimum length of 1 and maximum of 10.

Rule length = lhs + rhs

Lift is another important metric is deciphering the relationship rule. Lift is the proportion of the rows of data that meet the condition on both sides of the equation (X and Y). The higher the lift the stronger the relationship between the lhs and rhs. A lift of 1 means the antecedent and consequent are independent of one another.

1. **Data Pre-Processing Steps**

Various data pre-processing steps were conducted prior to analysis. Pre-processing differed depending on the analytical method.

*Word cloud data pre-processing* included the following steps (see HTML file; Hunfalvay\_Assignment\_3.R):

1. Install tm, wordcloud and Snowball IC for stemming. These R statistical packages are needed to process the following steps for text mining and developing word clouds.
2. Take a grouping of text (positive) and make it a vector source which is made for working with character objects in R. Vector source takes the group of texts and makes each element of the resulting vector a document within the R workspace.
3. Remove the white space. This is done to separate the text within the tweets in order to better assist the analytical process by creating independent objects.
4. Remove URL’s. As the URLs are not additive to the word cloud analysis they are removed.
5. Remove non-ASCII characters. ASCII is the American Standard Code for Information Interchange and is the most common character encoding format for text data (Loshin, n.d.). Any characters that are not ASCII are removed such as Chinese characters in order to clean the data.
6. Remove punctionation in a similar vein to removal of non-ASCII characters, punctuation is removed as it is not addative to the text analysis.
7. Remove numbers as again they are not addative to the text analysis.
8. Transform toSpace is created using the tm\_map() function to replace special characters in the text such as &, $, # with space.
9. Remove stop words. Stop words are commonly occuring words such as ‘as’, ‘and’, ‘the’ that are useless for classification and if not removed can degrade the performance of the analytical model (Bramer, 2013).
10. stemDocument. Text stemming is used to reduce words to their root form. The purpose of this is to not over frequent words with the same meaning such as “happy”, “happier”, “happiest”.
11. Remove white space that occurs during the prior-preprocessing steps.
12. Covert the file into a matrix. The document matrix is a table that contains the frequency of words. Column names are words. Rows are the documents. Cells are numbers showing the frequency of a word within a document (see Figure 23)

Table

Description automatically generated

Figure 23: Shows the first 10 rows and first 10 columns of the matrix *prior* to counting frequency of terms

1. Set the conditions for word cloud:
   1. Words must appear at least 35 times (this was iterated upon, see results section)
   2. Remove sparse term: maximum of 60% empty space
   3. Find the word frequencies
   4. Determine the color scheme
   5. Set the minimum and maximum number of words in the word cloud (minimum = 35, maximum = 100). This was iterated on see the results section.
   6. Set the rot.per which is the proportion of words with 90 degree rotation
   7. Scale =c(0.9,0.9) which is the vector length of the word size. This was iterated upon, see results section.

*Trend analysis* required no further steps in *data pre-processing.*

*Apriori rules data pre-processing* included the following steps (see HTML file; Hunfalvay\_Assignment\_3.R):

1. Numeric variables were descretized and made into factors to be used in the model analysis. Retweet\_count, candidate\_confidence, sentiment\_confidence and subject\_matter\_confidence was all discretized into three groups Candidate, sentiment, and subject\_matter, and tweet\_location variables were changed to factors in preparation for analysis (Figure 24).

Text

Description automatically generated

Figure 24: Data Pre-processing Steps for Apriori rules analysis

1. Run the Apriori algorithm with the default arguments. Default parameters are confidence of 80%, support of 10%. Minimum length of 1 and maximum length of 10.
2. Appearance, by default shows all item sets. Inspect first 15 rules (Figure 25).

Table

Description automatically generated

Figure 25: First 15 rules after running the Apriori algorithm with default settings

1. Show a summary of the rules for inspection (Figure 26).

A screenshot of a computer

Description automatically generated with medium confidence

Figure 26: Shows summary rules for the default settings

Inspect the rules generated from the algorithm (Figure 25 and 26). When inspecting the rules, the first rule has a maximum length of 1. As we are looking for associations, we need to prune this parameter and only show rules that have at least 2 items. This was conducted (see figure 27 for code), then the rules were reinspected (see Figure 28 and 29). Note that the first four rules in the summary command now show two or more rules.

Text

Description automatically generated

Figure 27: Code for removing rules that only have one item

Table

Description automatically generated

Figure 28: Inpecting the first 15 rules after removing rules with only one item

A screenshot of a computer

Description automatically generated with medium confidence

Figure 29: Summary of rules after removing rules with only one item

1. Remove any redundant rules and inspect (Figure 30).

Table

Description automatically generated

Figure 30: Shows first 15 rules after redundant rules were removed

1. Show summary. Note the rules have now reduced from 52 in figure 26, to 30, in figure 31.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 31: Shows a summary of the rules after the redundant rules are removed.

1. Sort rules by lift and inspect

Once these pre-processing steps are completed for Apriori rules methods further analysis and visualization of the data is conducted (see next section on data analysis). These steps were iterating on many times to determine the best set of results.

**c. Data Analysis**

*Word Cloud Analysis*

Using the word cloud analysis, results revealed that two terms appeared more frequently than others throughout the twitter content (Figure 21). “gopdeb” and “trump” were used most frequently. Regarding Mr. Trump, this is consistent with the prior SQL queries showing he garnered the most amount of attention compared with other canddiates during this debate and showing further evidence that “any attention may be good attention.”

The “gopdeb” is an interesting insight and may help to explain the number of tweets that were plentiful yet neutral (n = 2087, see Figure 8). This may indicate the level of interest in the debate.

To further explore the data various parameters were examined. Figures 32-35 show adjustments in the minimum and maximum word values. This is helpful in understanding a) more effective ranges in word frequency, b) a further visual illustration of the final remaining words, c) as words reduce some limitations in the visualization can also be seen as there is less “clutter” in the visual presentation (limitations are discussed in the next section).

Text

Description automatically generated

Figure 32: Word cloud analysis results. Parameters minimum words 35, maximum words 100

Further analysis of the word cloud shows other terms that occurred frequently. One example being the Fox Moderator Megyan Kelli. This is consistent with Figures 9 and 10 which show that Fox and moderators became part of the talking points on twitter. With 2900 tweets the the moderator was higher than any one candidate including Mr. Trump!

A picture containing timeline

Description automatically generated

Figure 33: Word cloud analysis results. Parameters minimum words 100, maximum words 1000

As parameters of frequency are further adjusted to include a minimum of 200 words (Figure 34) only 5 terms remain. Two of the five terms relate to Mr. Trump, again showing he garnered a high level of attention. The general interest in the debate remains with two common terms: gopdeb and debat. The fifth word is rwsurfergirl who has a large following on Twitter for conservative issues (see: <https://twitter.com/hashtag/RWsurfergirl?src=hashtag_click>).

A picture containing timeline

Description automatically generated

Figure 34: Word cloud analysis results. Parameters minimum words 200, maximum words 1000

Final words remaining with more than 400 tweets are Trump and gopdebate (figure 35). Therefore, the word cloud along with SQL queries have answered the following specific analytical questions:

2. Who got the most attention from the debate? Answer: The GOP and Mr. Trump

Chart

Description automatically generated with medium confidence

Figure 35: Word cloud analysis results. Parameters minimum words 400, maximum words 1000

*Trend Analysis*

The trend analysis was developed to show the number of posts of each sentiment over time. The purpose of this analysis is to look across time, and not just at the day of the debate to determine the sustainability of sentiments. Given that the debate was of interest to people throughout the United States (see Figure 19) this information may influence future debates should sentiments persist. In turn, such an analysis may shed light on the primary analytical question, which is to help explain the unexpected success of Mr. Trump.

The negative sentiment was sustained across time to include before and after the debate (Figure 36). This could indicate a general displeasure, not just with the debate and the candidates but with the state of affairs in the country at the time and/or the GOP. The trend analysis results show that across the timeframe of the debate, irrespective of topic, people showed negative sentiments before, after, and during the debate.

In trying to understand Mr. Trumps eventual win, perhaps Mr. Trump was relatable to the general public as the sentiments generated by tweets occuring with Mr. Trump were mostly negative (n = 1756) out of a total of 2813 (see Figure 8). It may also be true that Mr. Trump fueled the negative sentiment and in turn “activated” people to become engaged in voting and, in the political process. This may help to explain the first sub-question of the analysis “What types of sentiments resonated with the audience? Did Mr. Trump use those sentiments?”

Chart, line chart

Description automatically generated

Figure 36: Trend analysis of tweets by sentiment across time.

*Apriori Rule Analysis*

The Apriori rule analysis is used to learn associations between frequently occuring items. The rules from the analysis are shown in Figure 37. Note that RETWEET\_COUNT = 0 is the first descritized bucket of tweets and ranges from no tweets (0) to 100 tweets.

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

Figure 37: Shows rules generated from the Apriori method of analysis

Rules are examined in the follwing manner; *rule 1 & 2:* Moderate sentiment confidence (between 0.185 and 0.6628) coupled with both moderate subject matter confidence (0.221-0.662) and no retweets (=lhs) was an antecedent for low candidate confidence (0 – 0.49; rhs). Rule 2 is the same as rule one without the retweets. Interpretation of Rule 1 and 2 is that moderate confidence regarding a sentiment (positive, negative, neutral) and moderate confidence on subject matter preceeds a low level of candidate confidence.

Lift is another important metric is deciphering the relationship rule (Figure 38). Lift is the proportion of the rows of data that meet the condition on both sides of the equation (X and Y). The higher the lift the stronger the relationship between the lhs and rhs. A lift of 1 means the antecedent and consequent are independent of one another. Rule 1 has the highest lift value (3.92, see Figure 38). Rule 2 also has a very high lift (3.90).

Diagram

Description automatically generated

Figure 38: 3D Matrix of the rules showing those with the highest lift

Relationships are *not* causal. Instead they show items from the left hand side (lhs) that preceed, that is, are antecedents to, items of the right hand side (rhs). Therefore, when viewing the first six rules generated from the apriori analysis method three variables are intertwined: candidate confidence, sentiment confidence and subject matter confidence. These variables are further explained using the parallel coordinates plot (line 1, Figure 39) whereby retweet-count=0 (3rd position), subject matter confidence (0.221-0.662; 2nd position) sentiment confidence (0.183-0.6628, 1st position) are antecedents to candidate confidence (0-0.49; lhs). This is seen as a strong relationship due to the red coloring in the plot (Figure 39).

Chart, line chart

Description automatically generated

Line 1

Figure 39: Coodinates plot showing the lhs to rhs strongest relationships

In rules 4, 5, and 6, low candidate confidence was an antecedent to moderate subject matter confidence with a very high level of confidence in the results (Figure 40). Interestingly, no rules were generated for high confidence.

Chart, scatter chart

Description automatically generated

Figure 40: Scatterplot of rules by confidence, support and lift value.

The two-key plot (Figure 41) shows a high level of confidence in items that were ordered 3rd in the rule, yet a high level of support for items second in the rule. Support is how often the rule is applied to a specific data set. Whereas confidence determines how frequently items in Y appear in transactions with X (Han, Kamber, Pei, 2011). Both support and confidence are important factors in understanding the FIRST\_GOP\_DEBATE dataset.

Chart, scatter chart

Description automatically generated

Figure 41: Order, support and confidence two-key plot

When digging deeper into the data using a relationship rules map (Figure 42) high sentiment confidence and the candidate Donald Trump are coupled closely together. The relationship is indirect and weak however. Sentiment confidence is defined as “confidence in the sentiment”, that is positive, negative or neutral. Perhaps people felt that although Mr. Trump was mostly negative in his sentiments, they knew what they were getting?

Diagram

Description automatically generated

Figure 42: Relationship map

Rule 7 is interesting and shows when no candidate is mentioned and the subject matter of the tweet is Fox News or Moderators this is an antecedent to a negative sentiment. Perhaps the audience did not like something about the moderators or, did not like the way the moderators were controlling (or not controlling) the debate.

Rule 24 and 26 mention Mr Trump specifically (Figure 43). Rule 24 states that the candidate Donald Trump coupled with a negative senitment was an antecendent to retweeting up to 100 times. Rule 24 is consistent with Figure 8 showing that tweets related with Mr. Trump were negative *and* were an antecedent to retweeting. This is further evidence that Mr. Trump received the most attention of any of the candidates and that attention was negative.

A picture containing bar chart

Description automatically generated

Figure 43: Shows items in the LHS group via support and confidence metrics.

Rule 26 states that Mr Trump himself with no other factors also was an antecedent to retweeting up to 100 times. Rule 26 is yet another indicator that Mr Trump got the most attention.

Various other potentially important statistics can be used to examine results and include:

* 1. *Chi Squared:* is a hypothesis test for statistically significant relationship between nomimal and ordinal variables and is used to test whether the variables are independent of one another (Brin et al., 1997a). The critical value for α=0.05 is 3.84; higher chi-squared values indicate that the null-hypothesis of independence between lhs and the rhs should be rejected (i.e., the rule is not spurious). Larger chi-squared values indicate stronger evidence that the rule represents a strong relationship. The statistic can be converted into a p-value using the χ2 distribution. Chi squared ranges from 0 to infinity. It is calculated by the following formula:

Table

Description automatically generated

* 1. *Conviction:* is an alternative measure to confidence as confidence was found to not capture the direction of associations adequately. Conviction compares the probability that X appears with Y if they were independent on the actual frequency of the appearance of X without Y. It is similar to lift except it is a directional measure. Conviction ranges from 0 to infinity and 1 indicates independence, rules that always hold have a value of infinity. Conviction is calculated by the following formula (Brin et al, 1997b) and where  is the event that Y does not appear in the transacation.

Diagram

Description automatically generated with medium confidence

* 1. *Cosine:* is the null-invariant measure of correlation betweent he items in X and Y. It ranges from 0-1 and 0.5 means no correlation. It is measured using this formula (Hahsler, n.d.):

Text

Description automatically generated

* 1. *Coverage:* also known as lhs support, measures the probability that a rule X -> Y and applies random selection transaction and therefore is often called the antecedent support of LHS support ((Hahsler, n.d.). It ranges from 0 to 1 and is calculated using the following formula:



* 1. *Leverage:* measures the difference of X and Y appearing together in the data set and what would be expected if X and Y were statistically dependent. It ranges from -1 to 1 and 0 indicates independence (Platesky-Shapiro, 1991). Leverage is calculated using the following formula:



* 1. *Odds Ratio:* is a measure of the relationship between two binary variables. It is the ratio of the odds of a transaction containing Y in the groups of transactions that do and do not contain X. It ranges from 0 to infinity, whereby 1 indicates that Y is not associated with X. It is calculated using the following formula (Tan et al., 2004):

Letter

Description automatically generated with low confidence

Results utilizing these additional metrics are seen in Figure 44. According to the Chi Squared metric 1, 2, 3, 4, 5, 6 are statistically significant. This indicates stronger evidence that the rule represents a strong relationship. Rules 3 and 4 have a low correlation with a cosine of close to 0.5. Of the first 6 rules, rule 6 is the most highly correlated at 0.79. Values for the first 6 rules for coverage are low. Leverage for the first 6 rules ranges from 0.09 to 0.1 indicating a high level of independence between X and Y. Furthermore, this is validated by the odds ratio which indicates Y is not associated with X as values are greater than 1.

Given these results, low candidate confidence, moderate subject matter confidence and moderate sentiment confidence are strongly (and independently) related. In plain language, if a candidate is moderately confident in their subject matter and the sentiment perceived is also moderate, then they are perceived as having low (candidate) confidence.

This *may* be interpreted as anything other than a high level of percieved subject matter confidence means the candidate is seen as having low confidence. Unfortunately we can not directly say that a high level of subject matter confidence was an antecedent to high candidate confidence as no rule was generated relating to high confidence. Nevertheless, we can make strong associations regarding low levels of expertise and confidence.

Text

Description automatically generated with medium confidence

Figure 44: Shows results of the 30 rules using additional metrics for evaluation

**d. Discussion**

The *purpose* of this analysis was to find associations within the FIRST\_GOP\_DEBATE dataset that may help explain the unexpected success of Mr. Trump in the 2016 election. Specifically, to answer the following questions from the dataset:

*Question 1:* What types of sentiments resonated with the audience? Did Mr. Trump use those sentiments?

*Answer 1:* The overwhelming majority of sentiments were negative (62%). Tweets associated with Mr. Trump were also overwhelmingly negative (n = 1758, 62%).

*Question 2:* Which candidate garnered the most attention from the debate?

*Answer 2:* Throughout several analysis Mr. Trump garnered the most attention compared with any other candidate as evidenced by the number of tweets and number of retweets.

*Question 3:* Does subject matter confidence relate to candidate confidence?

*Answer 3:* Low levels of subject matter expertise formed a strong antecedent to low candidate confidence. This result was strong.

It seems high confidence *may* be associated with Mr. Trump via relationship mapping (Figure 24). However, the relationship is indirect and weak at best. Nevertheless, some political pundits stated that “Trump was clearly the most confident” (Neffinger in Politico, 2015).

Rules from the Apriori analysis show sentiment confidence, subject matter confidence and candidate confidence as elements of rules.

**Limitiations**

There are several limitations to the analysis. First, the data is from only one debate and there were many debates for the GOP primary. Second, the data only looks at Twitter responses which tends to be a certain demographic. According to Pew Research (Odabas, 2022) this includes people under 30 with a minority of twitter users producing the majority of tweets (for example: swsurfergirl). Such data can skew results.

Limitations in the analysis are also evident. First, the analysis is not causal so any outcome can not be stated as “This factor *caused* that outcome.” We can only infer.

Furthermore, it is impossible to directly address a candidates “credibility” which is an important factor in evaluating the debate. Input variables indirectly examine confidence via tweets but confidence and credibility, although related are also different. For instance, a person with high confidence, does not mean they are also highly credible.

Specific limitaions in the word cloud include terms that were related being categorized as different. For example, Megynkelli and megyn; Gopdeb and gop time and debate group (see figure 32 and 33). Additional cleaning of stem words could be used to overcome this limitation. Futhermore, certain nonsensical words were included such as ‘amp”, ‘tcot’. This causes clutter and reduces the understanding of the word cloud.

Limitations in the Apriori algorithm were found when trying to examine rules that targeted “donald trump”. No rules resulted. Furthermore, the 30 rules that did result were somewhat difficult to interpret. For instance, we can not make any associations of high candidate confidence and high subject matter or sentiment confidence. Many different input variables, input parameters and visualizations were experimented with but the method failed to produce rules with more clarity.

Future analysis should consider more data sets from different debates and from different social media tools. Causal analysis such as a regression to determine weighted factors may be interesting in trying to uncover important elements in Mr. Trumps unexpected win.

In conclusion, Mr. Trump was, by all standards, an unconventional candidate. The “newness” of his style caused many people to find it difficult to understand his impact and predict his success. Some experts in the political field were woefully incorrect in their predictions, as reported by Politico on August 7th, 2015 Neffingger, J. (2015). What is clear from this analysis however, is that:

1. Mr. Trump garnered the most attention
2. The vast majority of tweets were negative
3. Tweets associated with Mr Trump were also negative
4. The debate had national interest
5. Topic areas of the debate were infrequently discussed (only 18% of the time).
6. More than 50% of the tweets were not associated with any of the candidates
7. FOX News and the moderators garnered a lot of attention during the debate.

**e. References**

Alida. (2012). *The Pros and Cons of Word Clouds as Visualizations.* Retrieved on July 4th, 2022, from: <https://www.alida.com/the-alida-journal/pros-and-cons-word-clouds-visualizations>

Bramer, M. (2013). *Principles of Data Mining.* 2nd Ed. Springer, New York, NY.

Brin, S., Motwani, R. & Silverstein, C. (1997a). Beyond market baskets: Generalizing association rules to correlations. In *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data,* pages 265-276, Tucson, Arizona, USA, May 1997.

Brin, S., Motwani, R., Ullman, J.D., & Turk, S. Dynamic itemset counting and implication rules for market basket data. In *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data,* pages 255-264, Tucson, Arizona, USA, May 1997.

Cillizza, C. (2015). Winners and losers from the first Republican presidential debate. *The Washington Post.* Retrieved on July 16th from: <https://www.washingtonpost.com/news/the-fix/wp/2015/08/06/winners-and-losers-from-the-first-republican-presidential-debate/>

Cambria, E., Das, D., Bandyopadhyay, S., Feraco, A. (2017). *A Practical Guide to Sentiment Analysis.* Springer, New York, NY.

Cran-R (n.d.). arules: Mining association rules and frequent itemsets. Retrieved on July 16th, 2022 from: [https://cran.r-project.org/](https://cran.r-project.org/web/packages/arules/index.html#:~:text=arules%3A%20Mining%20Association%20Rules%20and,mining%20algorithms%20Apriori%20and%20Eclat)

Flores, R. (2016). *In a Stunner, Donald Trump Wins the Presidency.* Retrieved on July 16th, 2022, from: <https://www.cbsnews.com/news/in-a-surpising-upset-donald-trump-wins-the-presidency/>

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts and Technique, 3rd Ed.* Ch.3. Elsevier Science. New York, NY.

Hasler, M. (n.d.). A probabilistic comparison of commonly used interest measures for association rules. Retrieved on July 16th, 2022 from: <https://mhahsler.github.io/arules/docs/measures>

IBM (2022). *Big Data.* Retrieved on July 16 from: <https://cognitiveclass.ai/courses/what-is-big-data/>

Loskin, P. (n.d.). *ASCII (American Standard Code for Information Interchange).* Retrieved on July 16th, 2022 from: [https://www.techtarget.com/](https://www.techtarget.com/whatis/definition/ASCII-American-Standard-Code-for-Information-Interchange#:~:text=ASCII%20(American%20Standard%20Code%20for%20Information%20Interchange)%20is%20the%20most,additional%20characters%20and%20control%20codes)

Neffingger, J. (2015). The Debates Biggest Winners and Losers.Retrieved on July 16th, 2022 from: <https://www.politico.com/magazine/story/2015/08/republican-debate-2015-winners-losers-121139/>

Odabas, M. (2022). *10 Facts about Americans and Twitter. By Pew Reseach Center.* Retrieved on July 16th, 2022 from: [https://www.pewresearch.org/](https://www.pewresearch.org/fact-tank/2022/05/05/10-facts-about-americans-and-twitter/#:~:text=Also%2C%20U.S.%20adults%20under%2030,the%20vast%20majority%20of%20tweets).

 Piatetsky-Shapiro, G., (1991). Discovery, analysis, and presentation of strong rules. *Knowledge Discovery in Databases*, p. 229-248.

R Documentation (n.d.(a)). *Wordcloud: Plot a word cloud.* Retrieved on July 16th from: <https://www.rdocumentation.org/packages/wordcloud/versions/2.6/topics/wordcloud>

R Documentation B. (n.d.(b)). *Ggplot.multistats.* Retrieved on July 16th, 2022, from: <https://www.rdocumentation.org/packages/ggplot.multistats/versions/1.0.0>

Tan, P-N., Kumar, V., and Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems, 29(4):*293–313.

**Appendix**

**Appendix A**

**Additional Analytical Processes**

> # Terms Cluster dendogram

> library("cluster")

> dtms <- removeSparseTerms(dtm, 0.98) #Remove sparse terms

> d <- dist(t(dtms), method="euclidian") #Build the dissimilarity matrix

> fit <- hclust(d=d, method="ward.D2")

> plot(fit, hang=-1)

A picture containing diagram

Description automatically generated

|  |
| --- |
| > #Network of terms  > library ("igraph")  Attaching package: ‘igraph’  The following object is masked from ‘package:arules’:  union  The following objects are masked from ‘package:dplyr’:  as\_data\_frame, groups, union  The following objects are masked from ‘package:stats’:  decompose, spectrum  The following object is masked from ‘package:base’:  union  > tdm<-TermDocumentMatrix(docs) # Term document matrix  > tdm <- removeSparseTerms(tdm, 0.96) # Remove sparse terms  > termDocMatrix <- as.matrix(tdm) # Convert tdm to matrix  > termDocMatrix[termDocMatrix>=1] <- 1 # Set non-zero entries to 1 (1=term present, 0=term absent)  > termMatrix <- termDocMatrix %\*% t(termDocMatrix)  > View (termMatrix) |
|  |
| |  | | --- | | > | |

Table

Description automatically generated

g <- graph.adjacency(termMatrix, weighted=T, mode="undirected")

> g <- simplify(g) # Remove the self-relationships

> # V(g) is a graph vertex

> V(g)$label <- V(g)$name # Label each vertex with a term

> V(g)$degree <- degree(g)

> set.seed(3952)

> plot(g, layout=layout.fruchterman.reingold(g), vertex.color="cyan")

Diagram

Description automatically generated

|  |
| --- |
| plot(g, layout=layout\_with\_gem(g), vertex.color="pink") |
|  |
| |  | | --- | | > | |

Diagram

Description automatically generated

plot(g, layout=layout\_as\_star(g), vertex.color="yellow", vertex.shape="square")

Diagram

Description automatically generated with low confidence

plot(g, layout=layout\_on\_sphere(g), vertex.color="magenta")

Chart, diagram

Description automatically generated

plot(g, layout=layout\_randomly(g), vertex.size=10)

A map of a city

Description automatically generated with low confidence

plot(g, layout=layout\_in\_circle(g), vertex.color="pink", vertex.size=35)

Diagram

Description automatically generated

plot(g, layout=layout\_nicely(g), vertex.color="plum", vertex.size=25)

Chart, bubble chart

Description automatically generated

plot(g, layout=layout\_on\_grid(g), vertex.color="green", vertex.size=20)

Chart, bubble chart

Description automatically generated

plot(g, layout=layout\_as\_tree(g), vertex.color="brown", vertex.size=20)

Shape

Description automatically generated

> cl <- maximal.cliques(g)

> cl

[[1]]

+ 13/25 vertices, named, from 33f3542:

[1] bush gopdeb gop carson job candid cruz amp trump debat foxnew great night

[[2]]

+ 13/25 vertices, named, from 33f3542:

[1] bush gopdeb gop carson job candid cruz amp trump debat rubio great night

[[3]]

+ 12/25 vertices, named, from 33f3542:

[1] bush gopdeb gop carson job candid cruz amp trump

[10] debat rubio rwsurfergirl

[[4]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] watch rate candid realdonaldtrump megynkelli

[[5]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] watch rate candid realdonaldtrump carlyfiorina

[[6]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch rate candid carson

[[7]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] watch rate news megynkelli realdonaldtrump

[[8]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch

[10] tedcruz cruz love carlyfiorina

[[9]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch tedcruz cruz love carson

[[10]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch tedcruz cruz love news

[[11]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch

[10] tedcruz cruz candid carlyfiorina

[[12]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch tedcruz cruz candid carson

[[13]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch tedcruz question love carson

[[14]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] watch tedcruz question love realdonaldtrump carlyfiorina

[[15]]

+ 15/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] watch tedcruz question love realdonaldtrump news megynkelli

[[16]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew watch tedcruz question candid carson

[[17]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] watch tedcruz question candid realdonaldtrump megynkelli

[[18]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] watch tedcruz question candid realdonaldtrump carlyfiorina

[[19]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew job cruz carson candid

[[20]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew job cruz carson love

[[21]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew job cruz news love

[[22]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] job question megynkelli realdonaldtrump candid

[[23]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew

[9] job question megynkelli realdonaldtrump love news

[[24]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew job question carson candid

[[25]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat fox foxnew job question carson love

[[26]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch

[9] foxnew rate candid realdonaldtrump megynkelli

[[27]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch

[9] foxnew rate candid realdonaldtrump carlyfiorina

[[28]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew rate candid carson

[[29]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch

[9] foxnew rate news megynkelli realdonaldtrump

[[30]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew tedcruz cruz love news

[[31]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew tedcruz cruz love carson

[[32]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew

[10] tedcruz cruz love carlyfiorina

[[33]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew tedcruz cruz candid carson

[[34]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew

[10] tedcruz cruz candid carlyfiorina

[[35]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew tedcruz question love carson

[[36]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch

[9] foxnew tedcruz question love realdonaldtrump carlyfiorina

[[37]]

+ 15/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch

[9] foxnew tedcruz question love realdonaldtrump news megynkelli

[[38]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch foxnew tedcruz question candid carson

[[39]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch

[9] foxnew tedcruz question candid realdonaldtrump megynkelli

[[40]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch

[9] foxnew tedcruz question candid realdonaldtrump carlyfiorina

[[41]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch rubio carson candid cruz

[[42]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch rubio carson candid question

[[43]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch rubio carson candid rate

[[44]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch rubio carson love question

[[45]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop watch rubio carson love cruz

[[46]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job foxnew cruz carson love

[[47]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job foxnew cruz news love

[[48]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job foxnew question carson candid

[[49]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job foxnew question carson love

[[50]]

+ 13/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job

[9] foxnew question megynkelli realdonaldtrump candid

[[51]]

+ 14/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job

[9] foxnew question megynkelli realdonaldtrump love news

[[52]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job rubio carson cruz love

[[53]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job rubio carson question love

[[54]]

+ 12/25 vertices, named, from 33f3542:

[1] night gopdeb great amp trump debat gop job rubio carson question candid

[[55]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch candid

[9] realdonaldtrump rate

[[56]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch candid

[9] realdonaldtrump question tedcruz

[[57]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch candid carson

[10] rate

[[58]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch candid carson

[10] tedcruz question

[[59]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch candid carson

[10] tedcruz cruz

[[60]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch news

[9] rate realdonaldtrump

[[61]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch news tedcruz

[10] cruz

[[62]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox watch news

[9] tedcruz realdonaldtrump question

[[63]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox job candid

[9] realdonaldtrump question

[[64]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox job candid carson

[10] cruz

[[65]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox job candid carson

[10] question

[[66]]

+ 9/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox job news cruz

[[67]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl fox job news

[9] question realdonaldtrump

[[68]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch candid

[9] realdonaldtrump tedcruz question

[[69]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch candid

[9] realdonaldtrump rate

[[70]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch candid carson

[10] tedcruz question

[[71]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch candid carson

[10] tedcruz cruz

[[72]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch candid carson

[10] rubio rate

[[73]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch candid carson

[10] rubio cruz

[[74]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch candid carson

[10] rubio question

[[75]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch news

[9] rate realdonaldtrump

[[76]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch news tedcruz

[10] cruz

[[77]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop watch news

[9] tedcruz realdonaldtrump question

[[78]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop job candid

[9] realdonaldtrump question

[[79]]

+ 11/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop job candid rubio

[10] carson question

[[80]]

+ 9/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop job news cruz

[[81]]

+ 10/25 vertices, named, from 33f3542:

[1] debat gopdeb amp trump rwsurfergirl gop job news

[9] question realdonaldtrump

> colbar <- rainbow(length(cl) + 1)

> for (i in 1:length(cl)) {V(g)[cl[[i]]]$color <- colbar[i+1] }

> plot(g, mark.groups=cl,vertex.size=.3, vertex.label.cex=1.2, edge.color=rgb(.4,.4,0,.3))

> #k-means clustering

> library(cluster)

Chart

Description automatically generated

> dtms <- removeSparseTerms(dtm, 0.99)

> d <- dist(t(dtms), method="euclidian")

> kfit <- kmeans(d, 5)

> clusplot(as.matrix(d), kfit$cluster, color=T, shade=T, labels=2, lines=0)

> kfit

K-means clustering with 5 clusters of sizes 1, 39, 21, 63, 4

Cluster means:

didnt gopdeb line night scottwalk walker car debat fiorina presidenti rate realdonaldtrump trump

1 53.263496 0.00000 53.131911 51.65269 53.16014 53.094256 53.28227 50.03998 52.943366 52.971691 51.12729 48.51804 47.66550

2 9.601234 52.68202 9.635250 15.97513 9.35853 10.025881 13.70726 18.08301 10.507921 10.318418 12.51264 18.93366 25.74531

3 12.642534 52.11833 12.676344 16.77000 12.51323 12.913735 15.14038 19.45861 13.279031 13.243270 14.05009 19.82783 26.87918

4 7.698391 53.13644 7.733253 15.05279 7.39051 8.171701 12.59072 17.23464 9.173836 8.669294 11.60880 18.22693 25.85707

5 19.901401 48.21510 19.989197 23.31134 19.92334 20.134894 22.24814 17.98324 20.457439 19.931489 19.58258 17.10116 20.66117

foxnew tedcruz govmikehuckabe love amp cruz truth vote fox megynkelli question randpaul presid

1 52.54522 51.30302 53.122500 52.53570 52.44998 51.99038 52.73519 53.094256 50.89204 52.68776 51.56549 53.06600 52.64979

2 13.31197 14.99970 10.210337 12.64064 15.93442 13.99361 10.95591 10.668427 14.12895 13.24954 12.81201 10.25567 11.48829

3 14.73201 16.56309 13.337466 14.36632 16.63426 15.15562 14.02648 13.680374 15.01645 14.63425 14.72325 13.18418 14.44738

4 12.14150 14.25180 8.832051 11.34941 14.92906 13.58623 10.08349 9.434676 13.20313 12.03801 12.03758 8.44451 10.69842

5 21.65069 21.98638 20.410287 21.44483 23.63218 21.07719 21.02026 20.557054 19.68042 21.53915 20.36565 20.29210 21.24783

carlyfiorina impress made perform donald peopl thing candid god great elect megyn time

1 52.64029 53.188345 53.047149 53.131911 53.084838 53.056574 53.037722 52.37366 53.216539 52.60228 53.235327 53.301032 52.829916

2 12.95362 9.522983 10.036327 9.470081 10.560833 10.270922 10.419779 14.43739 9.453863 13.64724 9.242119 9.904714 10.906428

3 14.58086 12.540853 12.994270 12.462401 13.671088 13.183835 13.161257 15.61758 12.560604 14.88351 12.385653 12.897881 13.832628

4 11.74299 7.609110 8.205173 7.588689 9.264599 8.497224 8.547908 13.20278 7.527608 12.47202 7.285890 8.000840 9.733017

5 21.82003 19.946467 20.123782 19.965949 20.095963 20.290574 20.372624 22.22838 19.973996 22.04998 19.790095 19.942651 20.526164

watch america job john johnkasich kasich bush huckabe republican rubio win give your

1 52.04805 53.028294 52.20153 53.197744 52.933921 52.69725 52.62129 53.122500 53.254108 51.88449 53.065997 53.169540 51.36146

2 13.08588 10.623078 13.08030 9.276325 10.671646 11.91698 12.04196 10.054465 10.547832 15.38183 10.906191 9.514559 11.30853

3 14.67201 13.665200 14.39774 12.524811 13.705885 14.77521 13.82964 12.944840 13.453749 16.17162 13.768718 12.618529 13.59350

4 11.80999 9.349906 12.58588 7.409216 9.474694 10.92010 11.55564 8.209627 9.073985 14.90744 9.606401 7.582542 10.43544

5 20.63628 20.531567 20.66076 19.898336 20.715540 21.39235 19.97132 20.174819 19.891486 22.14592 20.372386 19.875555 19.24474

chris christi dont gopd set marco nail rand immigr thought realbencarson winner bencarson

1 53.310412 53.122500 52.933921 53.684262 52.67827 53.037722 51.951901 53.150729 52.58327 52.99057 52.77310 52.792045 53.178943

2 9.342847 9.862376 10.376719 9.637367 10.93610 10.147545 10.028324 9.520081 10.36585 10.27070 11.20832 10.982354 9.252815

3 12.457464 12.871640 13.386801 12.601985 13.41730 12.878384 13.461582 12.615572 13.36393 13.04625 14.03646 13.870629 12.421726

4 7.399262 7.966902 9.019108 7.600860 10.58431 8.381271 9.285617 7.521179 8.55101 8.46537 10.13864 9.824567 7.286671

5 19.904047 20.082500 20.084386 19.487806 19.54028 20.327139 18.988419 19.923335 19.84073 20.19148 21.03331 20.696921 19.832918

donaldtrump marcorubio point lead make sens hes close ben carson tonight jebbush moder

1 53.113087 52.905576 53.10367 53.478968 53.000000 53.169540 52.962251 53.169540 52.896125 52.40229 52.839379 53.178943 53.075418

2 9.531263 11.038330 10.03611 9.849173 10.723122 9.242242 10.278951 9.294580 10.337752 13.03725 10.509056 9.366698 9.737113

3 12.591592 13.941226 12.99343 12.812361 13.768101 12.489406 13.434427 12.424840 13.180325 14.67377 13.515538 12.531890 12.618179

4 7.623153 9.914394 8.22705 7.847896 9.500849 7.335472 8.991138 7.356905 9.044907 11.89902 9.271958 7.418013 7.851469

5 19.802330 20.926496 20.19656 19.540423 20.684274 19.891969 20.440588 19.843206 20.539105 21.83681 20.588815 19.857112 19.863004

polit answer won show hillari agre donniewahlberg support ted imwithhuck good gop

1 53.103672 53.113087 52.952809 53.056574 53.150729 53.178943 52.87722 53.282267 52.82992 53.141321 52.64979 52.64979

2 9.755855 10.001441 10.780173 9.888151 10.014548 9.371910 10.10781 10.394490 10.80622 9.440055 11.67526 13.00924

3 12.797089 12.937648 13.729924 12.886266 12.945642 12.493425 13.78932 13.248689 13.66514 12.628225 14.43935 14.55914

4 7.891053 8.203542 9.572408 8.078317 8.167158 7.416733 9.44098 8.571297 9.57385 7.523993 10.62555 11.76049

5 19.953557 20.159944 20.366264 19.893091 20.145436 19.872349 20.74317 20.153381 20.35659 19.915674 21.07867 21.18619

call speak tcot american talk stage leader parti news democrat favorit person paul

1 53.207142 53.188345 52.933921 53.178943 52.364110 53.169540 53.450912 53.216539 51.02940 53.225934 53.02829 53.094256 53.084838

2 9.824925 9.486722 10.447541 9.747614 10.653367 9.501407 10.072667 9.748657 13.76530 9.410886 10.20526 10.137298 9.664509

3 12.866439 12.549544 13.276501 12.806899 13.696492 12.618200 12.982164 12.741002 14.74505 12.542873 12.99021 12.949011 12.714023

4 7.969095 7.576890 8.662774 7.876506 9.235391 7.607081 8.137655 7.879481 12.84603 7.467147 8.27163 8.213295 7.689689

5 19.877869 19.670668 20.321907 20.099513 20.018190 19.914456 19.695910 20.047655 19.77704 19.828847 20.18764 20.205964 19.919087

forward enjoy nomin expos ive head plan wallac countri attack hand brought rais

1 52.76362 52.75415 53.404120 52.69725 53.131911 53.084838 53.366656 51.99038 53.150729 53.188345 53.103672 52.640289 51.28353

2 10.60288 10.67099 9.222724 10.74774 9.763150 9.651137 9.641613 10.22626 9.307271 9.393049 9.566990 9.674369 11.23232

3 14.20350 14.20818 12.388993 13.26495 12.687285 12.777442 12.708578 13.55465 12.486874 12.426668 12.644301 12.823215 13.46328

4 10.05113 10.10399 7.190584 10.40121 7.742625 7.884075 7.678428 9.48242 7.359815 7.465384 7.683752 7.723126 10.35152

5 21.03687 21.02026 19.714731 19.41796 20.102037 19.721486 19.912214 19.08384 19.770237 19.512934 19.780570 19.486389 19.01833

control admit face frontrunn wont rwsurfergirl lrihendri democraticdeb legitim until band rid

1 52.488094 53.553711 51.961524 52.678269 53.572381 46.63690 52.68776 52.829916 52.019227 52.687759 52.71622 52.68776

2 9.759977 9.335152 9.973899 9.461012 9.327308 18.29419 11.14745 10.198590 9.787481 9.320690 10.55030 10.59945

3 12.738796 12.412434 13.429366 12.496446 12.404800 19.32606 14.16377 13.908434 13.283409 12.577401 13.11665 13.14542

4 7.885869 7.222233 9.226303 7.515729 7.211537 18.22096 10.26238 9.607061 9.013558 7.313051 10.21326 10.26059

5 18.820167 19.357359 18.960993 18.888493 19.371964 15.68044 20.79039 20.817010 18.878796 19.335235 19.36913 19.35544

Clustering vector:

didnt gopdeb line night scottwalk walker car debat

4 1 4 3 4 4 3 5

fiorina presidenti rate realdonaldtrump trump foxnew tedcruz govmikehuckabe

2 4 3 5 5 3 3 2

love amp cruz truth vote fox megynkelli question

3 3 3 2 2 3 3 3

randpaul presid carlyfiorina impress made perform donald peopl

4 2 3 4 4 4 2 4

thing candid god great elect megyn time watch

4 3 4 3 4 4 2 3

america job john johnkasich kasich bush huckabe republican

2 3 4 2 2 3 4 2

rubio win give your chris christi dont gopd

3 2 4 2 4 4 2 4

set marco nail rand immigr thought realbencarson winner

2 4 2 4 4 4 2 2

bencarson donaldtrump marcorubio point lead make sens hes

4 4 2 4 4 2 4 2

close ben carson tonight jebbush moder polit answer

4 2 3 2 4 4 4 4

won show hillari agre donniewahlberg support ted imwithhuck

2 4 4 4 2 4 2 4

good gop call speak tcot american talk stage

2 3 4 4 4 4 2 4

leader parti news democrat favorit person paul forward

4 4 3 4 4 4 4 2

enjoy nomin expos ive head plan wallac countri

2 4 2 4 4 4 2 4

attack hand brought rais control admit face frontrunn

4 4 4 2 4 4 2 4

wont rwsurfergirl lrihendri democraticdeb legitim until band rid

4 5 2 2 2 4 2 2

Within cluster sum of squares by cluster:

[1] 0.000 9587.973 9854.368 5836.383 7624.972

(between\_SS / total\_SS = 90.3 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss" "size" "iter"

[9] "ifault"

> #Plot the total squared distance between clusters and within clusters by k value

> bss<-integer(length(2:15))

> for (i in 2:15) bss[i] <- kmeans(d,centers=i)$betweenss

> plot(1:15, bss, type="b", xlab="Number of Clusters",

+ ylab="Sum of squares", col="blue")

> wss<-integer(length(2:15))

> for (i in 2:15) wss[i] <- kmeans(d,centers=i)$tot.withinss

> lines(1:15, wss, type="b" )

Chart, line chart

Description automatically generated

**Appendix B**

**Text

Description automatically generated**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CANDIDATE** | **SUBJECT\_MATTER** | **TOTAL** | **TWEET\_Count** | **%** |
| Ben Carson | Abortion | 404 | 1 | 0 |
| Ben Carson | FOX News or Moderators | 404 | 16 | 4 |
| Ben Carson | Foreign Policy | 404 | 9 | 2 |
| Ben Carson | Healthcare (including Medicare) | 404 | 4 | 1 |
| Ben Carson | Jobs and Economy | 404 | 6 | 1 |
| Ben Carson | LGBT issues | 404 | 5 | 1 |
| Ben Carson | None of the above | 404 | 274 | 68 |
| Ben Carson | Racial issues | 404 | 70 | 17 |
| Ben Carson | Religion | 404 | 10 | 2 |
| Ben Carson | Women's Issues (not abortion though) | 404 | 1 | 0 |
| Ben Carson |  | 404 | 8 | 2 |
| Chris Christie | Abortion | 293 | 2 | 1 |
| Chris Christie | FOX News or Moderators | 293 | 11 | 4 |
| Chris Christie | Foreign Policy | 293 | 35 | 12 |
| Chris Christie | Healthcare (including Medicare) | 293 | 2 | 1 |
| Chris Christie | Immigration | 293 | 6 | 2 |
| Chris Christie | Jobs and Economy | 293 | 7 | 2 |
| Chris Christie | LGBT issues | 293 | 1 | 0 |
| Chris Christie | None of the above | 293 | 221 | 75 |
| Chris Christie | Racial issues | 293 | 1 | 0 |
| Chris Christie | Religion | 293 | 1 | 0 |
| Chris Christie | Women's Issues (not abortion though) | 293 | 1 | 0 |
| Chris Christie |  | 293 | 5 | 2 |
| Donald Trump | Abortion | 2813 | 17 | 1 |
| Donald Trump | FOX News or Moderators | 2813 | 849 | 30 |
| Donald Trump | Foreign Policy | 2813 | 26 | 1 |
| Donald Trump | Gun Control | 2813 | 2 | 0 |
| Donald Trump | Healthcare (including Medicare) | 2813 | 6 | 0 |
| Donald Trump | Immigration | 2813 | 62 | 2 |
| Donald Trump | Jobs and Economy | 2813 | 40 | 1 |
| Donald Trump | LGBT issues | 2813 | 1 | 0 |
| Donald Trump | None of the above | 2813 | 1604 | 57 |
| Donald Trump | Racial issues | 2813 | 12 | 0 |
| Donald Trump | Religion | 2813 | 14 | 0 |
| Donald Trump | Women's Issues (not abortion though) | 2813 | 114 | 4 |
| Donald Trump |  | 2813 | 66 | 2 |
| Jeb Bush | Abortion | 705 | 12 | 2 |
| Jeb Bush | FOX News or Moderators | 705 | 156 | 22 |
| Jeb Bush | Foreign Policy | 705 | 21 | 3 |
| Jeb Bush | Gun Control | 705 | 2 | 0 |
| Jeb Bush | Healthcare (including Medicare) | 705 | 3 | 0 |
| Jeb Bush | Immigration | 705 | 23 | 3 |
| Jeb Bush | Jobs and Economy | 705 | 19 | 3 |
| Jeb Bush | LGBT issues | 705 | 1 | 0 |
| Jeb Bush | None of the above | 705 | 445 | 63 |
| Jeb Bush | Racial issues | 705 | 3 | 0 |
| Jeb Bush | Religion | 705 | 1 | 0 |
| Jeb Bush | Women's Issues (not abortion though) | 705 | 7 | 1 |
| Jeb Bush |  | 705 | 12 | 2 |
| John Kasich | Abortion | 242 | 2 | 1 |
| John Kasich | FOX News or Moderators | 242 | 4 | 2 |
| John Kasich | Foreign Policy | 242 | 1 | 0 |
| John Kasich | Healthcare (including Medicare) | 242 | 7 | 3 |
| John Kasich | Jobs and Economy | 242 | 10 | 4 |
| John Kasich | LGBT issues | 242 | 31 | 13 |
| John Kasich | None of the above | 242 | 179 | 74 |
| John Kasich | Racial issues | 242 | 1 | 0 |
| John Kasich | Religion | 242 | 4 | 2 |
| John Kasich |  | 242 | 3 | 1 |
| Marco Rubio | Abortion | 275 | 26 | 9 |
| Marco Rubio | FOX News or Moderators | 275 | 4 | 1 |
| Marco Rubio | Foreign Policy | 275 | 3 | 1 |
| Marco Rubio | Healthcare (including Medicare) | 275 | 1 | 0 |
| Marco Rubio | Immigration | 275 | 19 | 7 |
| Marco Rubio | Jobs and Economy | 275 | 17 | 6 |
| Marco Rubio | LGBT issues | 275 | 2 | 1 |
| Marco Rubio | None of the above | 275 | 181 | 66 |
| Marco Rubio | Racial issues | 275 | 2 | 1 |
| Marco Rubio | Religion | 275 | 12 | 4 |
| Marco Rubio | Women's Issues (not abortion though) | 275 | 3 | 1 |
| Marco Rubio |  | 275 | 5 | 2 |
| Mike Huckabee | Abortion | 393 | 18 | 5 |
| Mike Huckabee | FOX News or Moderators | 393 | 4 | 1 |
| Mike Huckabee | Foreign Policy | 393 | 45 | 11 |
| Mike Huckabee | Gun Control | 393 | 4 | 1 |
| Mike Huckabee | Healthcare (including Medicare) | 393 | 6 | 2 |
| Mike Huckabee | Immigration | 393 | 1 | 0 |
| Mike Huckabee | Jobs and Economy | 393 | 14 | 4 |
| Mike Huckabee | LGBT issues | 393 | 28 | 7 |
| Mike Huckabee | None of the above | 393 | 247 | 63 |
| Mike Huckabee | Racial issues | 393 | 6 | 2 |
| Mike Huckabee | Religion | 393 | 3 | 1 |
| Mike Huckabee | Women's Issues (not abortion though) | 393 | 10 | 3 |
| Mike Huckabee |  | 393 | 7 | 2 |
| No candidate mentioned | Abortion | 7491 | 181 | 2 |
| No candidate mentioned | FOX News or Moderators | 7491 | 1775 | 24 |
| No candidate mentioned | Foreign Policy | 7491 | 155 | 2 |
| No candidate mentioned | Gun Control | 7491 | 43 | 1 |
| No candidate mentioned | Healthcare (including Medicare) | 7491 | 31 | 0 |
| No candidate mentioned | Immigration | 7491 | 87 | 1 |
| No candidate mentioned | Jobs and Economy | 7491 | 120 | 2 |
| No candidate mentioned | LGBT issues | 7491 | 53 | 1 |
| No candidate mentioned | None of the above | 7491 | 4095 | 55 |
| No candidate mentioned | Racial issues | 7491 | 248 | 3 |
| No candidate mentioned | Religion | 7491 | 330 | 4 |
| No candidate mentioned | Women's Issues (not abortion though) | 7491 | 220 | 3 |
| No candidate mentioned |  | 7491 | 153 | 2 |
| Rand Paul | Abortion | 263 | 1 | 0 |
| Rand Paul | FOX News or Moderators | 263 | 5 | 2 |
| Rand Paul | Foreign Policy | 263 | 13 | 5 |
| Rand Paul | Gun Control | 263 | 9 | 3 |
| Rand Paul | Healthcare (including Medicare) | 263 | 1 | 0 |
| Rand Paul | Jobs and Economy | 263 | 6 | 2 |
| Rand Paul | LGBT issues | 263 | 2 | 1 |
| Rand Paul | None of the above | 263 | 220 | 84 |
| Rand Paul | Racial issues | 263 | 1 | 0 |
| Rand Paul | Religion | 263 | 1 | 0 |
| Rand Paul | Women's Issues (not abortion though) | 263 | 1 | 0 |
| Rand Paul |  | 263 | 3 | 1 |
| Scott Walker | Abortion | 259 | 29 | 11 |
| Scott Walker | FOX News or Moderators | 259 | 11 | 4 |
| Scott Walker | Foreign Policy | 259 | 25 | 10 |
| Scott Walker | Healthcare (including Medicare) | 259 | 2 | 1 |
| Scott Walker | Immigration | 259 | 3 | 1 |
| Scott Walker | Jobs and Economy | 259 | 10 | 4 |
| Scott Walker | None of the above | 259 | 145 | 56 |
| Scott Walker | Racial issues | 259 | 9 | 3 |
| Scott Walker | Religion | 259 | 12 | 5 |
| Scott Walker | Women's Issues (not abortion though) | 259 | 4 | 2 |
| Scott Walker |  | 259 | 9 | 3 |
| Ted Cruz | Abortion | 637 | 4 | 1 |
| Ted Cruz | FOX News or Moderators | 637 | 62 | 10 |
| Ted Cruz | Foreign Policy | 637 | 31 | 5 |
| Ted Cruz | Gun Control | 637 | 1 | 0 |
| Ted Cruz | Healthcare (including Medicare) | 637 | 4 | 1 |
| Ted Cruz | Immigration | 637 | 6 | 1 |
| Ted Cruz | Jobs and Economy | 637 | 2 | 0 |
| Ted Cruz | LGBT issues | 637 | 1 | 0 |
| Ted Cruz | None of the above | 637 | 500 | 78 |
| Ted Cruz | Religion | 637 | 19 | 3 |
| Ted Cruz |  | 637 | 7 | 1 |