

John Halloran

For this progress update, this project is now using both Donald Trump and Hillary Clinton's twitter dataset. We can now make some comparisons, however, as of right now, each is treated rather separately. The focus of my data has been involving cleaning Hillary Clinton's data set, favorite words of Hillary Clinton, and looking at hashtag frequency over a period of time.

Cleaning the data follows a similar pattern from cleaning Trump's data. Clinton's dataset does not have separate categories though for date and time, but rather a "created_at" category. This involves splitting up the String for each instance. Generally this means finding the index of when the year ends as follows,

```
def findyear(date):
    index = 0
    while(date[index:(index+4)] != '2016'):
        index = index + 1
    return index+4
```

I have also updated the dataset to use in other sets, so I wrote to new csv files that would allow me to carry this data over to any other notebooks that I choose to use. For example,

```
#Writing the tweets to a new data file
donald_tweets.to_csv('Donald_Trump_TweetsFinal.csv', sep=',', header='infer',
encoding='utf-8')
hilary_tweets.to_csv('Hilary_Clinton_TweetsFinal.csv', sep=',',
header='infer',encoding='utf-8')
```

It is important to encode the data in utf-8. Apparently there was some errors in the dataset that created strange unicodes that did not fit in the context of the tweet. This prevents seeing a Euro symbol or any other abnormal characters.

Now that we have an actual dataset to use with Hilary Clinton, I followed a similar process before with Trump's data. Again following the example of Bonzanini[2], we were able

to extract Clinton's most commonly used words, hashtags, and "@" mentions. For her most commonly used words, they tend to be names,

[('Trump', 816), ('Hillary', 759), ('Donald', 438), ('president', 253), ('America', 208)]

To see different results, I eliminated words that included, "Trump," "Hillary," and "Donald."

The results were as such,

[('president', 253), ('America', 208), ('vote', 200), ('make', 197), ('people', 180)]

It seems quite concerning how she mentions her name and Donald Trump's name the most.

As for her mentions, she tends to tweet at

[('@POTUS', 176), ('@HillaryClinton', 107), ('@FLOTUS', 101), ('@timkaine', 100), ('@TheBriefing2016', 91)]

She tends to tweet at the President, herself, the First Lady, and her Vice Presidential candidate.

With the President and First Lady, she could tweet at them a bit more due to her connection to the Executive Branch as the Secretary of State under President Obama.

These are the hashtags she uses the most,

[('#DebateNight', 54), ('#DemsInPhilly', 48), ('#VPDebate', 46), ('#RNCinCLE', 36), ('#debate', 29)]

Her hashtag usage is incredibly low compared to Donald Trump's dataset. This may be because her dataset only stretches back to the beginning of 2016. We will see in a moment, however, that her distribution of her hashtags is connected to events. They are only relevant for brief periods of time.

The next part of is the plotting use of hashtags over the period of time that the datasets cover. Here are Donald Trump's top five hashtags over the time his dataset covers.

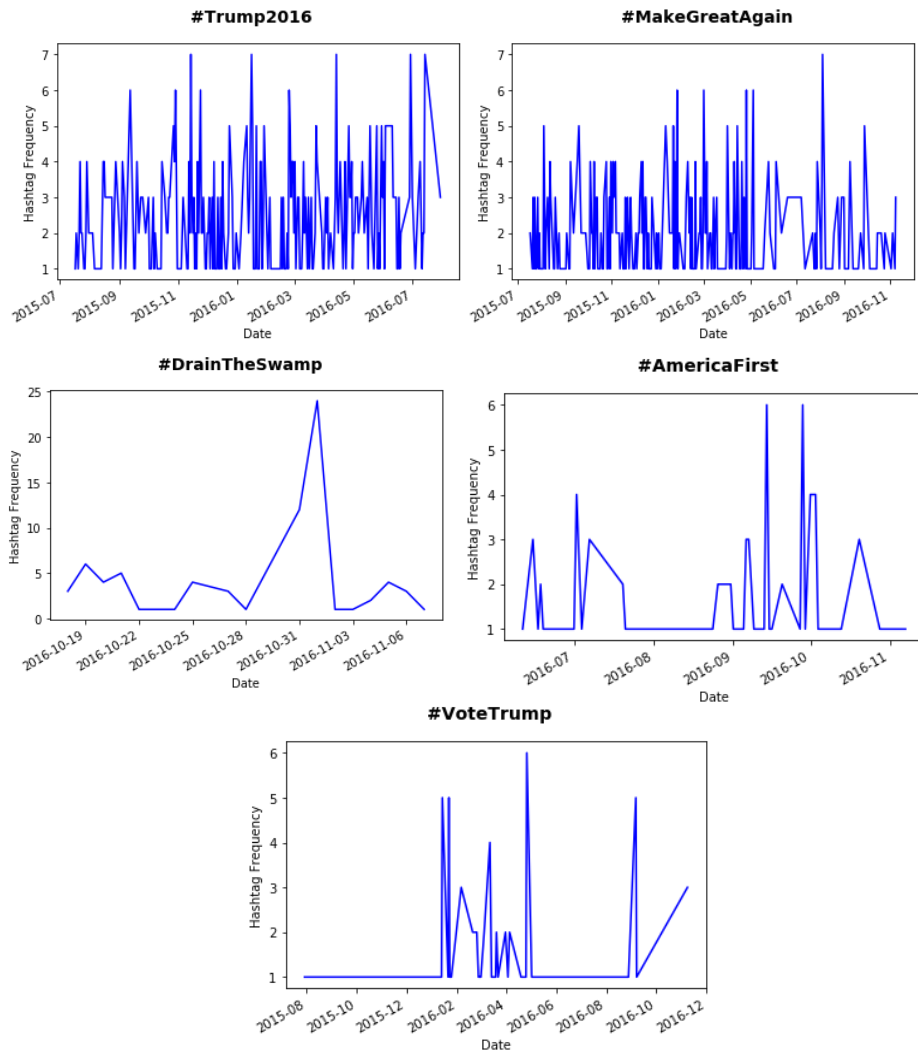


Figure 1 - Time Series of Trump's Hashtag

#Trump2016 and #MakeAmericaGreatAgain were definitely popular hashtags throughout the campaign season. This is probably because they are attached to the campaign. “Make America Great Again,” was Trump’s campaign slogan. The other three are a bit more interesting. They are not as frequent over a period of time. It was definitely popular for a specific period of time. #DrainTheSwamp refers to getting the government’s career and corrupt politicians out of Congress. A Slate article [5] describes in detail about the thoughts of Trump being an outsider considering he is a businessman and not a politician. This was also closer to voting day when

this hashtag was popular, probably referring to Hillary Clinton's shady dealings as a politician. The other interesting hashtag is #VoteTrump. This was most popular in during voting times in the primary season and then actual election day for the presidency.

Hillary has very different hashtags, and infact resonate more closely with #VoteTrump.

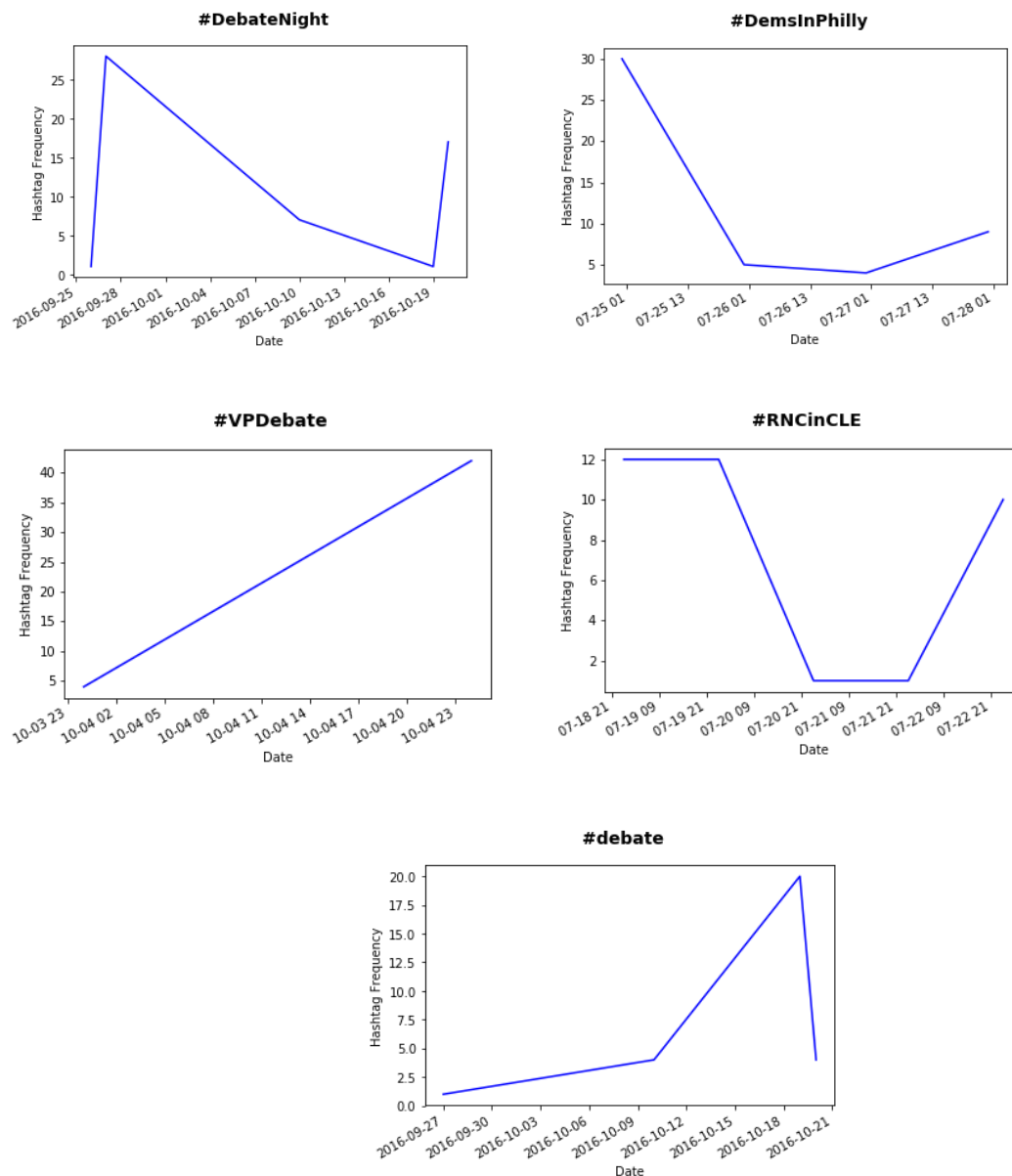


Figure 2 - Time Series of Clinton's Hashtag

Each of these hashtags live very short lives. #VPDebate, for example, was only used when Tim Kaine was about to debate. They increased in frequency per day leading up to the debate, but then stopped. It is similar pattern for each of these hashtags.

The main difference in these hashtag uses is that Trump actually wants to spread a slogan on social media. The hashtag most particularly, #DrainTheSwamp, might resonate with voters due to frustrations the public has had with the government. Clinton's hashtags are more like current event hashtags letting people know to pay attention to these debates or conventions that were happening at the times that Clinton was tweeting these hashtags.

In general from this data found so far, Trump may have been using social media more effectively. He was able to spread hashtags that would grab people's attentions. Clinton's tweets are not as trendy and would not go viral. Before the semester is over, my goal is to have a semantic analysis done as well as a comparison of the tweets against retweets and favorites to see what might resonate more with their audiences.

Works Cited

1. Alidina, Abbas. *Donald Trump's 11 Best Tweets (and 7,000 Losers)*. Retrieved February 14, 2017 from Crowdbabble:
<https://www.crowdbabble.com/blog/the-11-best-tweets-of-all-time-by-donald-trump/>.
2. Bonzanini, M. *Mining Twitter Data with Python*. Retrieved March 8th, 2017:
<https://marcobonzanini.com/2015/03/02/mining-twitter-data-with-python-part-1/>.
3. Ginns, Andrew. *Idp-twitter*. Retrieved February 14, 2017 from GitHub:
https://github.com/andrewginns/idp-twitter/blob/master/HillaryClinton_tweets.csv.
4. Halloran, John. *DataScience_Marquette*. Retrieved March 9, 2017 from GitHub:
https://github.com/johnnyguitar95/DataScience_Marquette/tree/master/Notebooks.
5. Kelly, John. *What's With All Trump's Talk About "Draining the Swamp."* Retrieved April 11, 2017 from Slate:
http://www.slate.com/blogs/lexicon_valley/2016/10/26/why_do_trump_and_his_supports_keep_talking_about_draining_the_swamp.html