Relationships Between Presidents Twitter and Political Issues

Jimmy Chan

July 10, 2018

Abstract

Twitter is one of the most popular social media in the U.S. and used by influential politicians including US presidents. Twitter has been proven to be a tool to predict the real world. There are many kinds of research on the relationship between Twitter and the stock market. The most famous study is Twitter mood predicts the stock market by J Bollen which has been cited more than 3000 times. In this project, we sought to answer a question: Do politicians' tweets match to their real-life behaviors? With the answer to this question, we can forecast the changes in political concerns by analyzing tweets.

1 Introduction

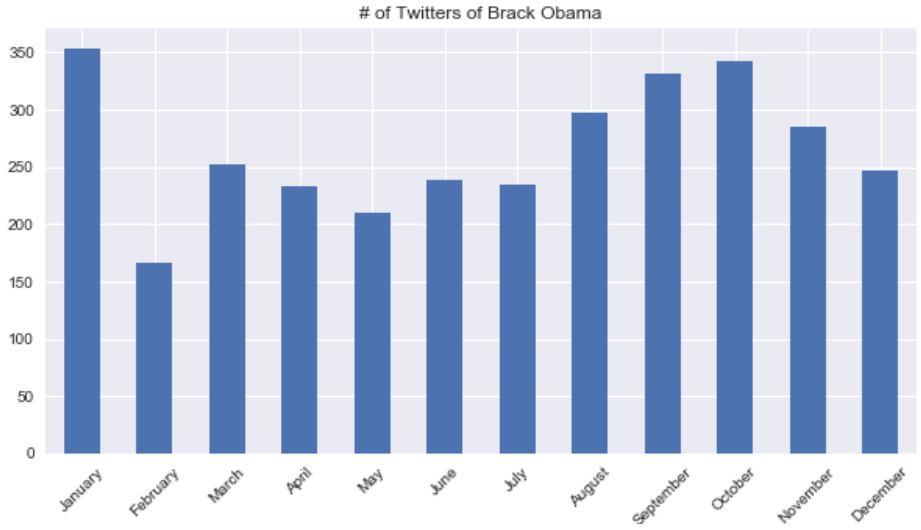
The purpose of the project is analyzing the Twitter behaviors of Donald Trump and Barack Obama and comparing their Twitter behaviors to what they do in the government. We will explore their tweets by a Natural Language Processing approach, and find out their characteristics. We will compare their tweets to find the changes in political concerns.

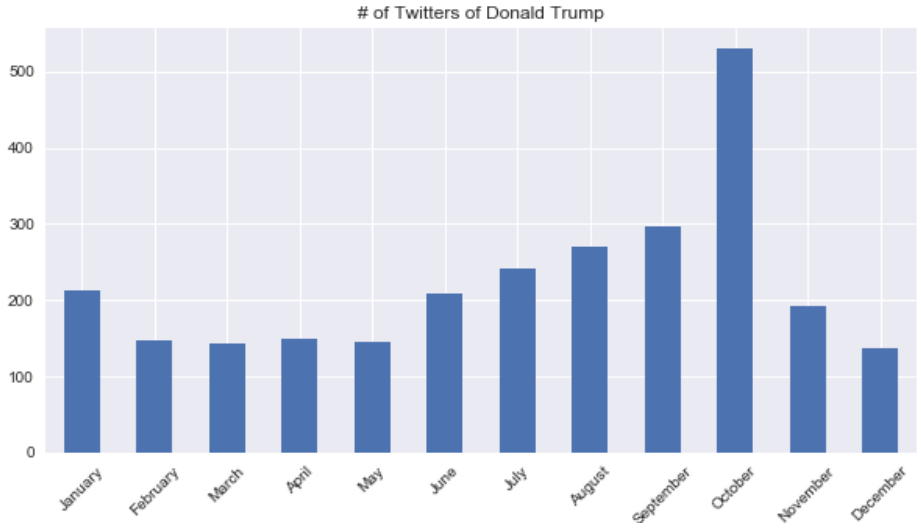
First, we will briefly explore the usage of the twitter of both presidents, for example, the frequency of tweets, popularity change, etc. Then we will process content of tweets with NLTK to explore aggregate text statistics of the speeches and visualize some of their properties. Also, we will analyze the relationships between the properties of the tweets, time, and provide a novel visualization of the relevant information. Lastly, we try to use Latent Dirichlet allocation model to find the political concerns change between the President Obama's last 10 Months of the presidency and President Trump's first 10 months of presidency in twitter in order to match to the actual changes of political concerns between two presidents in the real world.

2 Visualization of the Raw Data

2.0.1 Frequency of tweeting histograms

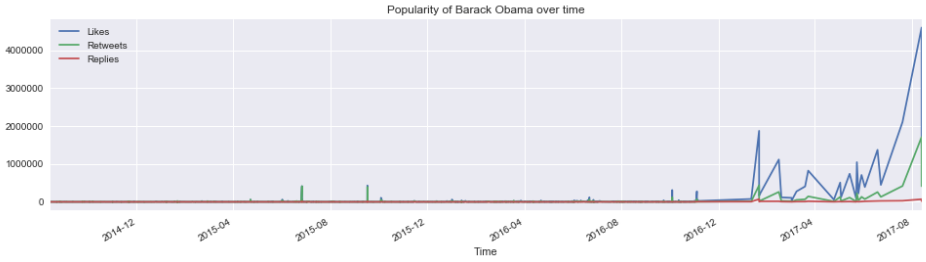
We gather the frequency of tweets of President Obama and President Trump by month and visualize the distributions by histograms.

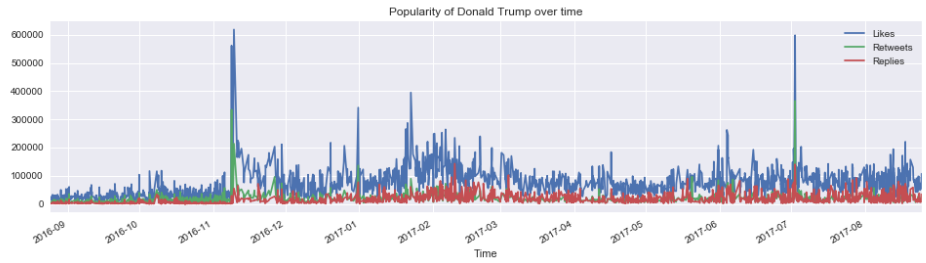




We can connect the frequency of tweeting with political events. The the presidential election of 2016 was held on November 8 and the number of tweets of Donald Trump in October is much higher than the other months. It seems the political events would influence the number of Twitters for Donald Trump. The frequency of tweeting of Brack Obama is very normal, we cannot generate insights based on the data.

2.0.2 Popularity of both presidents in Twitter





The popularity of President Obama is much better than President Trump. Trump's likes reached the peak twice, once in Nov 2016 and once in July 2017, those tweets got near 600000 likes. And the popularity of Obama suddenly changed after Jan 2017. We can easily guess Obama didn't always tweet before Jan 2017. His likes raised rapidly after Jan 2017, always got more than 2000000 likes. His likes reached the peak in Aug 2018, more than 4000000 likes, which was six times more than Trump.

3 Text Processing and Novel Visualization of the Tweets

We processed and cleaned up the content of their tweets with NLTK. And we counted the number of words, unique words, and characters of all tweets. We try to find the descriptive statistics(e.g. mean, count, sd, etc.) of cleaned tweets and explore insights. - ***n_words - number of words - n_wwords - number of unique words - n_chars - number of characters***

Descriptive statistics of Obama

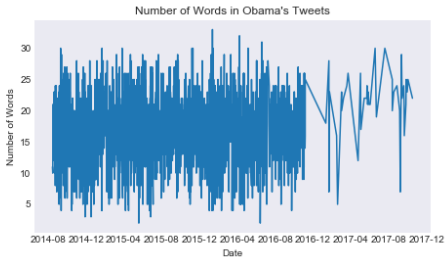
	n_words	n_uwords	n_chars
count	3207.000000	3207.000000	3207.000000
mean	17.289679	16.475834	116.560337
std	4.826105	4.326429	22.661028
min	2.000000	2.000000	28.000000
25%	14.000000	14.000000	102.000000
50%	17.000000	17.000000	122.000000
75%	20.000000	19.000000	135.000000
max	33.000000	30.000000	157.000000

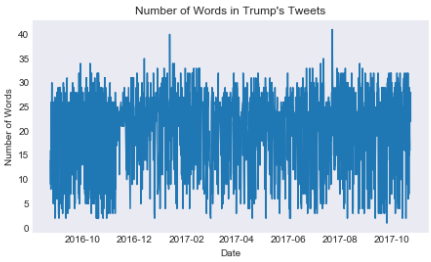
Descriptive statistics of Trump

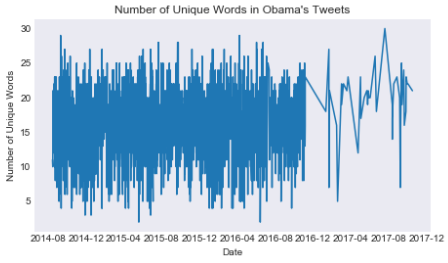
	n_words	n_uwords	n_chars
count	3232.000000	3232.000000	3232.000000
mean	20.697401	19.395421	124.827042
std	6.906415	6.157699	31.060590
min	1.000000	1.000000	10.000000
25%	17.000000	16.000000	112.000000
50%	22.000000	21.000000	137.000000
75%	26.000000	24.000000	142.000000
max	41.000000	31.000000	208.000000

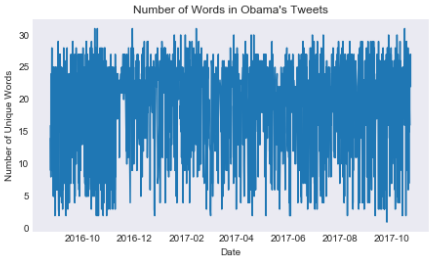
3.0.1 Aggregate text statistics with NLTK

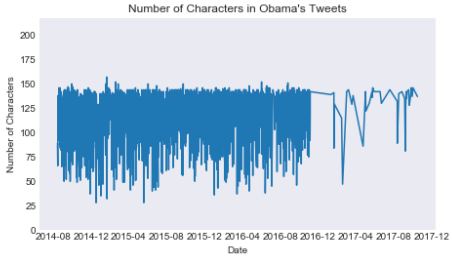
We did the text processing with NLTK to compute the aggregate text statistics with NLTK and visualized their properties.

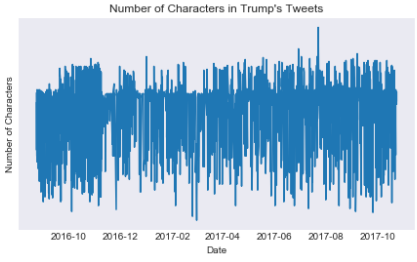












We can find both presidents' tweeting behaviors when we compare the aggregate text statistics of both presidents. Trump uses more words and characters than Obama on average, but the standard deviations of Trump's text statistics are much higher than Obama. It means trump either writes a long tweet or short tweet while Obama is somewhere in the middle.

4 Change of political concerns between the President Obama last 10 Months of presidency and president's Trump first 10 months of presidency in Twitter

We will use the LDA to analyze the political concerns of President Obama last 10 Months and President Trump first 10 months. Latent Dirichlet allocation (LDA) is a topic model that generates topics based on word frequency from a set of documents

```
              Top 10 Topic of Obama
0  obama,fair,suprem,peopl,continu,enter,read,wan...
1  court,time,chang,live,economi,get,make,http,ke...
2  senat,hear,leader,watch,deadlin,chanc,will,mee...
3  judg,garland,give,climat,american,add,nation,o...
4  doyourjob,million,record,new,call,nomine,one,w...
5  support,need,chicago,take,day,america,voic,tea...
6  presid,now,name,speak,tune,vacanc,unit,board,e...
7  ,scotu,play,growth,believ,sector,qualifi,white...
8  actonclim,vote,job,merrick,deserv,say,see,heal...
9  today,fight,equal,agre,nomin,work,process,week...

              Top 10 Topic of trump
0  great,presid,can,border,long,call,prime,happen...
1  fake,bad,new,interview,tax,take,support,forwa...
2  countri,just,get,year,come,fail,enjoy,good,ord...
3  job,democrat,day,back,vote,hous,must,want,begi...
4  nation,work,tonight,unit,watch,don,wonder,neve...
5  thank,american,make,obama,honor,secur,obamacar...
6  big,meet,win,join,senat,allow,washington,toget...
7  will,now,russia,court,much,alway,see,statement...
8  peopl,today,america,trump,even,look,elect,cong...
9  news,media,state,time,mani,healthcar,republica...
```

4.0.1 Compare the topic change by LDA

The important word from the top 5 topics of Obama's political concerns

- First topic are **fair, suprem, peopl, continu**
- Second topic is **court, time, chang, live, economi**
- Third topic is **senat, will, leader, think**
- Forth topic is **judg, climat, american, obamacar**

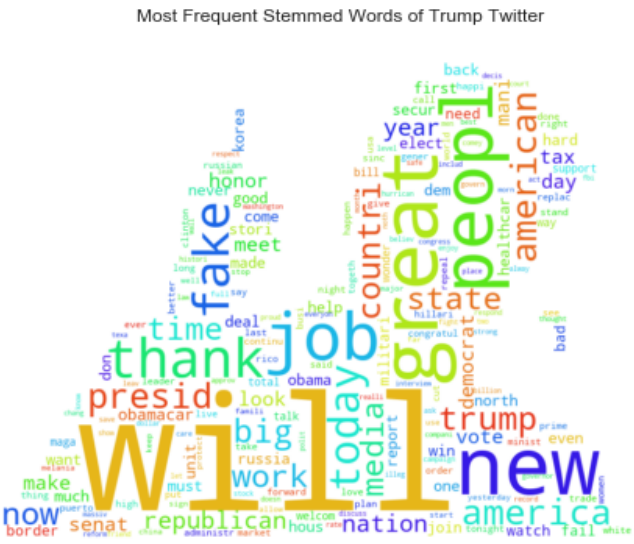
- Fifth topic is **doyourjob, new, women, gun**

The important word from the top 5 topics for Trump's political concerns

- First topic are **great, border, presid, iran, terribl**
- Second topic is **fake, new, interview, tax, administr,hard**
- Third topic is **countri, just, come, made**
- Forth topic is **job, democrat, vote, white**
- Fifth topic is **work, execut, women, help**

Most topics from Obama's tweets are related to social issue and the communtiy, e.g. peopl, fair, obamacar, economi, senat, climat, women, gun. We can guess from the LDA analysis, his topics more about the economic, climate change, women rights, obama care and gun control. Actually, he did work on these topics in his presidency. Most topics from trump's tweets are related to national security, tax, media and his slogan "Make America Great Again" , e.g. great, border, iran, fake, new, interview, tax. We can guess from LDA analysis, his topics are related to his slogan "Make America Great Again", building wall along the Mexican border, blaming the media and Iran nuclear deal. These topics actually are his political concerns. From the above result, we can say that the LDA performed a well analysis of the topics of twitter data.

4.1 WordCloud - Most Frequent Stemmed Words



4.2 Compare the word cloud

We can see that the most frequent stemmed words of Obama's Twitter are *presid*, *senat* , *Obama*, *need*, *change*, *job*, *leader*, *American*.We can see that the most frequent stemmed words of Trump's Twitter changed to *will*, *new* , *great*, *thank*, *job*, *fake*, *presid*, *today*. We know 2008 is a huge challenge for the United States. Job problem, terrorism problem, and climate problem was what Mr. Obama try to focus. And for Mr. Trump, we know "Make America Great Again" (MAGA) is a campaign slogan used in American politics that was popularized by President Donald Trump in his 2016 presidential campaign.

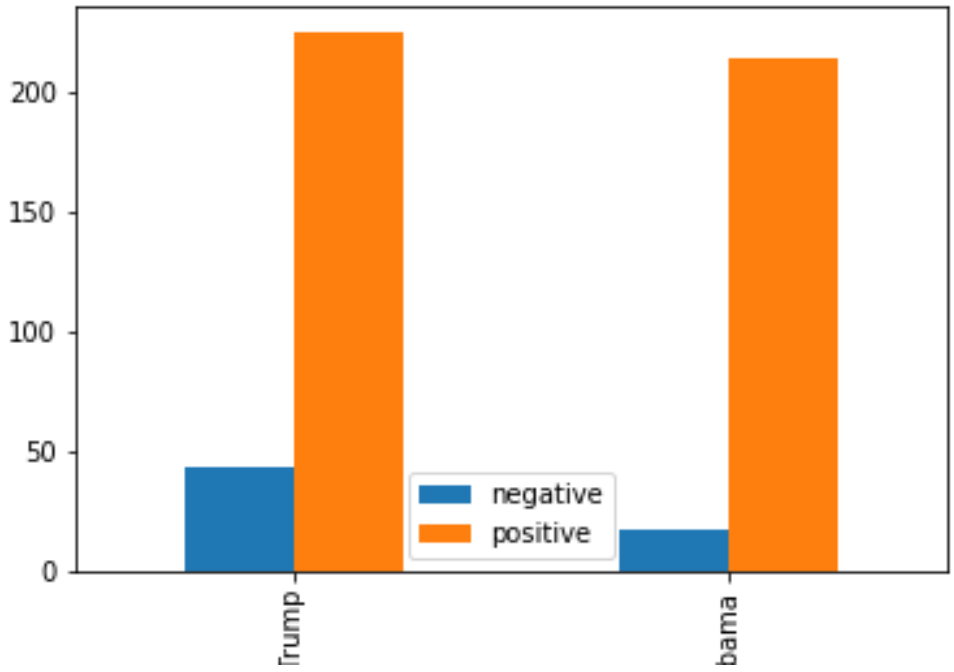
5 Sentiment Analysis

An analysis of Trump and Obama twitter data by using the Sentiment Analyzer from the NLTK package. We use the sentiment analysis to get the positive and negative on their tweets and compare their sentiment by the twitter data.

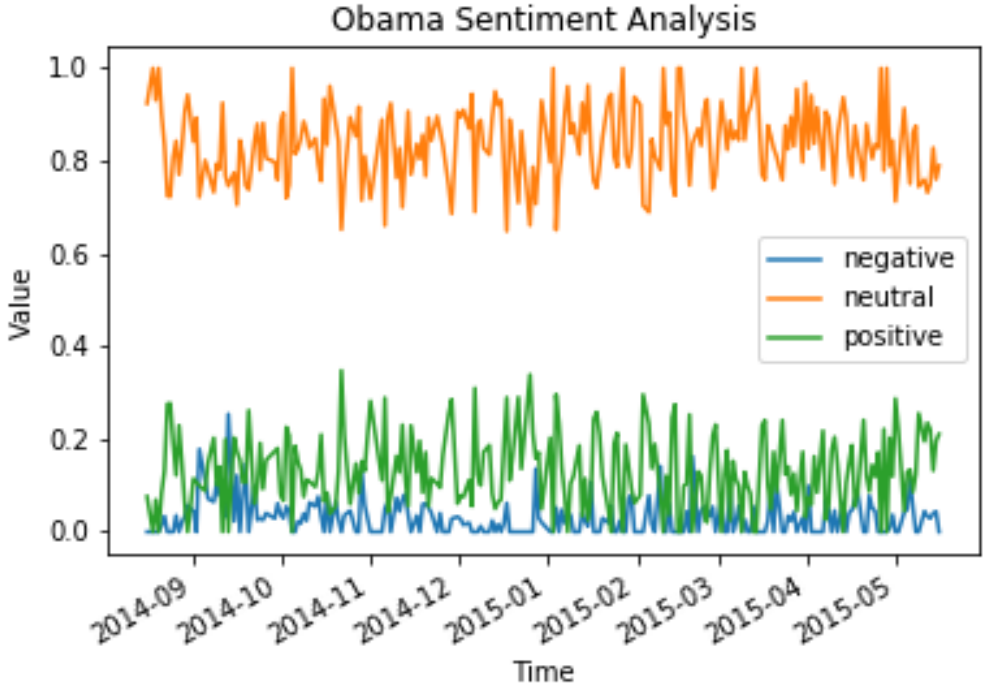
The VADER algorithm outputs sentiment scores to 4 classes of sentiments The sentiment score from 0 to 1 measuring the subjectivness of the text. 0 is objective, 1 is subjective. - Postive:The score of measuring positive words - Negative:The score of measuring negative words - Neutral:The score of neutral words - Compound: The aggregated score which is normalized of the sum of all scores .

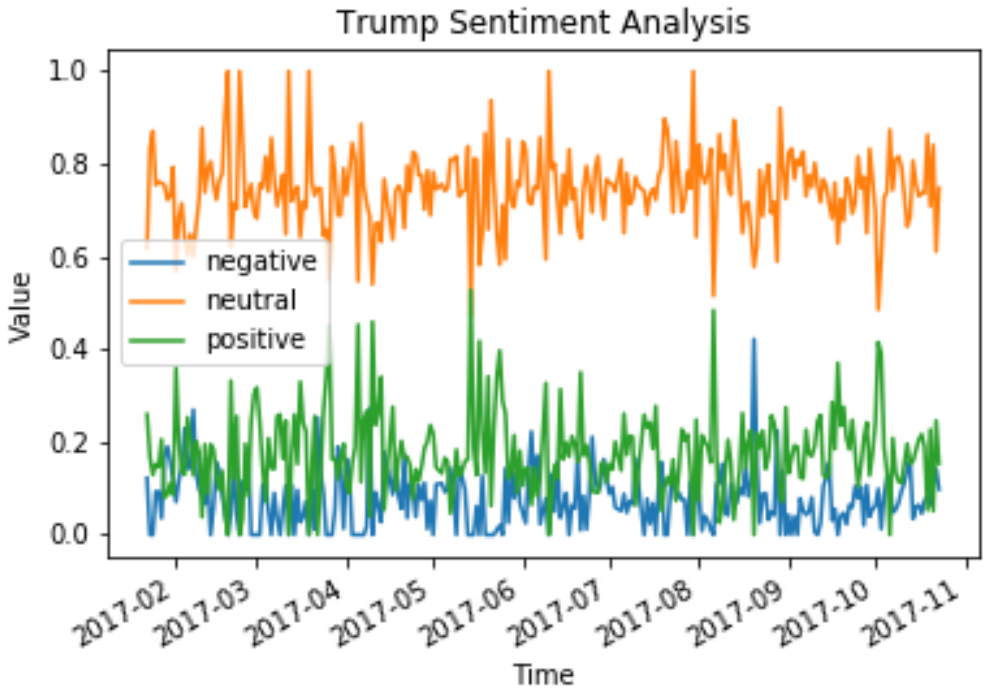
5.1 Visualization of the Sentiment Analysis

Postive and Negative Tweets of Both Presidents



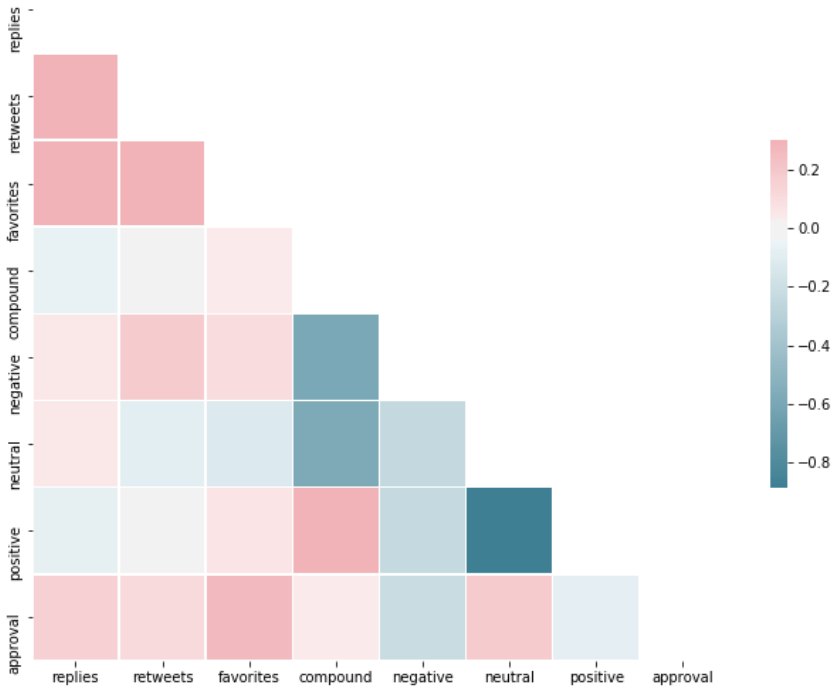
Change of Sentiment Over Time



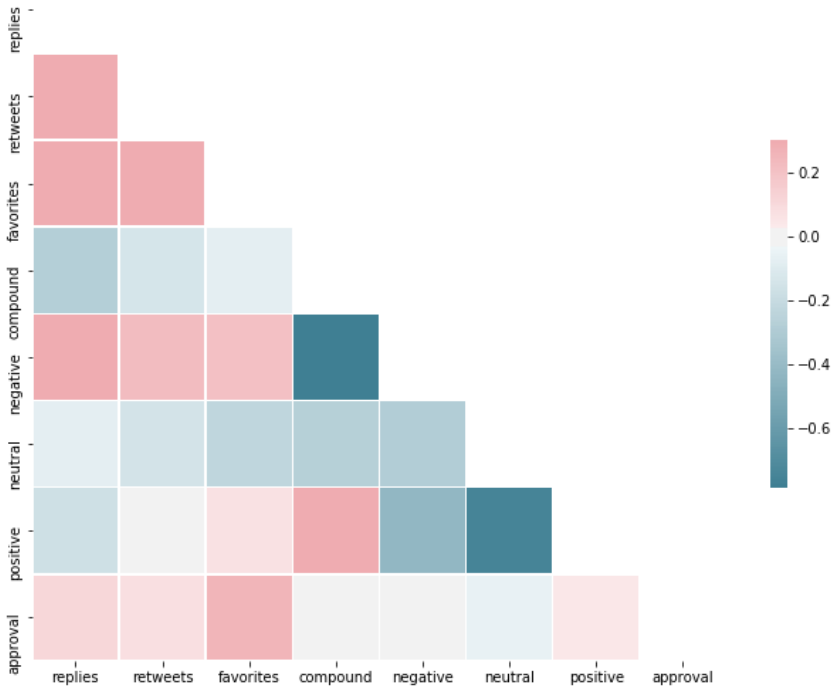


We can find the differences between the two presidents' sentiment based on the graphs above. But we can guess that President Trump is more likely an emotional person since he has more both negative and positive tweet than Obama, and the negative and positive curves fluctuated widely in the Trump sentiment analysis graph.

Correlation matrix of Sentiments of Obama



Correlation matrix of Sentiments of Trump



5.1.1 Summay of the Sentiment Analysis for Obama

	compound	negative	neutral	positive
count	231.000000	231.000000	231.000000	231.000000
mean	0.221540	0.031501	0.838107	0.130397
std	0.220897	0.038648	0.080143	0.080087
min	-0.624900	0.000000	0.648600	0.000000
25%	0.079325	0.000000	0.780350	0.075333
50%	0.214380	0.023762	0.836571	0.122000
75%	0.377946	0.049250	0.894833	0.188500
max	0.817600	0.254000	1.000000	0.348000

5.1.2 Summary of the Sentiment Analysis for Trump

	compound	negative	neutral	positive
count	269.000000	269.000000	269.000000	269.000000
mean	0.218289	0.076704	0.746446	0.176851
std	0.309979	0.062250	0.084951	0.089902
min	-0.921100	0.000000	0.472000	0.000000

25%	0.000000	0.025625	0.706500	0.123500
50%	0.219709	0.074200	0.747500	0.166600
75%	0.425900	0.110900	0.791200	0.223000
max	0.950100	0.422000	1.000000	0.528000

From the above results, we can see the compound value is for Trump is 0.218 while Obama is 0.221. The means of the postive and negative value of Trump's tweet are higher than Obama. Obama's mean of neutral is higher than Trump. Therefore, we can see Obama' tweets trend to neutral.