SQL for Data Analysis Capstone – Political Discourse Through Twitter

By Tyler Sims

Description

The Political Discourse Through Twitter project analyzes twitter data associated with elected officials from various United States Political Offices. The project aims to provide actionable insights on key political topics, elected officials, and relationships between members of congress. The insights from this research will be used by Lobbyist4America to strengthen and focus their lobbying efforts. The research project will also be available to the general public for anyone interested in political communications and relationships.

Questions

There are three primary questions this project seeks to answer. These questions are designed to gain insights valuable to lobbying, election, and research efforts.

1. **Who is most active on twitter?**
   a. What political office is most active?
   b. What political party is most active?
   c. What are the general sentiment trends amongst political parties?
2. **Who are congress members communicating with the most?**
   a. Who are the most mentioned congress members?
   b. What are the most mentioned usernames not associated with congress?
   c. What are the sentiments surrounding communications?
3. **What are the key topics congress members are discussing?**
   a. What are the most used words?
   b. What are the most used hashtags?
   c. What are the sentiments surrounding these topics?

Hypothesis

Tweets associated with communications between politicians of opposing parties, on average, will be more negative than general tweeting trends. Sentiments around political agendas will follow party lines.

Strategy

This project joins raw data associated with congress from Twitter, data provided by the United Stated Congressional Website, and independent research to link tweets to politicians, political offices, and political parties. It then uses aggregation functions, sentiment analysis, and time trend analysis to gain valuable insights on political trends, communications, and polarity.

A quick look at the general statistics of the data helps frame the entire project. There are 530 twitter accounts in observation with 529 unique individuals. While this does not encompass all politicians in office during the observed time the project assumes this is the true population of politicians active on Twitter. The only notable politician missing from observation is former President Barack Obama.

Democrats and Republicans count for over 99% of observed politicians. This project will only recognize Democrats and Republicans during party specific analysis.

Technical Challenges

Sentiment analysis is difficult to do. NLTK and TextBlob are two popular python packages that can do sentiment analysis. The project conducted sentiment analysis using both, tested results against a human eye, and then chose to use NLTK for the project.

To perform the test, we first use standard practices to clean the text of each tweet using NLTK. We then run sentiment analysis on the cleaned and uncleaned text using NLTK and TextBlob. After the sentiment analysis is complete, we pulled fifty random tweets, determined true sentiment with human judgement and compared the true sentiment to the NLTK and TextBlob sentiments.

Sentiment Analysis	Positive	False Negative	False Positive	False Neutral
TextBlob	58%	2%	20%	20%
NLTK	66%	6%	22%	6%

While NLTK outperformed TextBlob there is still noticeable room for improved accuracy. Both tests tend to generate false positives, however, TextBlob is significantly more likely to generate a false neutral. Because of this sentiment test, the Political Discourse Through Twitter project could be improved in future analysis by improving the text analysis programs used in the study.

Analysis

General Statistics

Party Representation

Party	Observed
Republican	292
Democrat	232
Independent	4
Libertarian	1
None/ President	1

Between the two major parties, there are more Republicans using the platform producing a larger quantity of Republican tweets. Democrats produce more tweets per politician. Senators are the most active political office followed closely by Governors.

Tweets Over Time
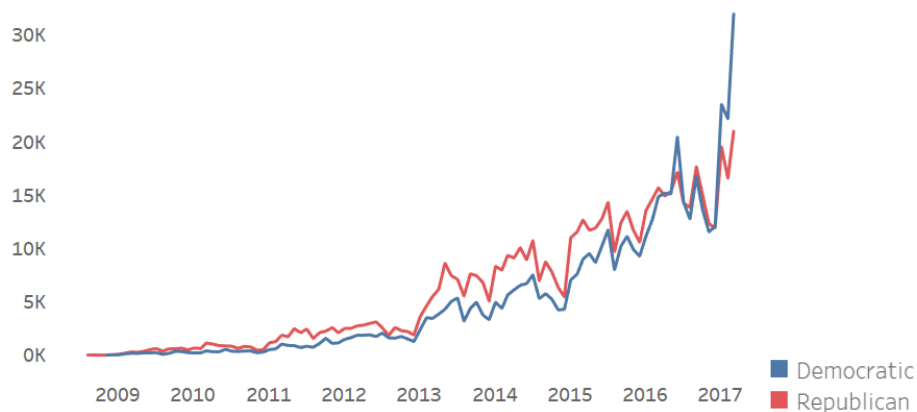
Twitter is becoming more popular amongst politicians. Each year more tweets are produced than the year before. This trend will likely continue as the United States population becomes more dependent on social media for news and general communication.

Tweet Statistical Analysis

|	Total Tweets	Mean	Min	Max	STD
Republican Party	657533	2251	27	3250	1061
Democratic Party	565600	2437	16	3258	965
Representative	842215	2181	16	3258	1046
Senator	272400	2808	305	3252	925
Governor	116402	2586	235	3250	925
President	3879	3879	3879	3879	-
All Tweets	1234896	2329	16	3258	1027

There is no huge variation between political party or political office tweeting trends. The only exception is the president. In this table Donald Trump's account is represented in the President's row. Trump is the first US president who started his presidency with a Twitter account.
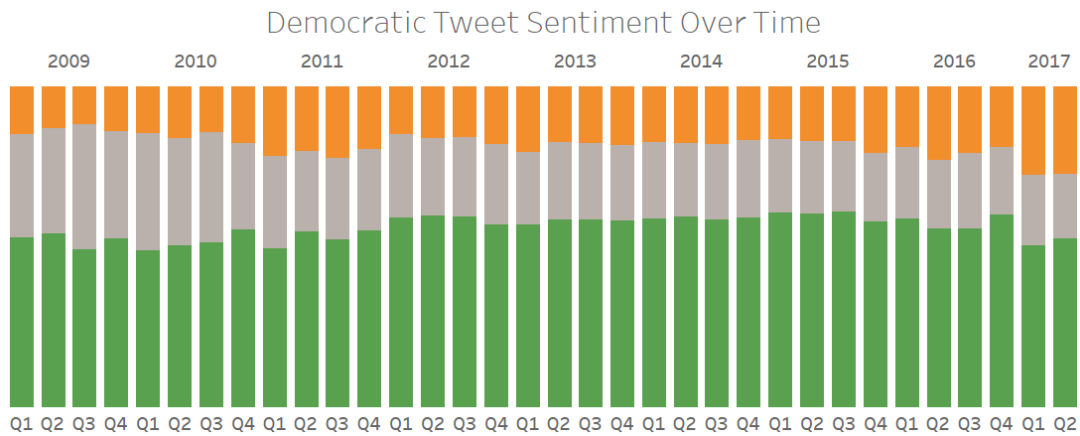
Sentiment

Sentiment Breakdown

| Sentiment	Percent of Total Tweets
Negative	19%
Neutral	24%
Positive	56%

Looking at all tweets in the observation, the majority tend to be positive. This is not surprising as a large role for political communications is for politicians to rally people behind a cause.
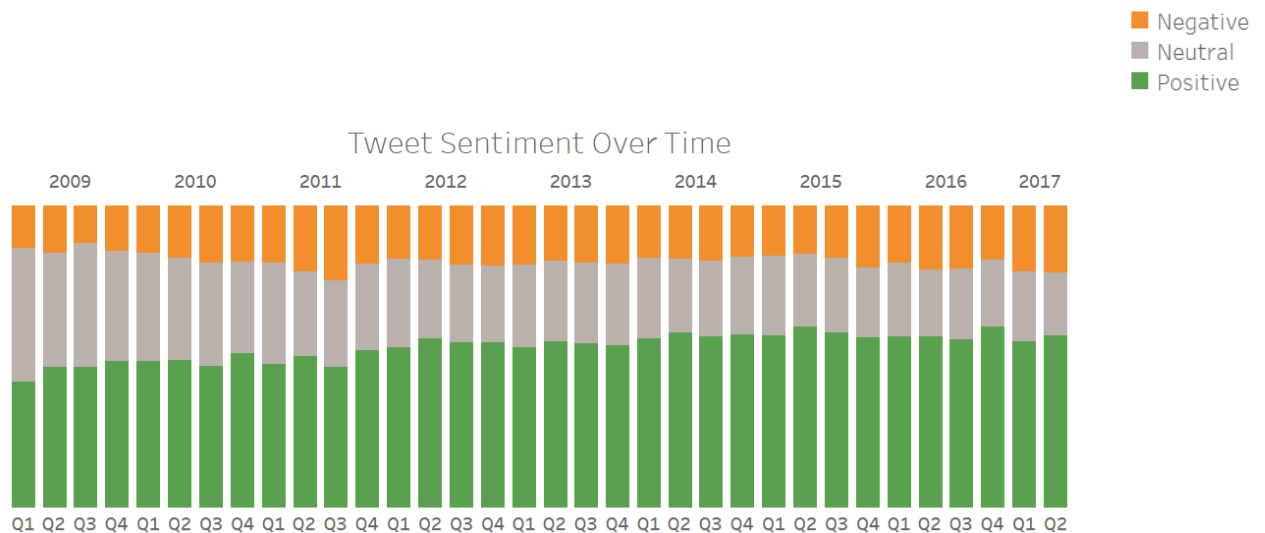
This chart shows that over time tweets have become more polarizing. Most tweets are becoming more positive, but more tweets are becoming negative as well.

Democratic Tweet Sentiment Over Time

This chart shows that over time Democrats have become more polarizing. Until the arrival and election of Donald Trump Democratic sentiment was trending positive.

Republican Tweet Sentiment Over Time
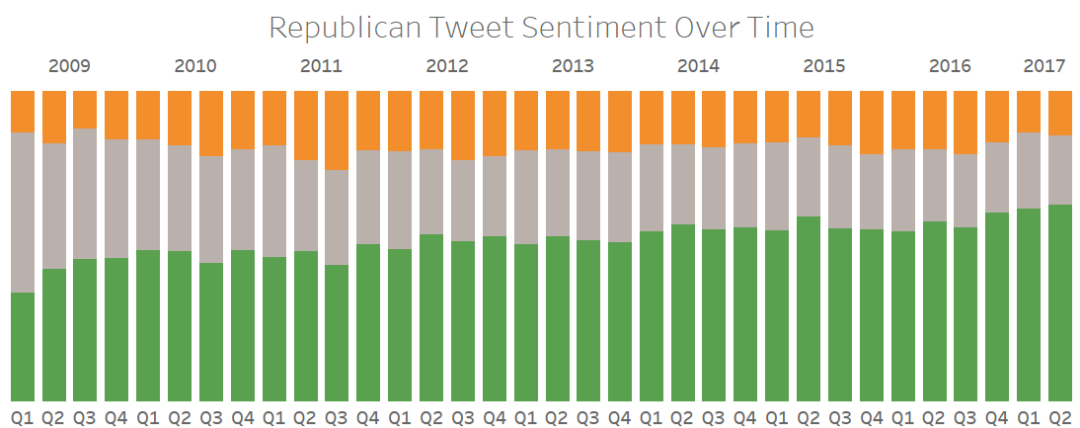
This chart shows that over time Democrats have become more polarizing. Until the arrival and election of Donald Trump Democratic sentiment was trending positive.
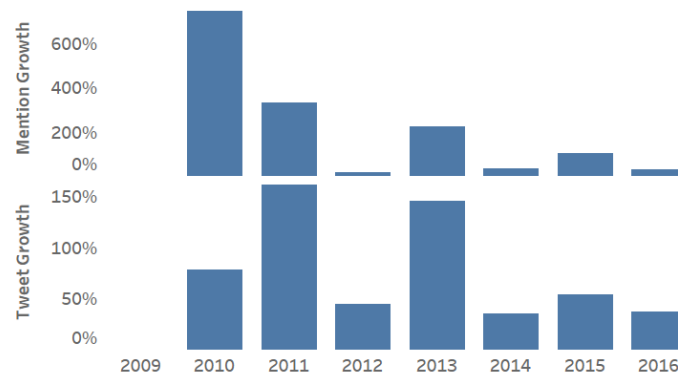
Mentions

To understand how communication within congress work, we analyze mentions on twitter. We will look at mentions of usernames not associated with members of congress and those associated with members of congress separately.

Mentions Over Time
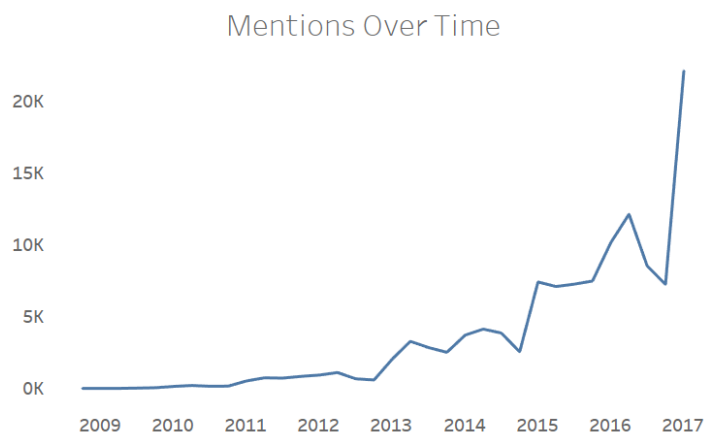
As social media and the internet becomes the primary source of communication for the public, we see politicians are directly communicating with other twitter accounts more and more.

Change in Mentions and Tweets 2010-2016

Mention growth rates far exceed the tweet growth rates year over year.

Mention Not In Congress.

For this study we will split all mentions into two groups. Mentions of accounts not associated with congress members and mentions of accounts associated with congress members.

Democrat Mentions Not in Congress

Screen Name	Count	Negative	Neutral	Positive
HouseGOP	7455	47	21	32
HouseDemocrats	4007	25	24	51
WhiteHouse	3111	19	24	58
OversightDems	1945	30	37	33
MSNBC	1843	32	40	28
EPA	1258	31	17	52
OfficialCBC	1248	18	38	44
DHSgov	1041	26	21	53
nytimes	1040	29	26	45

This figure shows the top ten mentions for Democrats. It includes total count and the percent of negative, neutral, and positive tweets associated with the mention.

Republican Mentions Not in Congress

Screen Name	Count	Negative	Neutral	Positive
HouseCommerce	3591	14	34	52
HouseGOP	3460	18	32	49
FoxNews	3457	31	40	29
GOPoversight	2942	29	42	29
WaysandMeansGOP	2710	21	30	49
HouseJudiciary	2251	25	38	37
HouseAppropsGOP	2080	14	35	50
WSJ	1973	33	33	34
FoxBusiness	1596	27	41	32
FinancialCmte	1591	25	35	40

This figure shows the top ten mentions for Republicans. It includes total count and the percent of negative, neutral, and positive tweets associated with the mention.
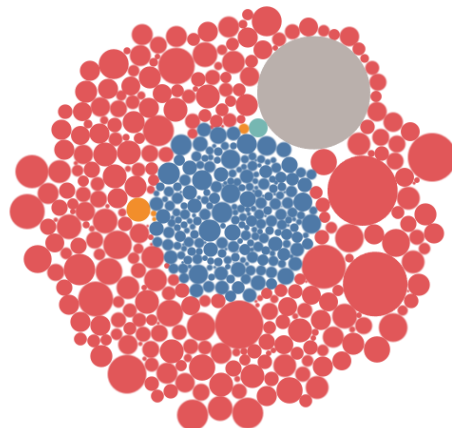
Mentions In Congress
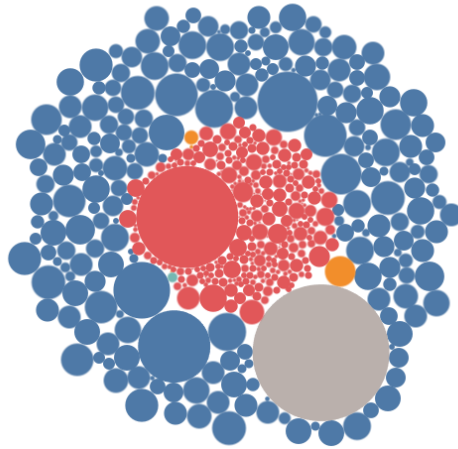
This figure shows all mentions associated with congressional members for Democrats. Democrats mention the president and Donald Trump more than any other account. This is a new phenomenon as Donald Trump redefined how presidents communicate with the country through twitter.

Democrat Mentions in Congress

Screen Name	Party	Count	Negative	Neutral	Positive
POTUS	None	8742	29	18	53
realDonaldTrump	Republican	4856	44	18	38
RepCummings	Democratic	2545	27	33	39
NancyPelosi	Democratic	2522	29	23	48
repjohnlewis	Democratic	1906	24	20	56
WhipHoyer	Democratic	1534	21	21	58
PramilaJayapal	Democratic	1379	8	27	65
SenWhitehouse	Democratic	1210	15	24	61
CoryBooker	Democratic	1206	17	29	54
RepJoeKennedy	Democratic	1138	15	24	60

When communicating with members of congress, Democrats primarily communicate with members of their own party. Specifically, the Democrats on this list are all Democrats with key leadership positions within congress. Democrats also tend to disproportionately mention Donald Trump with negative sentiment.

## Top Republican Contacts In Congrees

This figure shows all mentions associated with congressional members for Republicans. Like Democrats, Republicans mention the president and Donald Trump more than any other account.

Republican Mentions in Congress

screen_name	mention_party	count	negative	neutral	positive
POTUS	None	5265	24	22	54
GOPLeader	Republican	2289	17	35	48
realDonaldTrump	Republican	1917	14	23	63
LarryHogan	Republican	1578	18	30	52
RepEdRoyce	Republican	1387	22	36	42
SenRonJohnson	Republican	1336	14	55	31
RepHalRogers	Republican	1321	17	28	55
DarrellIssa	Republican	1240	19	33	48
KYComer	Republican	1173	4	35	61
RosLehtinen	Republican	1094	14	32	54

When communicating with members of congress, Republicans, like Democrats, primarily communicate with members of their own party. The Republicans on this list, like with the Democrats, are all hold key leadership positions within congress. Unlike Democrats, Republicans tend to mention Trump in a more positive sentiment compared to their average.
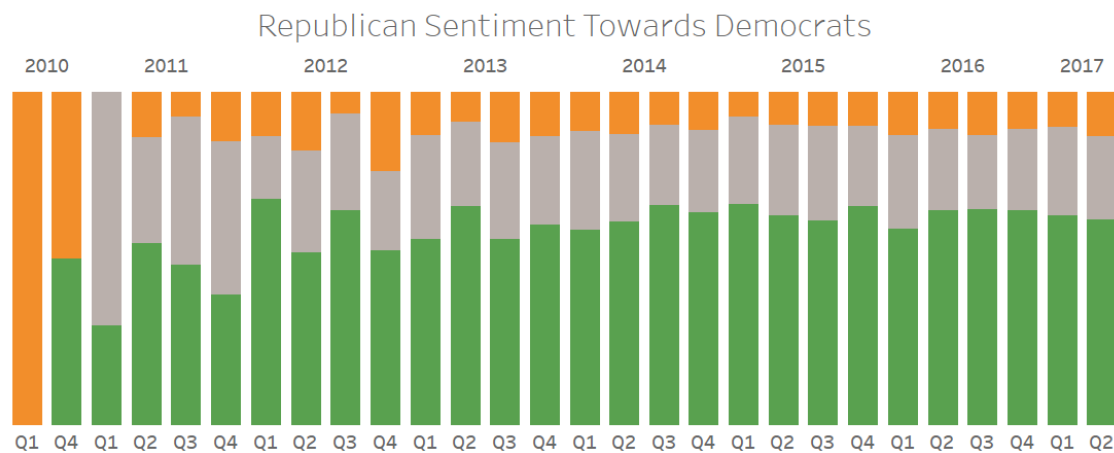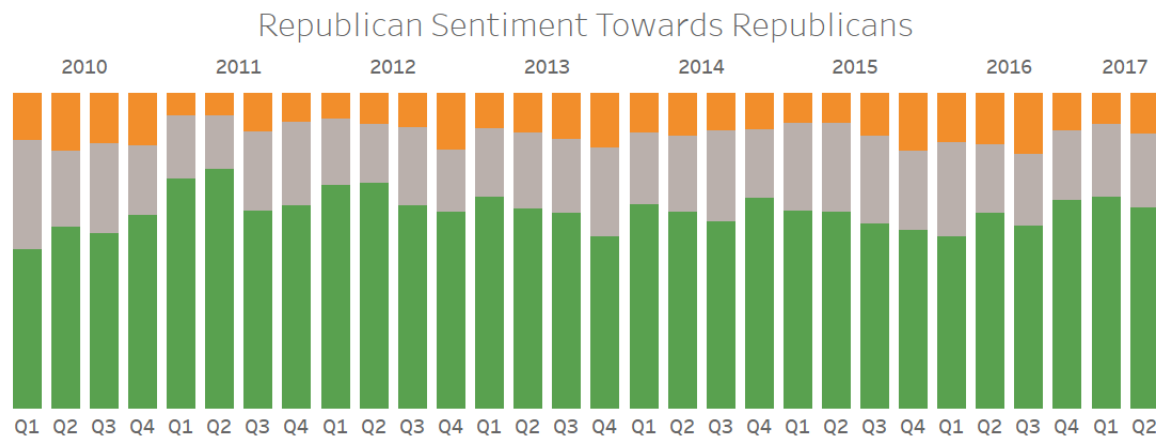
Mentions & Sentiment

Democrat Sentiment Towards Democrats
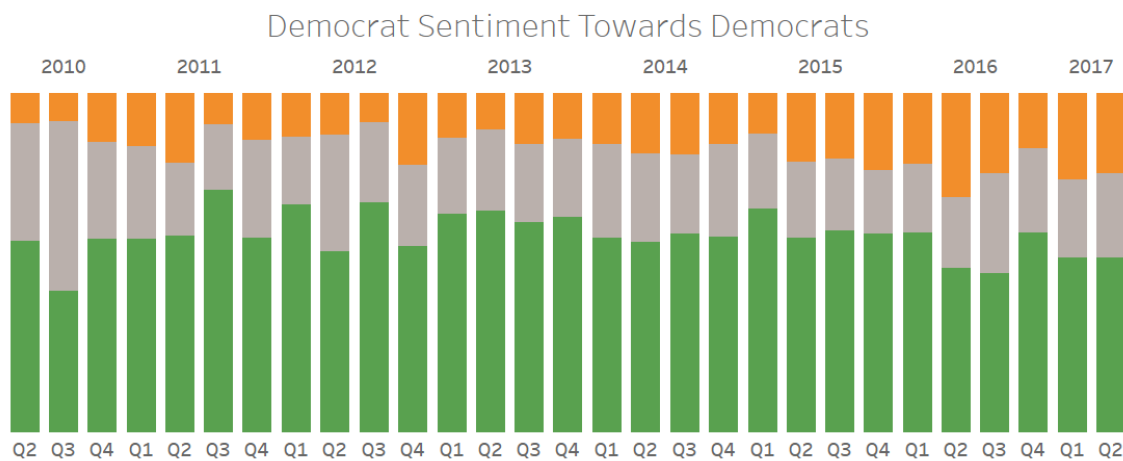
Democrats have slowly been mentioning more of their party members in negative tweets. This can be explained for a few reasons. First as they engage in more communication, they will mention their party members with negative sentiment aimed at another entity or subject. Secondly, 2016 was an election year and Democrats were having to compete against and be more critical of one another.

Democrat Sentiment Towards Republicans
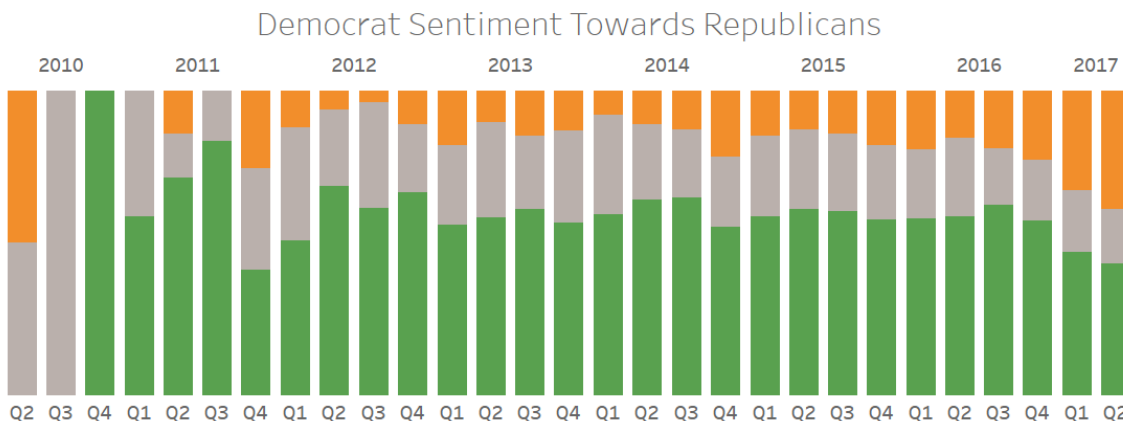
Democrats do not tend to mention Republicans with any huge difference in sentiment compared to Democrats until the 2016 election. Once Donald Trump becomes president Democrats become much more negative towards members of the Republican party.

Republican Sentiment Towards Republicans

Republican Sentiment Towards Democrats

Republicans tend to show consistent sentiments when communicating with Democrats and other Republicans. Republicans had a slight negative shift in their sentiments towards other Republicans during the 2016 election cycle. This could have been due Republicans being more critical of each other due to increased competition.

Word Frequency

To understand the conversations politicians are having on twitter we first looked at word frequency.

| Word	Count
great	70638
bill	47212
house	46785
new	45085
thanks	43808
thank	41136
day	37473
work	35888
act	35200
time	34627
support	34542
help	33189
congress	32548
health	31404
need	30645
watch	30200
must	29799
proud	29016
jobs	27824
people	27505

The most used word by politicians is great. This matches the general positive sentiment politicians have. Bill, house, and new are the next most used words. These words relate to basic job functions for politicians. Work could relate to the economy, employment, or welfare. Health could relate to healthcare, the economy, and more. Jobs also relate to the economy and employment.

Overall, the most used words tend to be positive and relate to political work or the economy.

Hashtags

The second part to understanding key topics is through hashtag analysis. We begin by looking at trending hashtags for republicans, then democrats, and then as a whole.


Top Democrat Hashtags
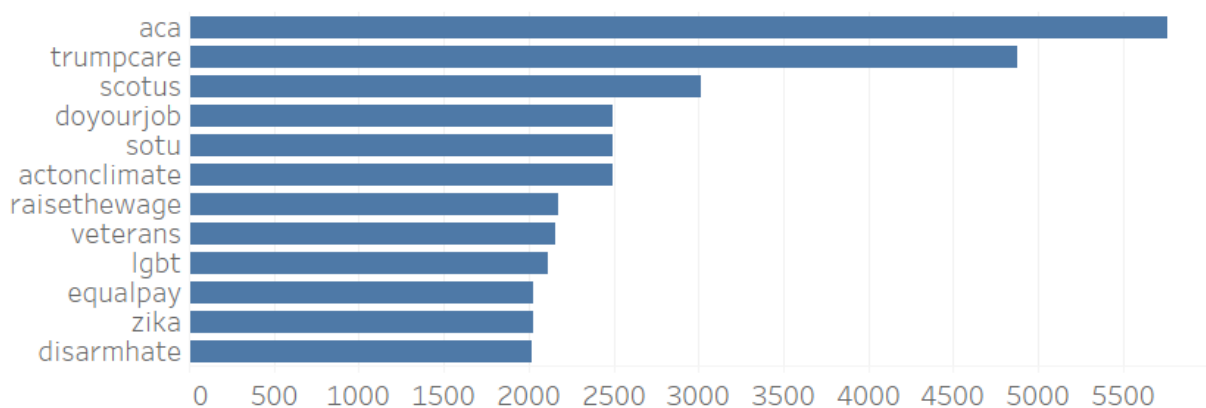
The most popular hashtags associated with Democratic accounts loosely follow party specific agendas. Looking at the hashtags used over 2000 times, only three are not associated with party agenda. First there is SCOTUS, Supreme Court of the United States. SOTU is the State of the Union Address. Lastly, Zika was a virus that was heavily active a few years ago.

ACA, the Affordable Care Act, was the healthcare plan proposed by former Democratic President Barack Obama. This was a controversial healthcare plan, and it is not surprising that it is referenced so often. TrumpCare, the healthcare plan proposed by Republican President Donald Trump, was another controversial

plan. The Doyourjob is a hashtag used by Democrats to highlight someone's or something's failure in doing all parts of their job description. The rest of the hashtags mean exactly what they say.

From this view the most important political topics for Democrats are healthcare, climate, minimum wage, veterans, gay rights, equal pay, and social justice.

Top Republican Hashtags
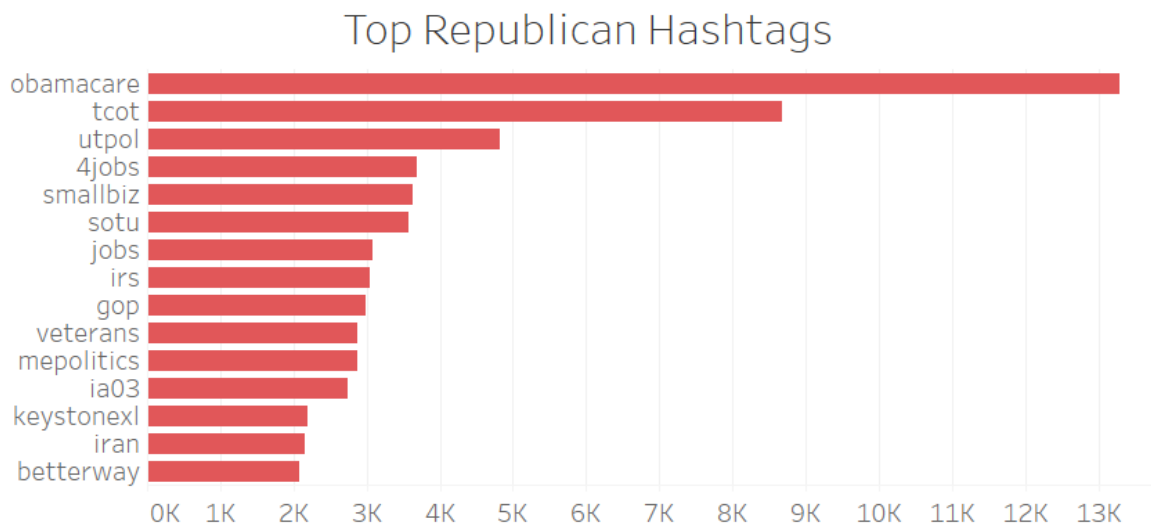
The most popular hashtags amongst Republicans loosely follow party agendas. Looking at hashtags with over 2000 observations, only one is not directly related to party agenda, SOTU.

Obamacare was a popular hashtag but should not persists into the future. TCOT is top conservative on Twitter and is a way for conservatives to support one another and created libraries of likeminded content. UTPOL and MEPOLITICS are both associated with Republican agendas for Utah and Maine respectively. The other hashtags are exactly what they sound like.

From this view the most important political topics for Republicans are healthcare, job creation, small business support, veteran, state support, and domestic violence.

Hashtags Related to Obamacare

Hashtags Related to TrumpCare
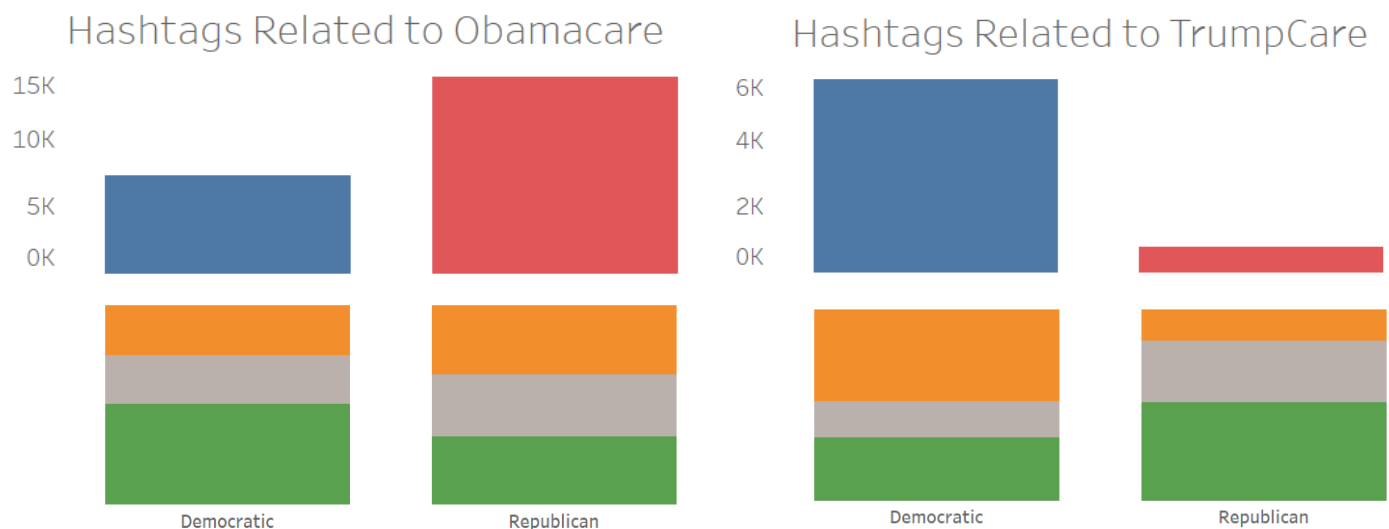
Taking a closer look at healthcare, we see that parties tend to speak about healthcare plans they oppose more than they do their own. They also tend to tweet with a negative sentiment when referencing the opposing healthcare plan.

Summary

In summary, tweeting behavior across political parties and political offices tend to be similar. The only political office with significantly different tweeting trends is the president. Donald Trump announced he was running for president in 2015 and since that time has redefined presidential communications. He has quickly become one of the most active political figures on twitter.

Excluding Donald Trump's presence, Senators are the most active relative to Representatives and Governors, and Democrats are more active than their Republican colleagues. The sentiments from all these groups are also similar and follow similar patterns.

Congress members mentioned the president and Donald Trump more than any other twitter account up to 2017. Virtually all these mentions were observed in the last three years. Aside from the former president, both parties communicate with members within their party more than those outside of their party. The party leaders are also the most mentioned.

Outside of congress, both parties mention party related accounts, governmental accounts, and media providers the most. In terms of media providers, MSNBC and NYTimes are the most mentioned news providers for Democrats. FoxNews and FoxBusiness are the most mentioned news providers for Republicans.

When Democrats mention other members of congress their sentiments are primarily positive. It is not until 2016 that Democrats become slightly more negative towards members of their own party. This is likely due to increased pressures from election campaigns. There is another noticeable change in 2017 once Donald Trump became president. Republicans became significantly more negative towards Republicans.

When Republicans mention other members of congress their sentiments are primarily positive. Like Democrats, Republicans had a slight negative spike around the 2016 election season. However, unlike Democrats, Republicans did not have a negative spike following Donald Trump's victory.

Congress members are primarily talking about healthcare on twitter during the observed time. This makes since as healthcare was one of the biggest points of discussion during the 2016 election season. Even before the election, Republicans also tweeted about Obamacare a significant amount of time.

Aside from healthcare, Democrats tend to talk about issues that align with general party agendas. Climate change, LGBT rights, social justice, and minimum wage are some of the most talk about topics. Outside of healthcare, Republicans tend to focus more on short term projects like the Keystone Pipeline, or with state politics. The major long-term topics include jobs and small business.

Lastly, my first hypothesis was not completely true. In the first eight years politicians did not have any significant differences in how they communicated with politicians in or out of their party. However, after the election campaign, Democrats became more negative in their communications with Republicans. Republicans did not follow this trend.

My second hypothesis was true regarding healthcare. Democrats were significantly more negative towards Trumpcare compared to their typical twitting trends. This is true for Republicans towards Obamacare. Regarding other political topics, there was not a clear shift in sentiment along party lines.