Congressional Tweet Analysis

SANJAY SHAH

Project Proposal

Project Goals

▶ Understand which politicians' tweets have been the most popular and the factors driving the popularity.

▶ Utilize results to help lobbyists decide where to focus their resources.

Approach

▶ Analyze data set of Congressional tweets from 2008-2017

▶ Import data into Python Pandas DataFrame, clean the data, perform descriptive statistics, look for correlations, conduct text analysis including LDA modeling
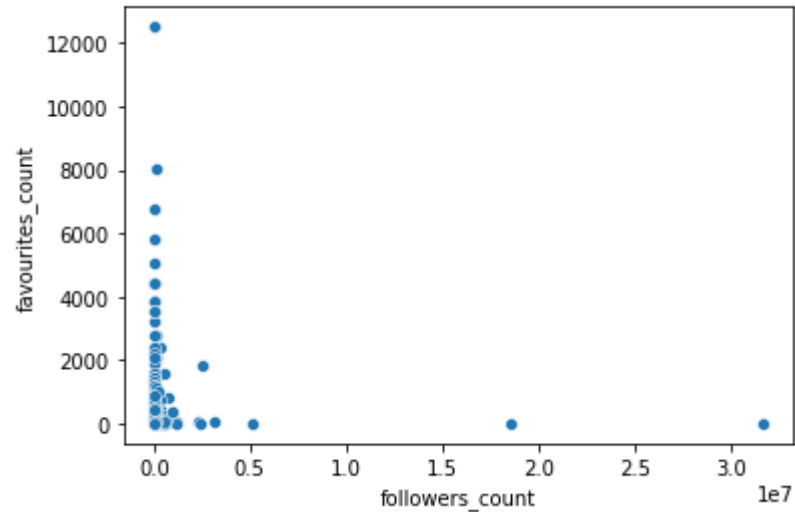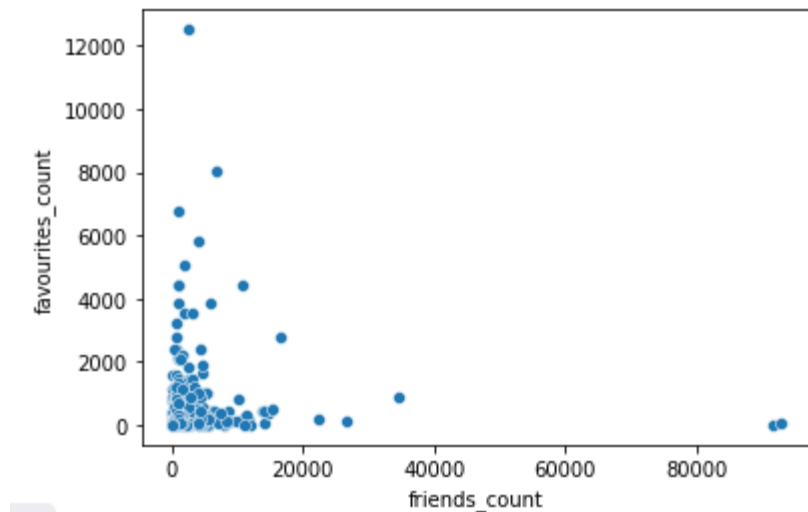
Hypotheses

▶ There are a group of politicians in the data set that tweet significantly more than others.

▶ There is a positive correlation between the number of friends a user has and favorites.

▶ There is a positive correlation between the number of followers a user has and favorites.

▶ Tweets with positive sentiment are retweeted more often

Descriptive Statistics
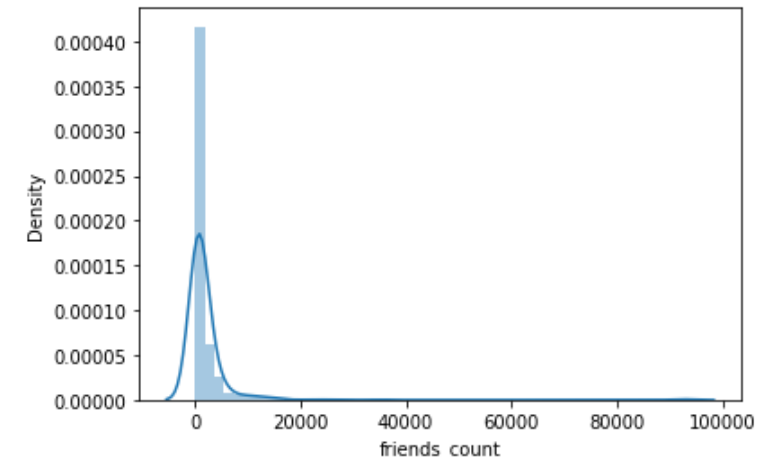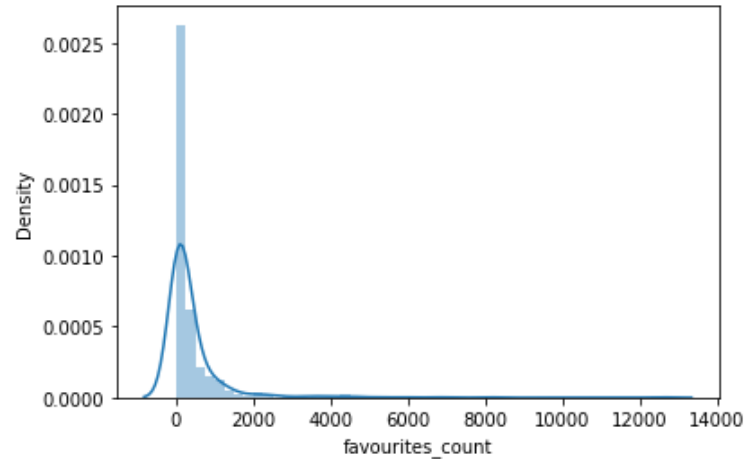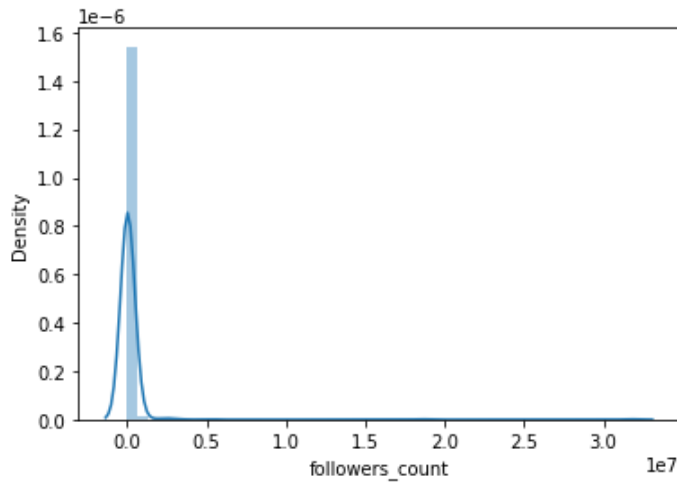
- ▶ The count of followers, favorites, and friends all follow normal distributions
- ▶ The distributions are skewed right by some very popular users
- ▶ Understanding what influences their popularity will be key

Friends and followers don't mean more favorites

- There is no correlation between the favorites count and friends or followers

- This disproves our initial hypotheses that the number of friends and followers a user has significantly impacts the number of favorites

- The actual content of the tweet may have more influence; we will investigate later

Positive sentiment has some impact
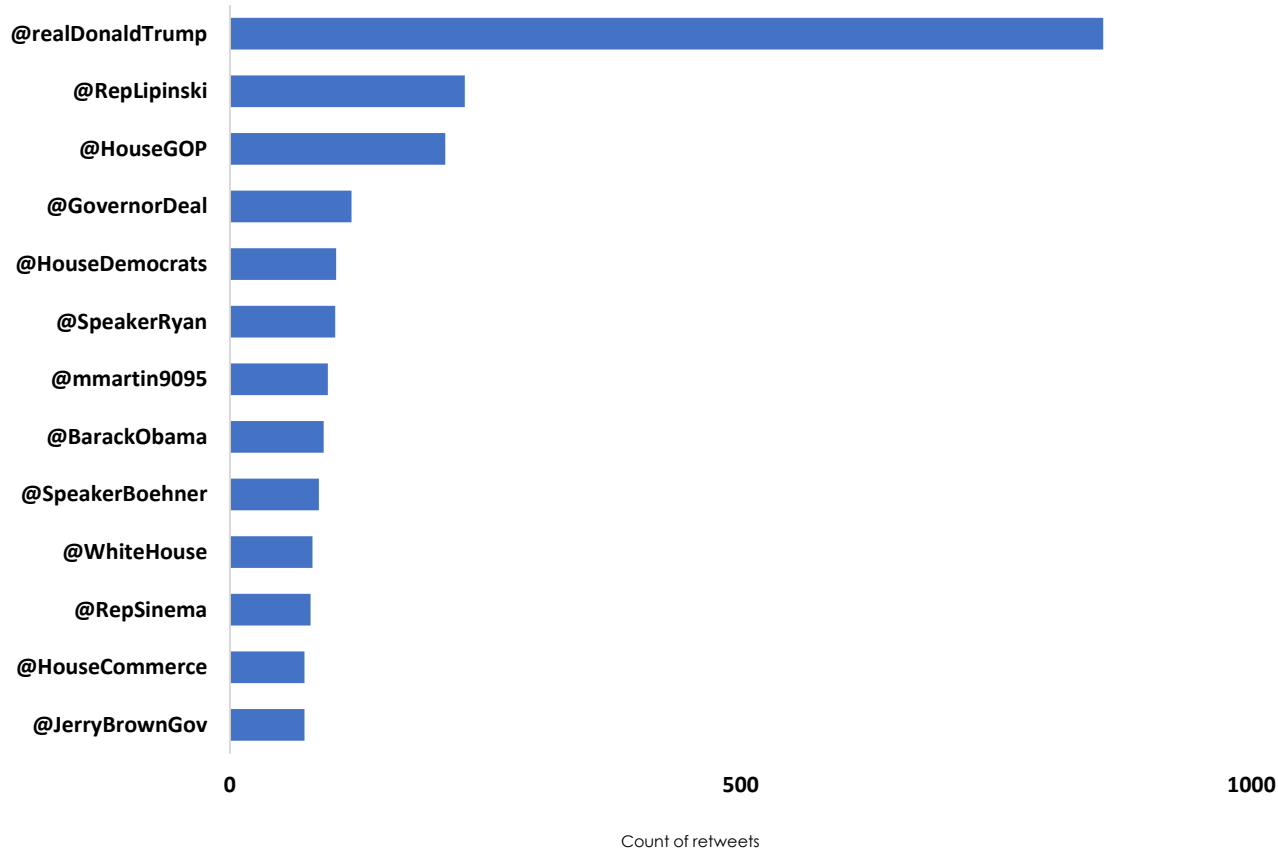
Retweets vs sentiment
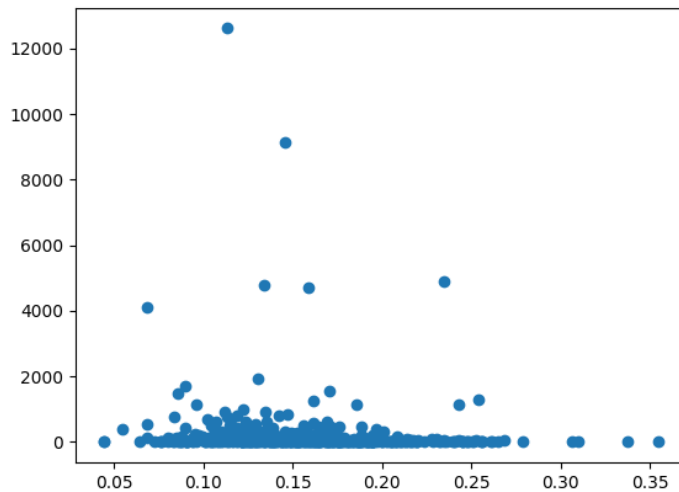
Favorites vs sentiment
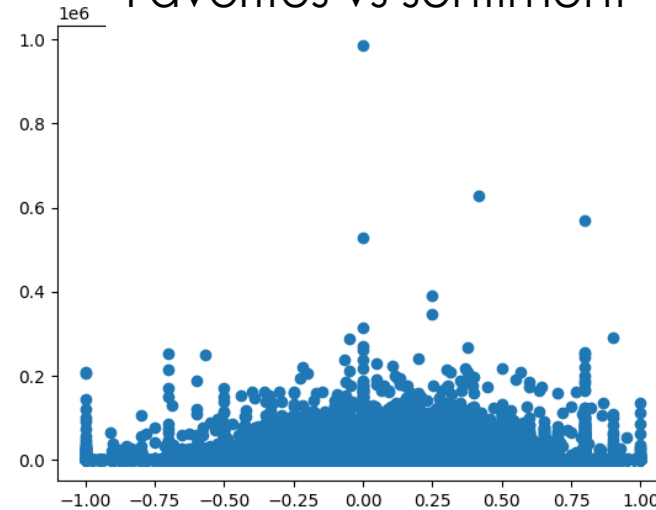
▶ Sentiment (as measured by polarity) seems to have a mild correlation with the number of retweets and favorites

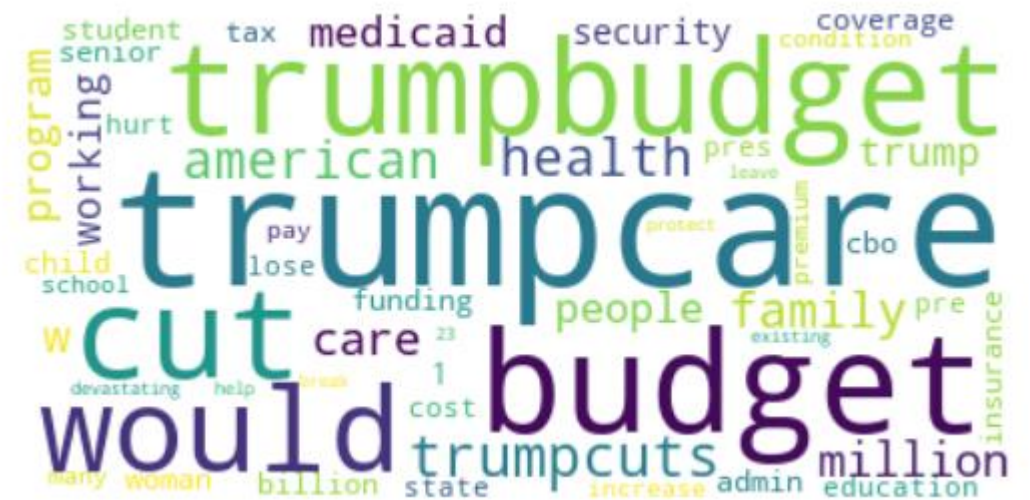▶ Polarity greater than 0 indicates positive sentiment

Trump dominates retweets

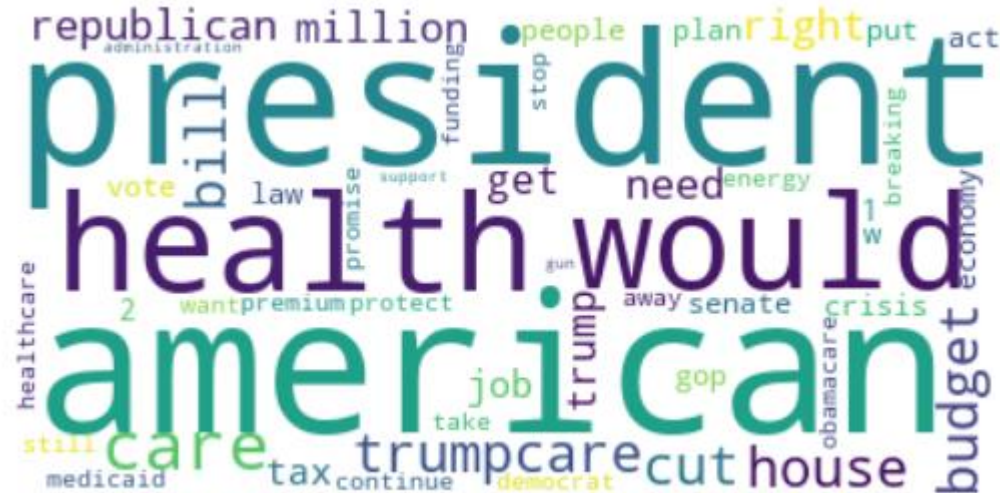## Who is retweeted to the most?

Count of retweets

- ▶ Trump clearly dominates retweets
- ▶ What topics drive the number of retweets?
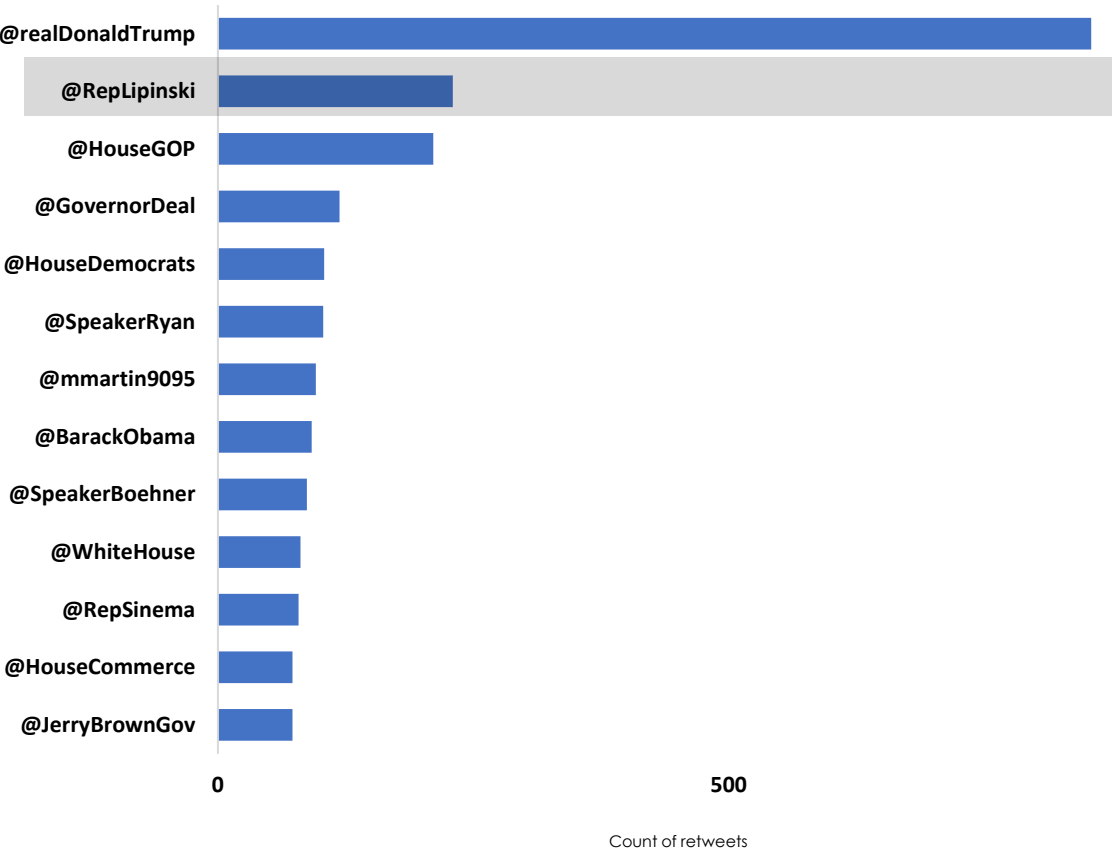
Word Cloud of tweets mentioning Trump

Topic analysis of most popular tweets

Word Cloud of Most re-tweeted tweets

Topic analysis of most popular tweets
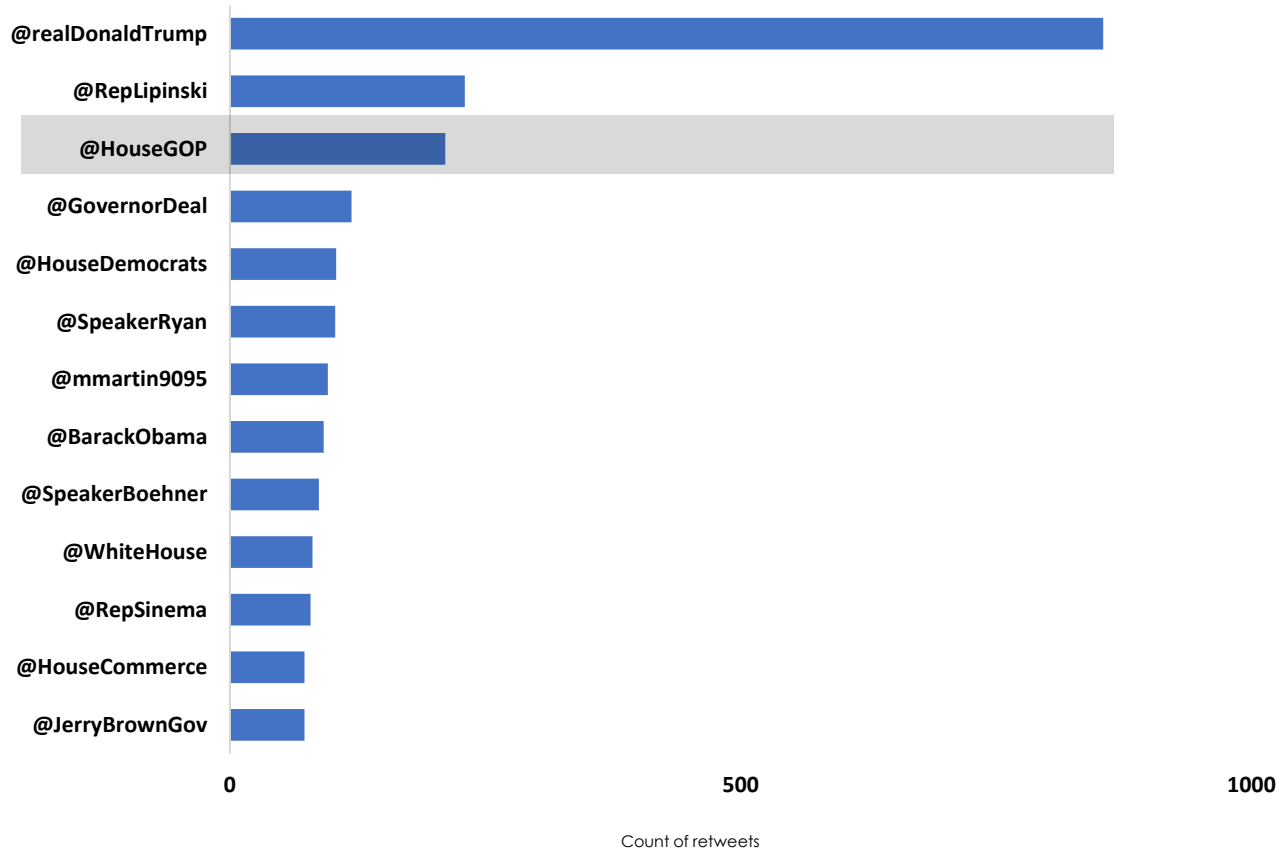
Who is retweeted to the most?

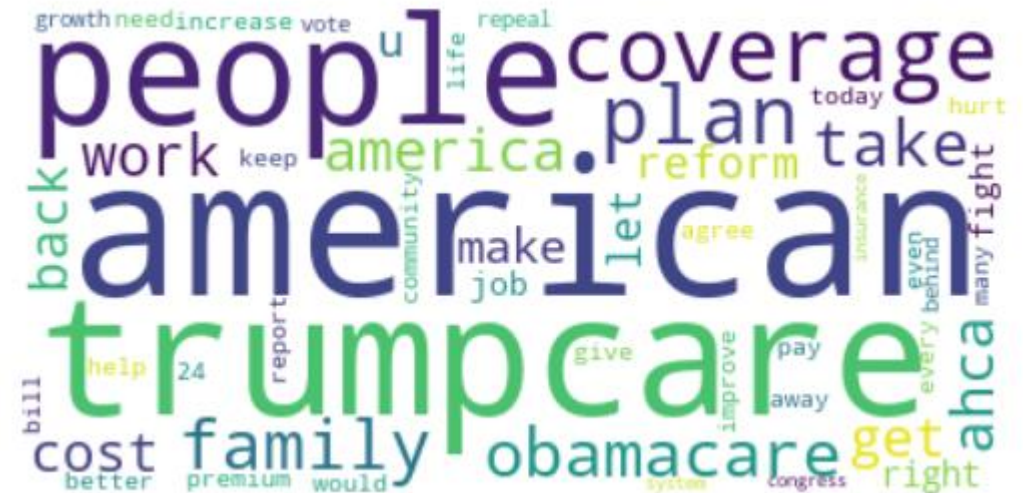## Word Cloud of @RepLipinski Tweets

Topic analysis of most popular tweets
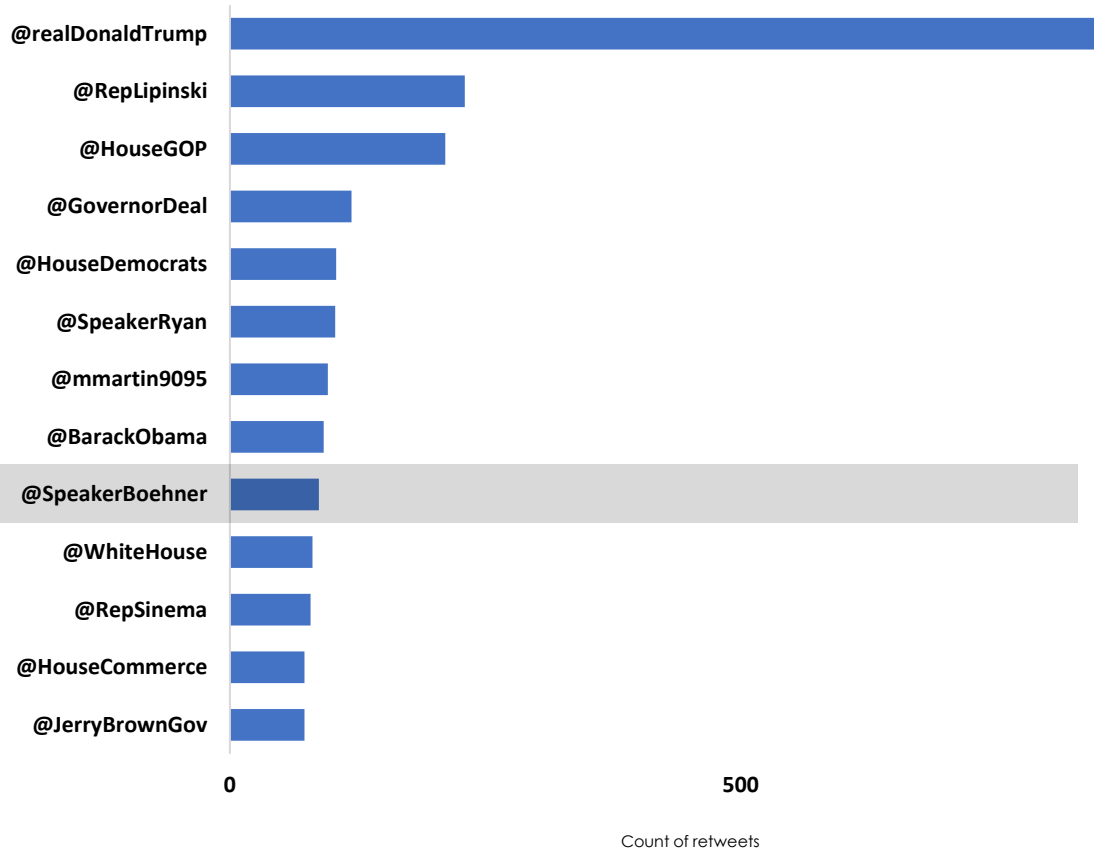
Who is retweeted to the most?

Word Cloud of @HouseGOP retweets

Topic analysis of most popular tweets
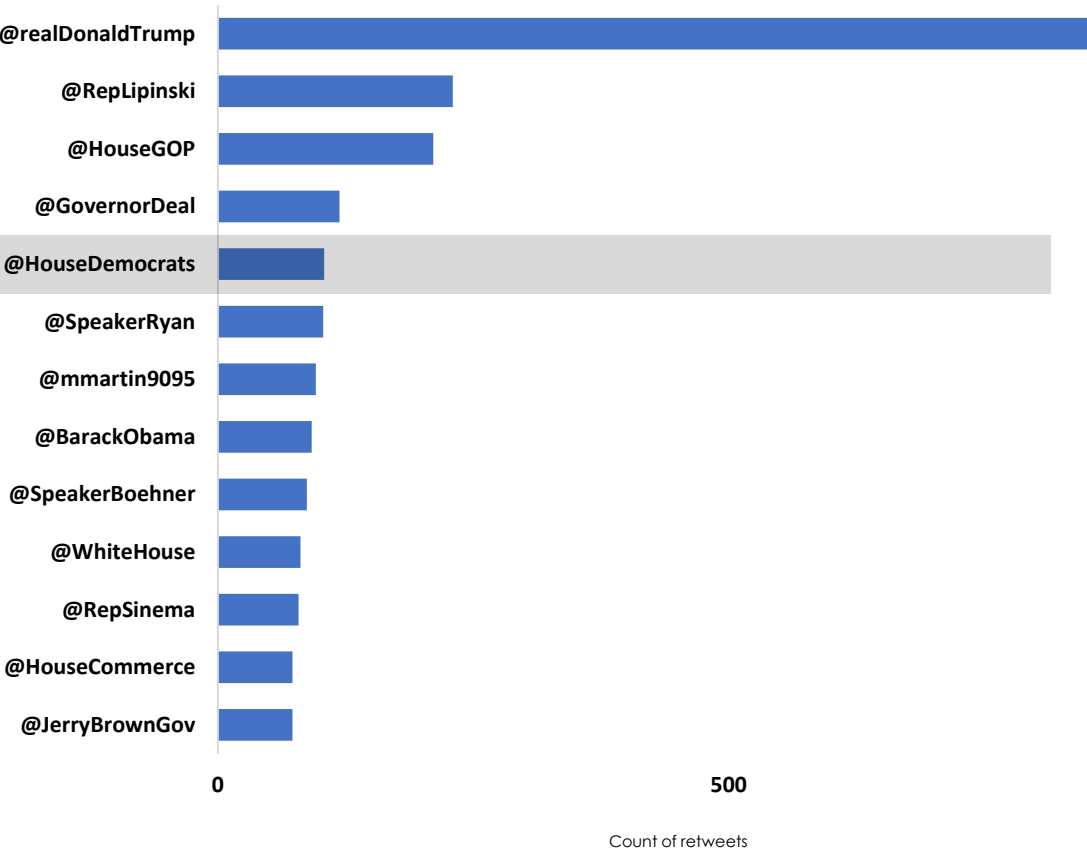
Who is retweeted to the most?
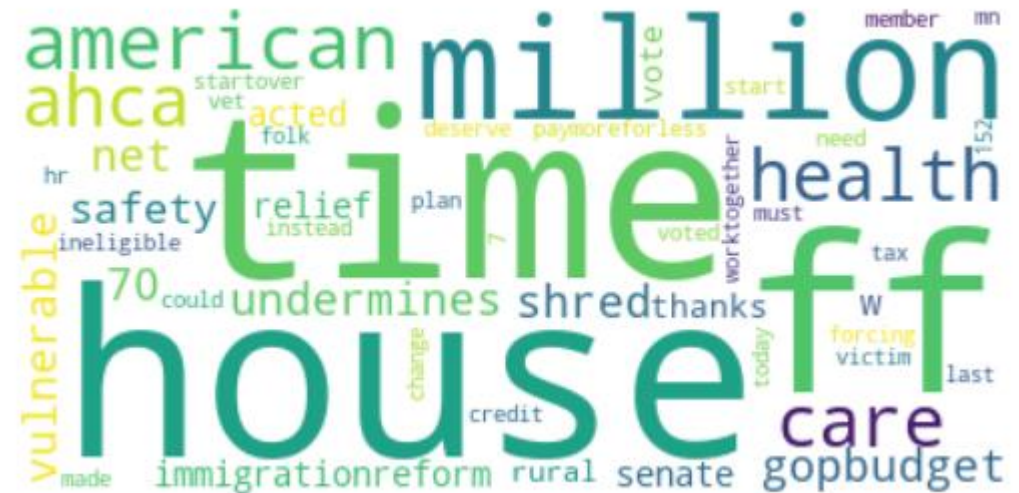

Word Cloud of @Boehner retweets

Topic analysis of most popular tweets

Who is retweeted to the most?

Word Cloud of @HouseDems retweets

Insights

- Our hypotheses about the number of friends and followers influencing popularity were disproved
- Our hypotheses about positive sentiment influencing tweets was confirmed, but we noted that the correlation is only mild
- A new hypotheses that the actual topic and content of the tweet was created and confirmed
- Trump was a wildly popular tweet topic and the subject of many retweets
- Discuss your hypotheses and any direct outcomes from whether you were right or wrong. Did you change your hypotheses? Or create new ones?
- The count of who is retweeted to the most was created to guide us on where to focus the topic analysis
- An LDA model in Python was constructed to analyze the topics of the most popular tweets; word clouds were generated from the results and are a good "metric" for textual analysis
- Healthcare was a recurring theme in the most popular tweets/retweets.
- Some, but not a lot of overlap in the topics from the most popular retweets

Recommendations / Actions

- The topic and content of a tweet are the most important drivers of a tweet's popularity (retweets and favorites)

- Positive sentiment may result in more popularity, so if possible be positive

- There are some common themes between the most popular retweets (e.g. Healthcare and Trump).

- An LDA model was developed to study the topics associated with the most popular tweets

- Further analysis with the LDA model could determine if we can predict the potential popularity of a tweet based on its content/topic. This could be used to help test tweets out first before posting them. The result could be increased popularity.

- We would need to test our model over different time frames; our hypothesis is that LDA topic analysis of tweets from the past few months may be a good predictor for how popular a tweet will be