

Lab #3: Social Media Data

Presidential Elections LAB

August 22nd 2017

Part 1: Social media!? Why do we care?

- In this lab, we will be working with data pulled directly from twitter. The most recent estimates show that Twitter alone has about 328 million active monthly users. Lucky for us, people tweet about everything! Although restaurant reviews and folks' opinions about "Game of Thrones" probably doesn't matter much for political scientists like us... We *can* use their opinions about candidates, and their policy preferences! In other words, it's a really cheap, and quick way to collect contextual information about elections and public opinion.
- Okay- so, the average Joe tweets! Big deal... But, guess what! So do Presidential candidates! Hillary Clinton and Donald Trump both utilized twitter during the 2016 election. As data scientists, we can collect these tweets and analyze them in R. For reference, Hillary Clinton's handle is: "@HillaryClinton" and Donald Trump goes by "@realDonaldTrump". What can this data tell us?

Part 2: The data...

Variable Name:	Definition:
text	The text of the actual tweet by the candidate
favoriteCount	The number of times a tweet has been "favorited"
created	The day and time of the individual tweet
retweetCount	The number of times a tweet has been Retweeted
location	What geographical location the tweet came from

- Note: These variable names will be the same for all of the datasets we work with today! We will have 3 in total!
- Jeb Bush's tweets are in the file called, "JebTweets".
- Donald Trump's tweets are in the file called "TrumpTweets".
- Hillary Clinton's tweets are in the file called "ClintonTweets".
- You already know how to import the JebTweets data (It's in your new script). Try to import Clinton's and Trump's now.

Part 3: Analysis...

- First, let's take a look at twitter data from Jeb Bush!

Jeb Bush	@JebBush
Tweet Dates:	December 2016 - August 2017
Tweets collected	189
Word used most often:	"great", "will", "educ"
Word used least often:	"Determined", "Environment", "Child"
News or Candidate Mentioned?	"Trump"

How do we know this? Check out your starter file! Run this first to see if you can get the same answers. Then fill out the first set of questions (Question 1)!

The code for this section is located in your new script: `posc149_lab3_starterfile.R`

Question 1:

- After running the script to look at the JebTweets data more closely, tell me the exact date and time range included within this data set.
- With the same JebTweets script: How many times is Trump mentioned? How many times is Clinton mentioned?
- Now, why might this be the case, given the information you have on hand?
- What tweet is most favorited? **HINT: You may need to use code you've already learned for this part!!!**

Question 2:

- Import Clinton's tweets (hint: look at how you got Jeb's tweets). How many tweets do we have?
- What is the exact date and time range that we're working with here?
- Why might this date range be important?
- What are the most and least words mentioned?
- Does she talk about healthcare? How do you know?
- Which tweet got the most retweets? **HINT: You may need to use code you've already learned for this part!!!**

Question 3:

- Import Trump's tweets (hint: look at how you got Jeb and Hillary's tweets). How many tweets do we have?
- What is the exact date and time range that we're working with here?
- Why might this date range be important?
- What are the most and least words mentioned?
- Does he talk about the military? How do you know?
- Which tweet is most favorited? **HINT: You may need to use code you've already learned for this part!!!**

Question 4:

- Look at the words "America" and "people" in context for all 3 candidates.

Example code::: (It's also in your script for today.)

```
kwic(JebCorpus,"America", valuetype = "regex", case_insensitive = T)
```

```
kwic(JebCorpus,"people", valuetype = "regex", case_insensitive = T)
```

- b. What do you make of this? What differences do you see?
- c. What can this tell you about the issues that each candidate wants to emphasize in their respective campaigns?