

Jocelyn Guerrero Lemus

***Setting up the environment:***

I analyzed an archive of Trump tweets using R. An excel sheet with all tweets that included the text, date, device, time, retweets, favorites, deleted, etc. was downloaded into R and then turned into a corpus (a collection of text). Through the quanteda library we first unitized our data sets.

***Preprocessing:***

I tokenized the corpus to clean the unit. Specifically we removed punctuations, numbers, symbols and urls, we also removed stop words and removed frequent and infrequent tokens, lastly stemming the text

***Topic Modeling:***

After preprocessing and cleaning the data I was able to create topic models through the quanteda library and calling the command topic model and extracting the topics (converting them) from the trump tweets data sets. We initially called for 10 topic models and they were shown in our output.

***Generate insights:***

Some significance we found from our word clouds included:

WB1, WB5: political focus and language; the tweets appear to be from 2016. Some of the major words include impeach, clinton, democrat and obama. There was interest in political rivalry.

WB2: positive interpretation; the bubble is alot smaller than others and features more leisure or happier words like, thank, birthday, love, happy. Seems to be tweets that were mostly used in salutary matters or celebration.

WB3, WB6: focus on Trump himself; looks to be tweets centering around his personal campaign. Major words include #makeAmericagreatagain, #trump2016, @foxnews, #maga. We can infer that twitter was an important avenue for his campaign because of the number of #'s used in this bubble.

WB4: focus on foreign affairs; talks about the border, Mexico, illeg, immigrant, obama, iran, iraq, kill, military, wall. Although there were enfaces on other foreign affairs the biggest words were about border with Mexico affairs. He appers to focus mostly on illegal immigration. WB7: focus on economy; specifically on china, trade, american products, deals, trade war, unemployment. Was one of the bigger bubbles

WB8: focuses on the media, words like, fake news, media, news, report, @foxnews, @nytm, @CNN. Although a concrete topic it is on the smaller side.

WB9 focuses on himself but as a private citizen. Seems to be from before his run for the presidency or not politically inclined.

WB10: focuses on domestic matters, He seems to be in office. The words he focuses on are military support, work, american, fight, love, nation, @whitehouse

### ***Interesting questions:***

Q1) Why are some bigger and some smaller?

- Could we infer the level of importance or frequency for each topic?
- How would we test that?

Q2) What does it mean when some words repeat?

- Are they used in different contexts?
- Are they frequently used?

Q3) What do these topics tell us about Trump's personality?

- Does his personality change over time?
- Why does “thank” so large in his word bubbles?
- 

Q4) What can these word clouds tell us about Trump's political agenda?

- Specifically WB10

Q5) Are there any intersections between the word bubbles?

- What does this mean?

### ***Hypothesis***

Based on my analysis I have identified that the word “thank” is found in multiple topics. Seeing how Trump has the reputation of being selfish or self centered this word stands out. Further analysis such as a sentiment analysis or a more in depth look into his tweets could showcase a more in depth look into his personality.

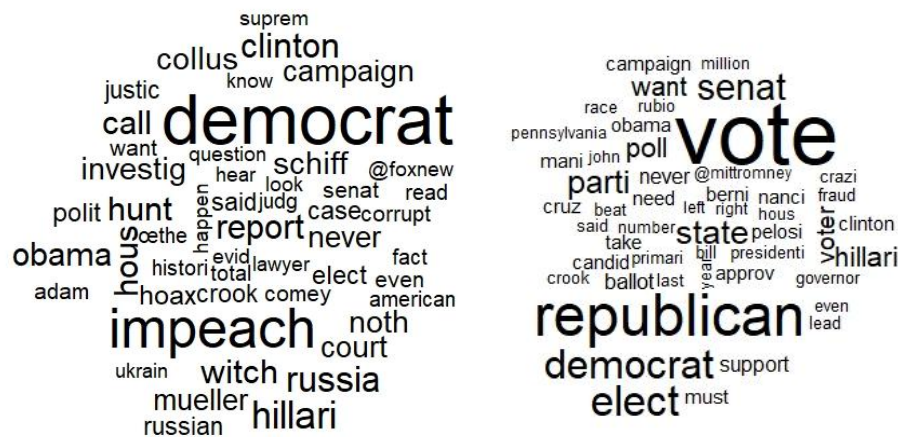
H1: Trump will showcase a more positive attitude towards his supporters and republicans

H2: Trump will use the word “thank” to generate support from others

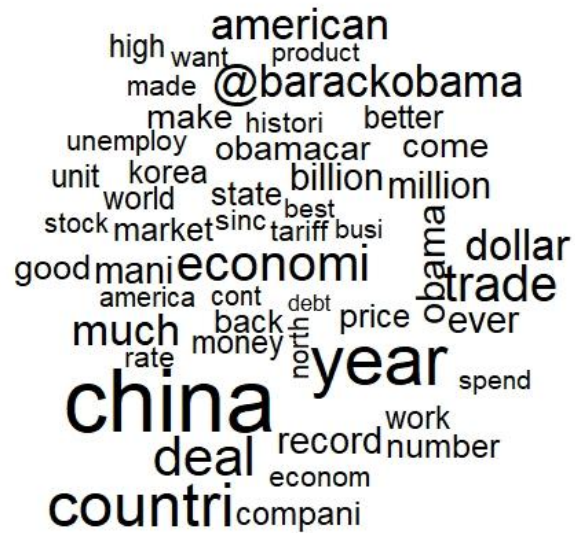
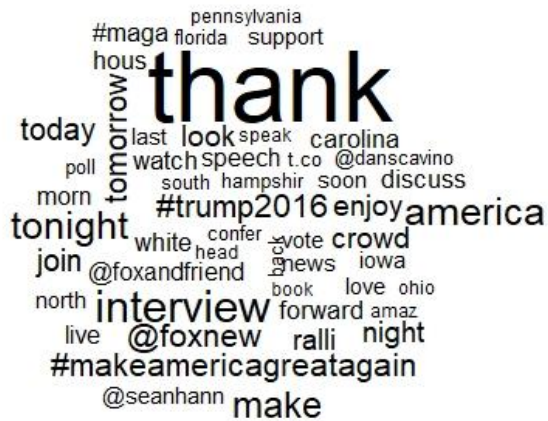
Looking back at Trump’s presidency his disdain for news organisations appears obvious. However the topic of “fake news” was present only in a small world cloud. Additionally, the vocabulary surrounding news did not appear in other topics. A more in depth analysis is needed to understand if Trump did not tweet about news organisations as much as it appeared at the time.

H2: Trump uses the same vocabulary when referring to news organisations.

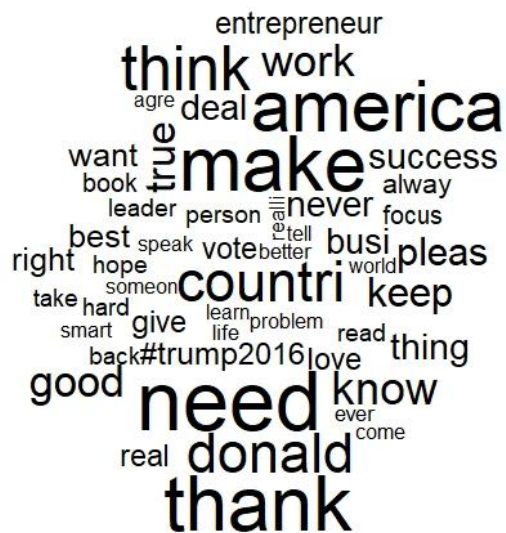
**Word bubble 1 & 5**

[illegible]

**Word Bubble 3 & 6**



**Word Bubble 7**



*Word Bubble 8*



*Word Bubble 9*



**Word bubble 10**

