

# VizRunner: Systematic Visualization Tool of 2016 Presidential Campaign Final Report

Yiheng Hu  
[yih44@pitt.edu](mailto:yih44@pitt.edu)

Ran Ding  
[rad112@pitt.edu](mailto:rad112@pitt.edu)

Kefei Tong  
[ket73@pitt.edu](mailto:ket73@pitt.edu)

Weihao Zhang  
[wez63@pitt.edu](mailto:wez63@pitt.edu)

## ABSTRACT

U.S Presidential Campaign in United State is running every 4 years, and 2016 Presidential Campaign became heat since both party has reached the end of debating and processed to nominate the president candidate within their party. In this project, we demonstrate a systematic way to track a candidate's status during campaign and compare the differences between two candidates. Our data source is based on different media/organization's poll, candidate's super PAC reports and third party observation, reports and journals from mainstream media. And three modules are expected to be implemented in our system.

## CCS Concepts

• Human-centered computing → Visualization

## Keywords

2016 President Campaign; Information Visualization; Political data Research; Media Information Visualization; Text-Mining;

## 1. INTRODUCTION

U.S presidential campaign is one of the most important issues for political trend study. The U.S presidential campaign is quadrennial, usually 2 years after mid-term election. All U.S registered voters will attend this general election to vote for their next president. Before the national election, two main parties will first nominate two candidates representing their party to run for the presidential campaign in June 2016. In order to win their inner party preliminary election, generally candidates will attend public TV debate, interview and other public activities in 2015 to gain enough support from their party members.

In this project, due to the nature of course period, we will only focus the early stage of presidential campaign. We plan to use data by December 2015 to ensure that we can cover enough data points as possible. Our goal in this project is to help user systematically track and analyse the performance of each candidate, based on poll results and his/her super PAC performance in different time periods. And one additional text analysis visualization module will help user understand the mainstream media's impression to a candidate.

## 2. Background and Data Explanation

United Presidential Election is a national election, which is hold for every 4 years. According to Article Two of the United States Constitution, the president is elected by United States registered voters. Normally, two main parties as Democratic and Republican use method called primary election to decided which candidate is best suit to run for the election. The primary election is an indirect election that cast the poll of voters who support a board of party delegates, and those delegates pledged to a specific candidate. And the primary election is always held in the early of election year, as early of 2016. Before primary election, candidates usually organise their super PAC to raise fund for campaign, and attend

various events and debate to earn better new coverages and more inner-party support.

The first direct campaign related data is poll and survey data, which represented a candidate's support rate and probability to win the presidential campaign. The main stream media uses a specific sampling method to ensure the sampling group is similar to real general election voter base, however, since the U.S election result is decided by election colleges in states, a single national survey sometimes may not accurately represent the real probability of candidate winning election. As many political studies revealed, some states (called as 'Swing State') played a more critical role in general election since they decided where undecided electoral votes going to. The geographical differences in survey are also useful for analyst to predict the result of election.

And, at current stage, the campaign focus on preliminary nomination for two parties, so most common survey only covers the support rate for candidates in their own party. Only very few specific surveys investigated the one vs. one poll results. Though those polls may not reflect the real result in general election, they still indicate a preference of voters for cross party competitors, so we included this data in our system.

Secondary, donation fund size is another important index to measure whether a candidate gain enough support, since nobody would donate to an unwelcomed candidate. Additionally, the early stage of superPAC funds were usually donated by enterprises rather than normal people. Hence, a visualization of superPAC funds can reveal a candidate's industry connections, which is not obviously available on poll data.

Another interesting part of superPAC is its geographical distribution and quarterly report. The geographical report could tell us which state prefer a candidate best, though in overall, the data is straight correlated with the economic status of states, it still presented a preference difference between northern U.S and southern U.S, east costa and west costa. For quarterly report, we can focus on the cash flow status for each candidate, and compared with their support rate, which may help user to understand whether candidate spend their cash in a reasonable way, or whether they gained support and donation at same time.

Thirdly, news media and mass media's opinions are very influential to election result. News media, on one hand, can forge their readers' opinion by broadcasting their value and ideas. On the other hand, to earn market and profit, a news media needs to keep closer value similar to its subscribers. Therefore, by researching the terms used by reports and journals, we could mine out news media's opinion from mass volumes of text, which will help other researchers understand whether a candidate performed good or bad on public media.

Excepted opinions, the vital events reported on news media is a good reference to help readers understand which factor contribute/demur candidate's supporting rate. Accordingly, we

included this data to track the relationship between poll and important event for candidates.

Based on all above consideration and discussion, we finally defined our system data sources in four approaches: 1. **Polls and survey for each candidate** 2. **Super PAC report and third party super PAC analysis** 3. **Mainstream new media comments and critics on high-support rate candidate** 4. **News media reported events for each candidates and important public events.** And each category employed following data source:

1. Polls and Survey data: we now using from [http://www.realclearpolitics.com/epolls/latest\\_polls/president/](http://www.realclearpolitics.com/epolls/latest_polls/president/) to collect all main stream poll result from January 2015 to December 2015 for all party candidates. And we specifically track three main polls as PPP, CNN and Fox News in 2-year period specific candidates. We also collected several state's survey results, and candidate vs. candidate and preliminary election polls.
2. Super PAC data: two data sources were included as <http://docquery.fec.gov/pres/>, <https://www.opensecrets.org>. We extracted the change of fund size, their donors' industry composite, top states that contribute the donations. We are also looking possible data that reveal the network of super PAC (the advisor panel) for network visualization but it doesn't go very well.
3. News Media comments and critics: We used script to capture only news comments text about Hillary Clinton, Donald Trump, Marco Rubio from Washington Posts. For each candidate, we finally collected 1500-2000 news reports and opinion articles. Those documents were then tokenized, removed stop-word and whites, stemming and transformed to terms. Finally, we used TF-IDF calculated each term's value.
4. Important Public Events and Events related to one candidate: we used Wikipedia schedule ([https://en.wikipedia.org/wiki/United\\_States\\_presidential\\_election,\\_2016\\_timeline](https://en.wikipedia.org/wiki/United_States_presidential_election,_2016_timeline)) to collected important public events such as debating time and candidate announced to attend campaign. Important events for each candidate are defined by the number of comments and popularity of such news on CNN, WSJ and New York Times. We referred some political blogs to define the important events for each candidate.

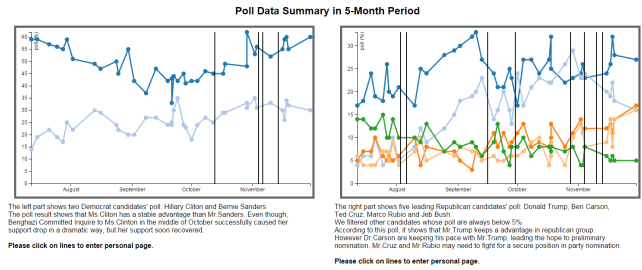
### 3. Visualization Design

#### 3.1 Framework and Module Design

Through our group discussion, we decided to design three modules in our visualization system. **Candidate poll and event tracker**, **Candidate super-PAC tracker** and **Media text analysis**. Except, Media text analysis module, both candidate poll and event tracker and super-PAC tracker has a summary page for comparison and personal page to illustrate personal specifications [5] [6]

#### 3.2 Candidate Poll and Event Tracker

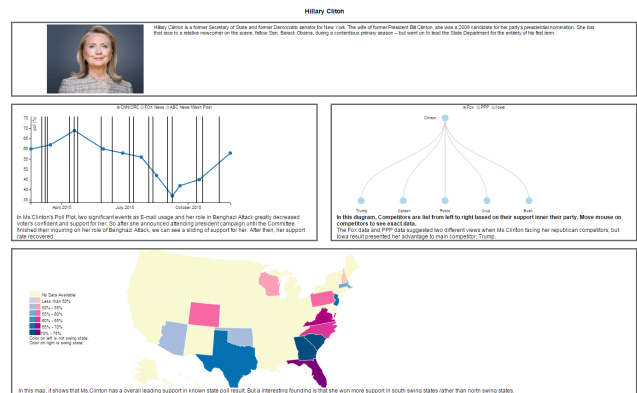
The Poll and Event Tracker is aimed to visualize two parties' candidate performance compared with their competitors (since in primary election candidate only needs to win competitors in same party) and events that could possibly influence their support rate. There are two parts were designed in this module, in part one, two parties' candidate support rate will be displayed on two separate time series plot, and significant events will be marked on the time series plot.[3]



**Figure. 1 Example of 6 Month Time Series Plot of Polls**

When user moved mouse on each candidate's plot, it will highlight the candidate's plot and fade others' plot. By clicking the plot, user can move to personal part of poll and event tracker.

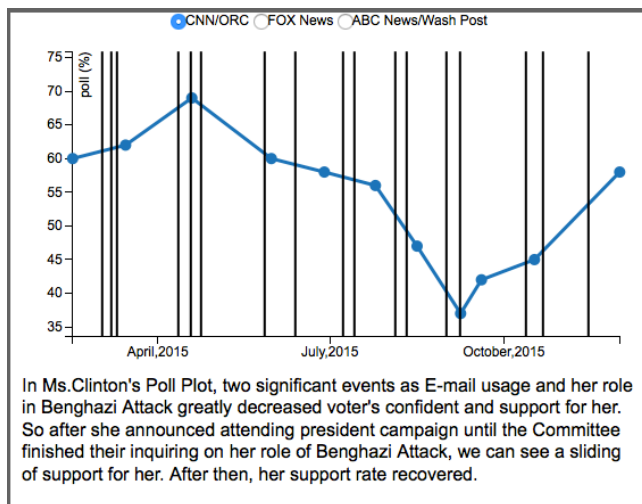
In the second part of poll and event tracker, it presented a specific candidate performance in the beginning of campaign to current time, his/her performance compared to other candidate based on match up survey, and his/her support rate by different states. The long time personal poll track will still use time series plot, but allow user to chose a preferred poll data source, and important event only related to this candidate will show on this time series plot, for example, he/she attended a television interview and gave a critical speech during campaign. For the 1 vs. 1 part, we considered to use a force directed network by different node size and edge to illustrate the possibility of wining and possibility that two candidates will meet in final election. The bottom part will use a state map to visualize the poll rate for a candidate in different states.



**Figure. 2 Personal Poll and Event Tracker Design**

##### 3.2.1 Supporting Rate Plot

The time-series plot on the top left of the page represents the supporting rate changes of the certain candidate. The vertical lines on these plots represent more detail personal events such as personal scandals and political positions on certain areas. There are three buttons for changing data sources from different polls with different political standings. Same as the summary page, when user put the mouse on the certain vertical lines or certain nodes on the time-series line, a sub-rectangle with details will appear.



**Figure 3. Screenshot of Supporting Rate Plot on Personal Page of Polls (Hillary Clinton)**

With the vertical lines to indicate the correlation of events and supporting rate changes and different polls to show the survey from different media with different political standings, we try to show more details of the supporting rate changes of the certain candidates. The crosses of vertical lines and time-series plot can help user identify some important events of the certain candidates, which may effectively produce a switching Here's visual encoding table: [8]

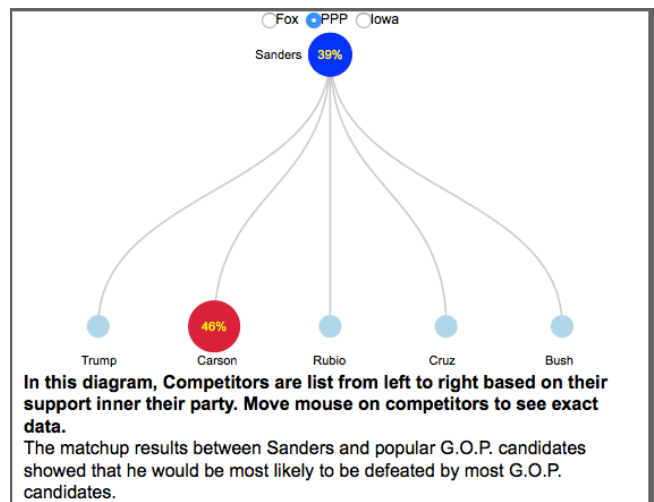
Variable	Type	Encoding
Support Rate	Quantitative	Y
Time	Ordinal	X
Individual Event	Normal	Line Parallel as Y
Data Source	Categorical	Widget (Radio)

**Table 1. Individual Poll Visualization (Time Series Plot)**

The result as we see is good, as well. For example, we can see a strong rising of Ms. Hillary Clinton's supporting rate after the Congress Hearing of her private email scandal in October of all the three polls.

### 3.2.2 One-To-One Matchup

On the top right of the personal page of polls, we put a network to show the one-to-one matchup of the certain candidate on the candidates from the other party. We use three buttons on the top to help users switch from different polls. When user put the mouse on the certain opponent, the certain node and the node on the top with represents the original candidate will be highlighted and information of one-to-one supporting rates will appear on each node, with red node represents the winner and blue one represents the loser.



**Figure 4. Screenshot of One-To-One Matchup Network on Personal Page of Polls (Sen. Bernie Sanders)**

The main purpose for this network is to show the one-to-one matchup competition results of the certain candidates when versus the candidates from the other party. So that to find the favor and victims of the certain candidates. The following table is our visual encoding analysis:

Variable	Type	Encoding
Candidate	Categorical	Shape Circle
Win/Lose	Categorical	Color Value
Support Rate	Quantitative	Color, Text
Match-Up	Normal	Connection
Data Source	Categorical	Widget (Radio)

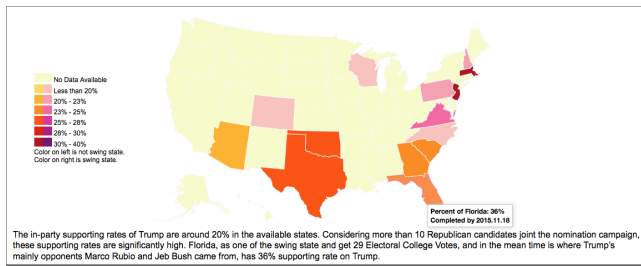
**Table 2. 1 vs. 1 Support Rate Comparison Visualization (Tree)**

We try to show the possible results of one-to-one matchups between every candidate from the two parties. Given the one-to-one matchup of Sen. Bernie Sanders as an example, we can see all the three polls indicates that he may lose when matchup with republican candidates.

### 3.2.3 Map of Supporting Rate by States

The map picture on the bottom part of the personal page of polls shows the supporting rates by states on certain candidates.

For Republican candidates, we use purple to illustrate the swing states and red for other states, while for Democratic candidates, we use purple to indicate swing state and blue for the remains. The swing states we selected are Pennsylvania, Wisconsin, Virginia, North Carolina, Florida and Colorado. In the meantime, we use six levels of hue to indicate the supporting rate levels of certain candidate on a given state. Similarly, when user put the mouse on a certain state, the state will be highlighted and the sub-rectangle with detail information will appear.[4]



**Figure 5. Screenshot of Map of Supporting Rate by Sates on Personal Page of Polls (Donald Trump)**

We try to visualize the supporting rate verify on different states of a certain candidates, and indicates some key states, such as swing states in the same time. We hope it will help user to better understand on the supporting rates of certain candidates on some key states. For example, Donald Trump earns a high supporting on Florida, which is a famous swing state. And our visual encoding table is shown below:

Variable	Type	Encoding
State	Normal	Geography Size and Position
Support Rate	Quantitative	Color Hue
Normal/Swing State	Categorical	Color Value

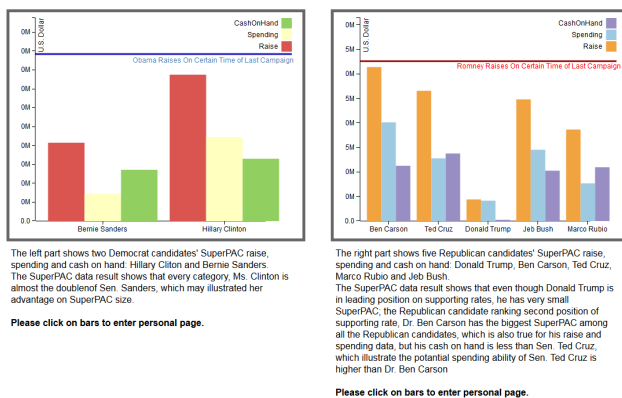
**Table 3. State Support Rate Visualization (Map)**

### 3.3 Candidate super-PAC tracker

Candidate super-PAC tracker is module designed to help user understand candidate's fund status and compare two candidates based on their super-PAC composite. In this module, the outline is very similar to poll and event tracker, which make user easier located information.

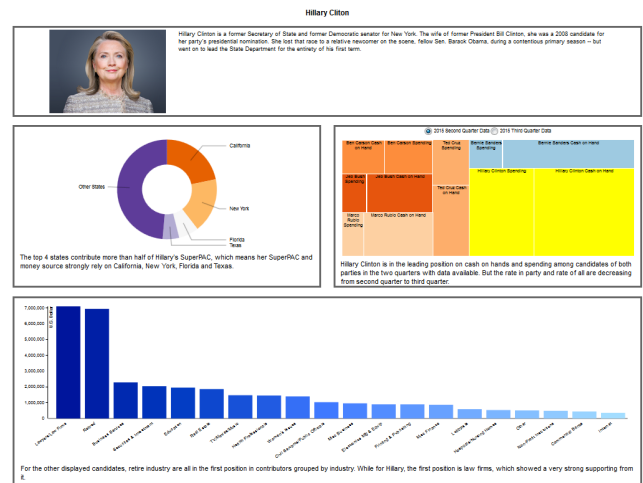
In this module, summary will show two parties' candidate already received fund size in a bar chart, compared with the fund raised by 2012 presidential candidates Barack Obama and Mitt Romney's final raised fund. This will help user understand the current progress of election.

**SuperPAC Data Summary of The Year 2015 Till October**



**Figure. 6 Summary Comparison of Super-PAC Tracker**

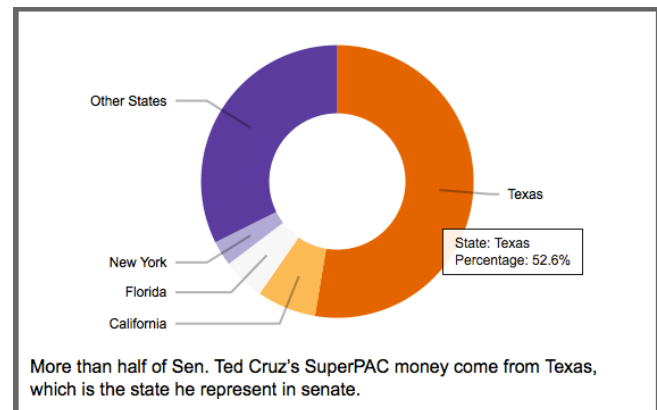
The second part of the module is about a specific candidate's super-PAC visualization. In this part, user can check the composite of donors to super-PAC, either by type of industry or by geographical distribution (states). In the bottom of this part, a tree map can dynamic shows the fund size of target candidates compared with other candidates in a historical view.



**Figure 7. Part 2 of Super-PAC Tracker**

#### 3.3.1 Pie Chart of Super PAC Constitution by States

A Pie Chart is showed on the top left of the page. It shows the occupying rates of the constitution of the candidate's Super PAC by the highest 4 donation states and others. When user put the mouse on certain part of the pie, it will show the detail data of the state:



**Figure 8. Screenshot of Pie Chart of Super PAC Constitution by States (Sen. Ted Cruz)**

We can see for most candidates, Texas, California, Florida and New York are the most frequency states appeared on the top 4 states. So the appearance of other states on a certain candidate may represent an unusual supporting. Meanwhile, the rates of the top 4 states are varying among different candidates, for example, Sen. Ted Cruz get more than half of his Super PAC from Texas, which also match his supporting rate by states visual result on the personal page of supporting rate, that Texas records the highest supporting on him, comparing to other states with data. And the visual encode for this visualization is shown below:

Variable	Type	Encoding
Fund Size	Quantitative	Size
State	Categorical	Color Value

**Table 4. State Support Rate Visualization (Pie Chart)**



### 3.3.2 Tree Map of Candidates from the 2 Quarter to Third Quarter of 2015

We put a tree map on the top left of the personal page of Super PAC. We use red and blue, which are traditional color for each party, with different hue to represent the candidates of each party. Meanwhile, we use yellow to highlight the current candidate. For each candidate with data, we have two sub-rectangles separately representing the spending and cash on hand amount rates of in-party and between-parties' changes. There are dynamic changes when users switch from certain quarters by clicking on the buttons on the top.[2]

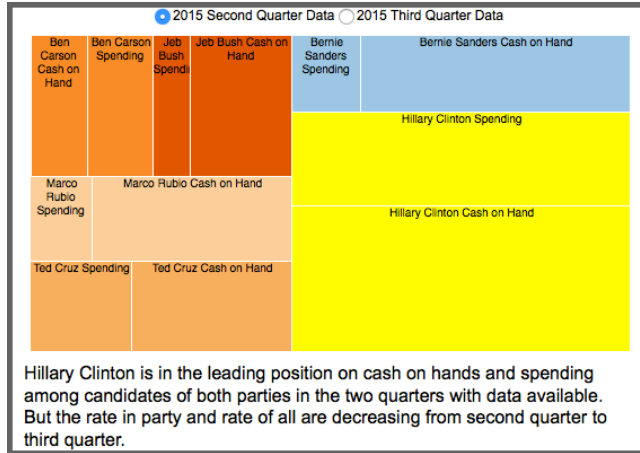


Figure 9. Screenshot of Tree Map of Candidates' Spending & Cash On Hand (Hillary Clinton)

We try to highlight the spending and cash on hand rate of the certain candidate's SuperPAC on both the in-party and between-parties' levels on different quarters. We find a huge rate of Ms. Hillary Clinton's SuperPAC size comparing to others, but her spending and cash on hand rate are decreasing from second quarter to third quarter on both in-party and between-parties' levels. Even though we only select two Democratic candidates versus five Republican candidates, the spending and cash on hand rate of Democratic are obviously relatively equal or even bigger than Republican. And this is a very interesting visual idea inspired by NY times visualization, the visual encoding is shown below:

Variable	Type	Encoding
Fund Size	Quantitative	Size
Fund Type	Categorical	Shape (Rectangle)
Candidate	Categorical	Color Value (And Highlight)
Time	Ordinal	Widget (Radio)

Table 5. Quarterly Super PAC Fund Comparison Visualization (Tree Map)

### 3.3.3 Bar Charts of Super PAC Constitution by Industries

We put a 20-bars bar chart on the bottom of the personal page of Super PAC to illustrate the candidate's Super PAC constitution by the top 20 industries.

Similarly, we use blue bars on Democratic candidates' pages, red on Republican candidates' personal pages, with hue from deep to light representing the contribution level from most to least.

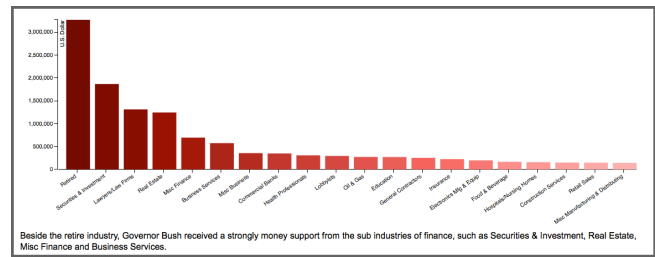


Figure 10. Screenshot of bar Charts of SuperPAC Constitution By Industries (Gov. Jeb Bush)

We find the retire industry ranking for most first position, except for Hillary Clinton, which was law firms at that position. Besides that, we find that most candidate were performed by certain industries, for example, Governor Jeb Bush receive more money from financial and related industries, Doctor Ben Carson receive more money from medical industries and so on. Although this is a very simple visualization, it still used two color attribute to emphasis the difference of fund size, the visual encode is shown below:

Variable	Type	Encoding
Fund Size	Quantitative	Y and Color Hue
Industry Type	Categorical	X
Party	Categorical	Color Value

Table 6. Industry Super PAC Fund Comparison Visualization (Bar Chart)

### 3.4 Media analysis

As it stated before, for the media report data, we have collected the articles from Washington Post by the search term of each candidate, and then calculated the appearance of each term, remove stop words, then reweight them by TF-IDF, and sort them by top-down. Then we select the adjective words and the special nouns, which will label the candidates specially from others, and use these words to create word clouds to describe certain candidate.

Data is collected using Python as programming language, scrapy as crawling spider framework, selenium as font-end browsers manipulation tools, mongodb as Nosql database to store data. We chose to collect comments from the Washington Post within 12 month using relevance criterion. Data is collected by program automatically. And Here are the numbers we collect for each candidate:

Candidate	Number
Marco Rubio	1137
Ben Carson	1561
Donald Trump	1935
Hillary Clinton	2499

Table 7. Raw Text Data Summary

Once the data is collected, we use word cloud open source library and modify its ranking algorithm to use TF-IDF as criterion. After choosing proper color as background and font size, then word cloud diagrams can be generated automatically. And we selected four candidates: **Sen. Marco Rubio, Hillary Clinton, Donald Trump and Dr. Ben Carson**. For each of them, we also added some comments and term introductions on the bottom.[1]



Figure 10. Screenshot of Word Cloud Page

The initial idea of this part is to find the objective words on medias to describe the candidates. But unexpectedly, we find some special nouns and names on the lists. For example, we find “Michigander” for Dr. Ben Carson, “D-Maryland” and “Jordan” for Hillary Clinton, “Science” for Sen. Marco Rubio and “Mom” for Donald Trump. All these terms have special meaning for the certain candidates, for example, media may point on the term “Mom” when talking about Donald Trump as he criticizes on the current immigration policy while his mother is a foreigner. We try to explain these special terms under each picture. Because these terms work well on shaping and labeling the certain candidates. Word Cloud technique is very simple but intuitive visualization method. The visual encoding table is shown below:

Variable	Type	Encoding
Word	Categorical	Color Value
Frequency	Quantitative	Size

Table 8. News Media Words Visualization (Word Cloud)

## 4. Discussion

In this visualization project, our original idea to try to design an integrate system that can visualize different types of data related to 2016 presidential campaign. This idea is constructed based on our understanding for U.S political environment, since the support rate of normal citizens and large enterprises are two keystones for modern U.S presidential campaign. There are several aspects we are interested in but did not implement in our project, due to the lack of data or too complicated to visualize in one figure:

- Network of Super PAC advisors: This visualization is done by New York Times. And we considered to reconstruct this visualization but found we did not have such data.
- Donald Trump’s Data: Since Donald Trump announced to attend campaign very late, there’s very few data about how his super PAC constructed and how much fund he raised for campaign. It caused a problem in our super PAC visualization part.
- Restriction for News Access: In our original plan, we want to collected documents and reports from difference media source. Unfortunately, we found most important medias are only open for subscribers such as New York Times and

Wall Street Journal. So finally, we have to limited our data source to Washington Posts, which is open for crawlers.

In our future plan, we hope to continue track and collect data until April 2016 when the preliminary election has result. By that time, after we analyze data on our hand, we may have new idea to give final general election candidates a more innovative and completed visualization for them.

## 5. Conclusion

In this project, we try to present the current information we have about 2016 presidential campaign in three different attributes: supporting rate polls, Super PAC and media sentiments. Specially, we also present the candidates’ personal information on supporting rate and Super PAC parts. The main idea of this project is to help users form a complete and objective cognition on the main candidates of the two parties. We try to make every color and hue in this project meaningful, so instead of making a colorful and gorgeous, we approached many red and blue colors and purple, which is the mix of red and blue. As red and blue are the traditional representing colors of Republican and Democratic. As we have summarized make interesting points in the Layout section, which are the points we formed after we have this plots, we hope our users may enjoy having this project as will and can get some interesting conclusion[7].

## 6. Team Contribution

Yiheng Hu: Design and Implement the visualization code, apply and adjusted webpages.

Ran Ding: System Framework Design, Visualization Method Evaluation and Data analysis

Kefei Tong: Background Knowledge Provider, Documentation and Super PAC visualization design.

Weihao Zhang: Text Mining Module Design and Implementation.

## 7. REFERENCE

1. News Media comments and critics:

<http://www.wsj.com/>

<http://www.nytimes.com/>

<http://www.washingtonpost.com>

2. Tree Map example, which inspire our design:

[http://www.nytimes.com/interactive/2009/09/12/business/financial-markets-graphic.html?\\_r=0](http://www.nytimes.com/interactive/2009/09/12/business/financial-markets-graphic.html?_r=0)

3. Important Public Events and Events related to one candidate:

[https://en.wikipedia.org/wiki/United\\_States\\_presidential\\_election\\_2016\\_timeline](https://en.wikipedia.org/wiki/United_States_presidential_election_2016_timeline)

4. GeoTime information visualization, which inspired our design on map part:

<http://ivi.sagepub.com/content/4/2/136.full.pdf>

5. Business Information Visualization example, which inspire our design:

[http://s3.amazonaws.com/academia.edu.documents/31045961/BusInfoVizTut.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1447016668&Signature=eYFfx7ELsXZ0yuSrJCYpWskhL4U%3D&response-content-disposition=inline%3B%20filename%3DBusiness\\_information\\_visualization.pdf](http://s3.amazonaws.com/academia.edu.documents/31045961/BusInfoVizTut.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1447016668&Signature=eYFfx7ELsXZ0yuSrJCYpWskhL4U%3D&response-content-disposition=inline%3B%20filename%3DBusiness_information_visualization.pdf)

6. An Evaluation of Space-Filling Information Visualizations for Depicting Hierarchical Structures  
<https://smartech.gatech.edu/bitstream/handle/1853/3414/00-03.pdf>
7. The Challenge of Information Visualization Evaluation

- <http://triton.cc.gatech.edu/hci-seminar/uploads/1/The%20Challenge%20of%20Information%20Visualization%20Evaluation.pdf>
8. Noah Iliinsky: Visual Encoding & Best Uses