Sophie Thesis Update 06/03 Meeting

Whole Dataset

Number of Tweets in Dataset:
137,986 matches

Number of Unique Users in Dataset:
22,540

Number of Unique Stories They've Shared:
6,911

Users that follow at least one candidate:
14,382

Users that only follow one candidate:
7,003

Only Stories with 10+ Tweets

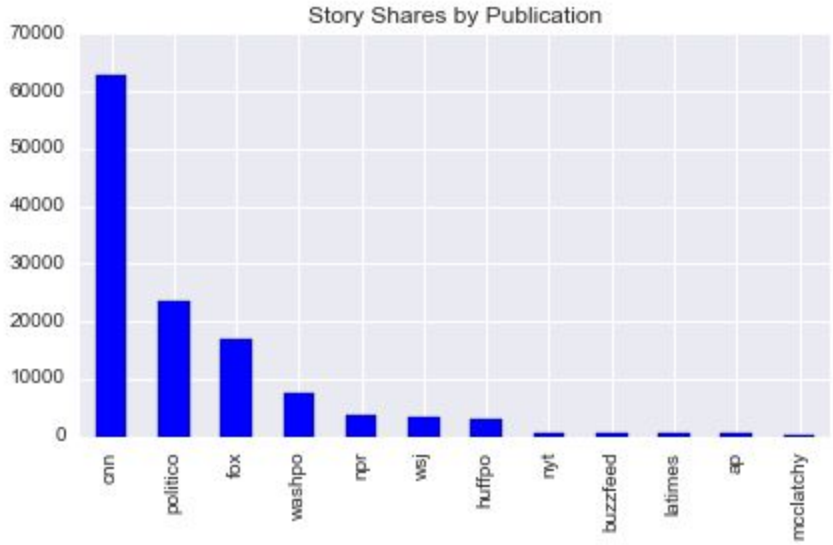
Number of Stories:
2,658 (38%)

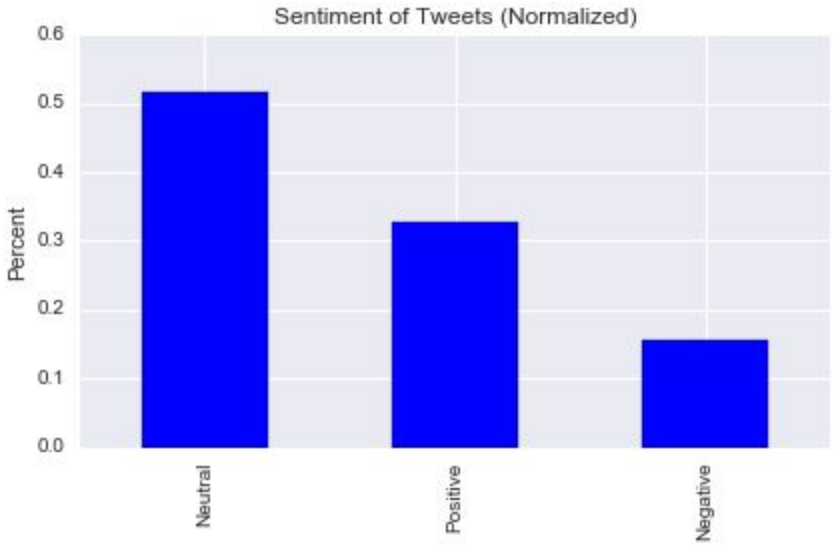
Number of Tweets:
123,113 (89%)

Number of Unique Users in Dataset:
20,956 (93%)

Users that follow at least one candidate:
13,472

Users that follow *only* one candidate:
6,567



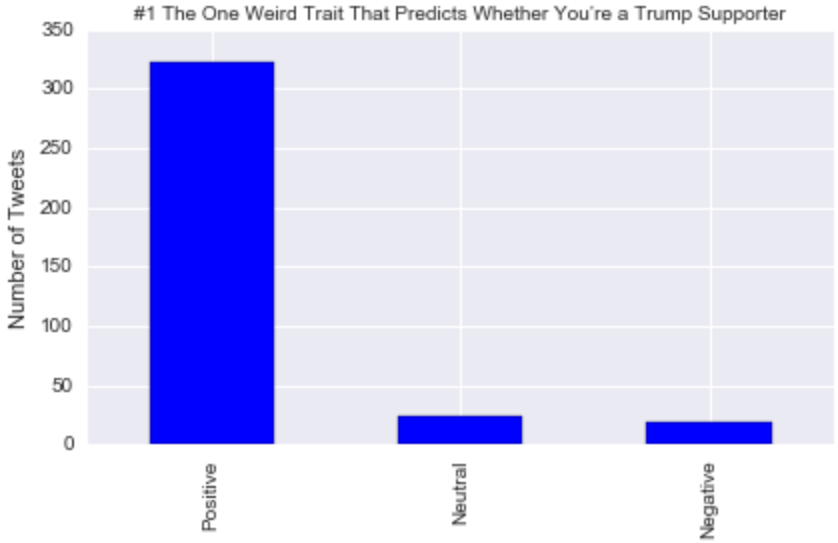


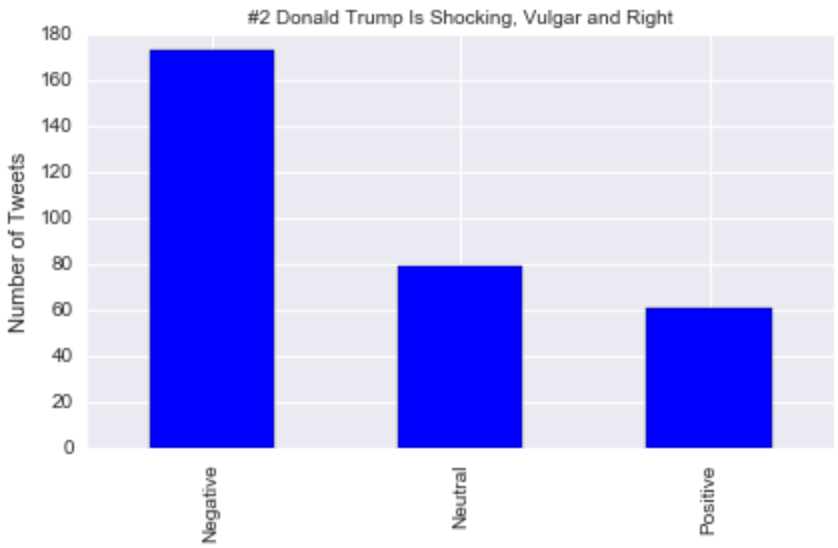
Tweets per Author

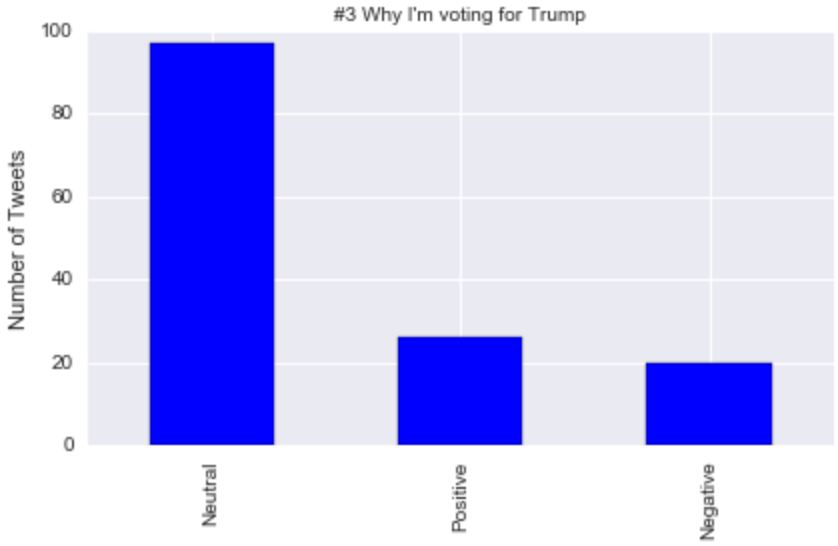
count	20956.000000
mean	5.874833
std	35.145778
min	1.000000
10%	1.000000
25%	1.000000
50%	1.000000
75%	2.000000
99%	124.000000
max	1246.000000

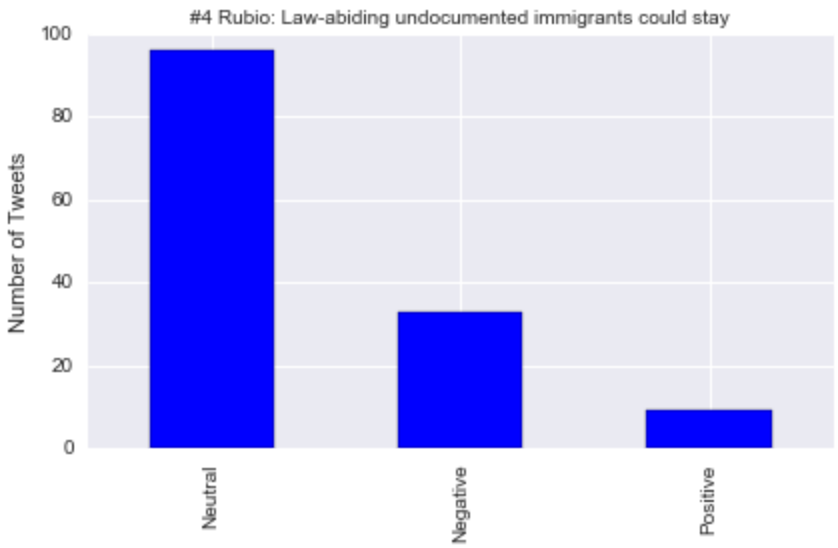
Top 10 Most Popular Stories

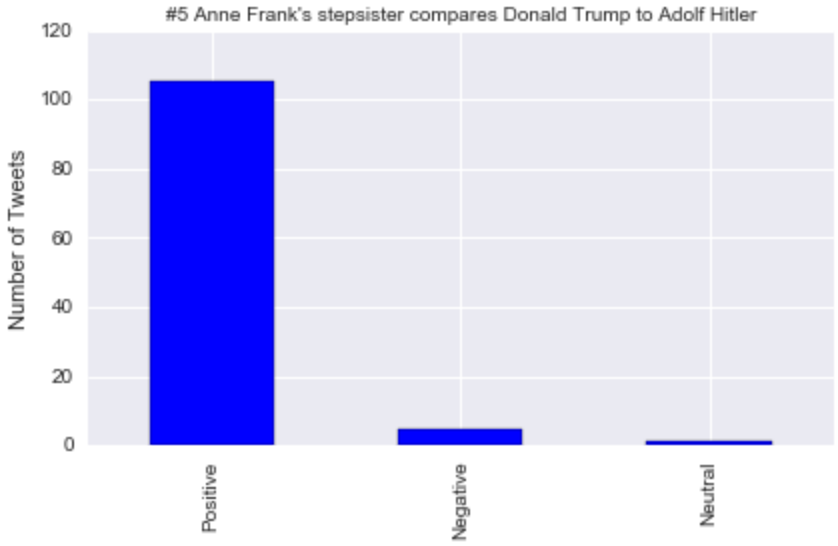
1. The One Weird Trait That Predicts Whether You're a Trump Supporter 364
2. Donald Trump Is Shocking, Vulgar and Right 313
3. Why I'm voting for Trump 143
4. Rubio: Law-abiding undocumented immigrants could stay 138
5. Anne Frank's stepsister compares Donald Trump to Adolf Hitler 111
6. Biden praises Sanders on income inequality 110
7. On January 20, 2017 Bernie Sanders Will Be Sworn In as America's 45th President 109
8. Poll: Trump dominates GOP field 108
9. Bernie Sanders Is a Once in a Lifetime Presidential Candidate. The Time Is Now. 105
10. TODD STARNES School caught trying to get students to work for Hillary 105

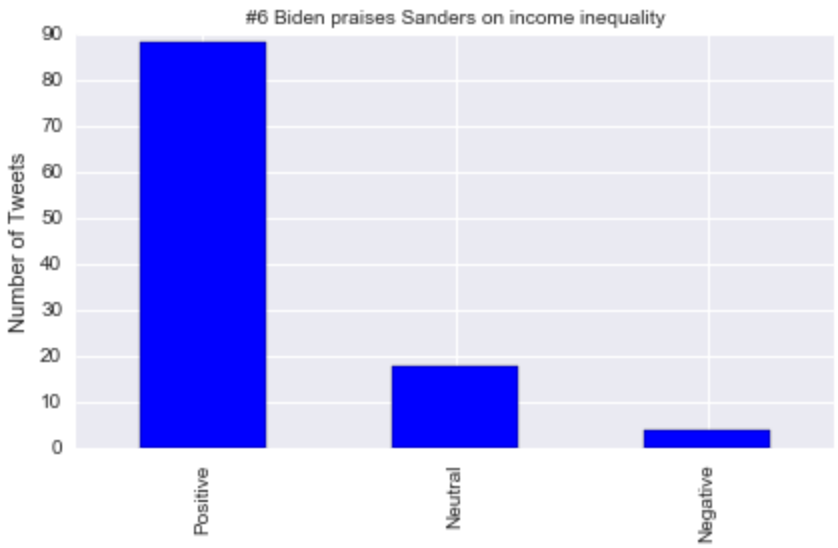


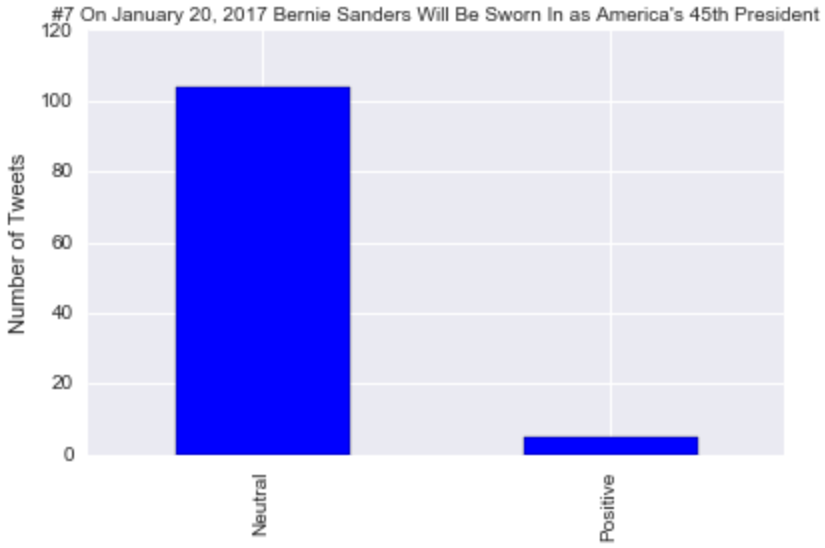

#2 Donald Trump Is Shocking, Vulgar and Right
Number of Tweets
180
160
140
120
100
80
60
40
20
0
Negative
Neutral
Positive

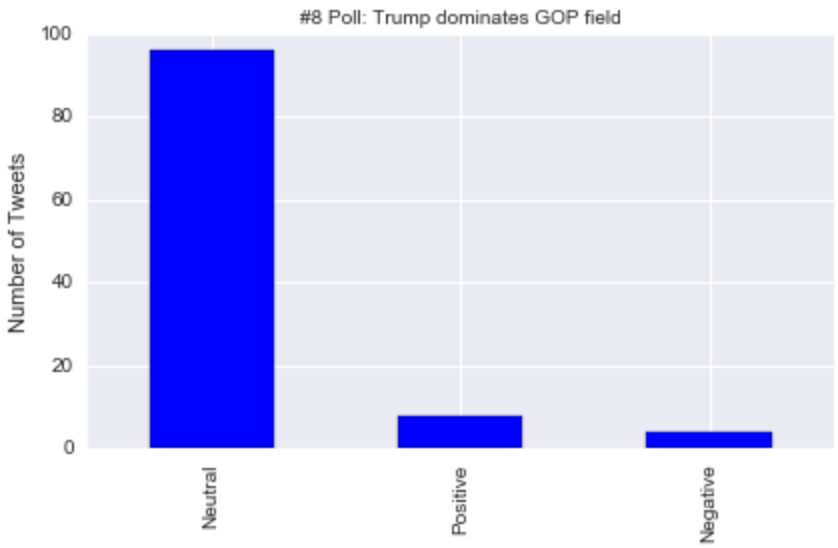

#3 Why I'm voting for Trump
Number of Tweets
100
80
60
40
20
0
Neutral
Positive
Negative

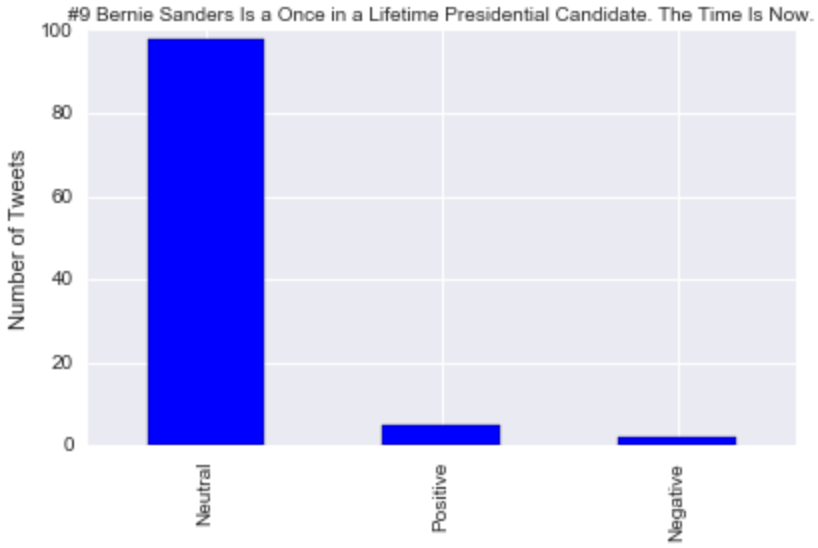

#4 Rubio: Law-abiding undocumented immigrants could stay
Number of Tweets
100
80
60
40
20
0
Neutral
Negative
Positive

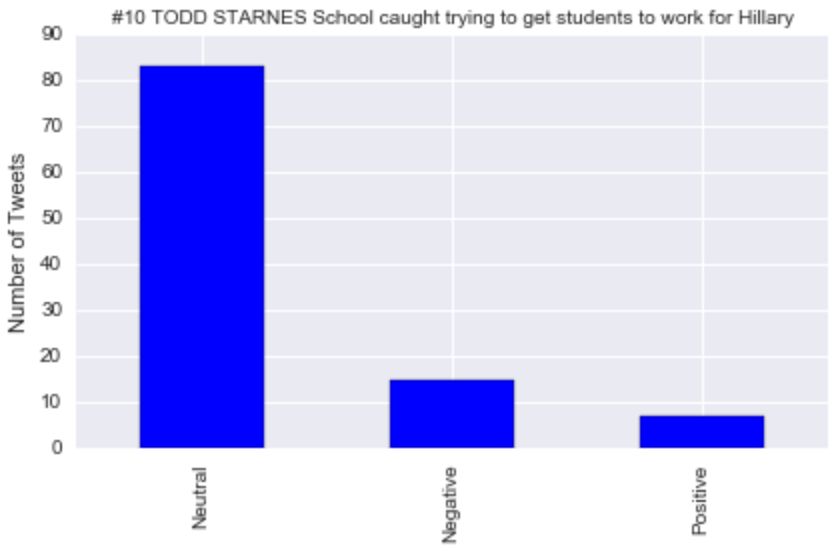

#5 Anne Frank's stepsister compares Donald Trump to Adolf Hitler
Number of Tweets
120
100
80
60
40
20
0
Positive
Negative
Neutral


#6 Biden praises Sanders on income inequality
Number of Tweets
90
80
70
60
50
40
30
20
10
0
Positive
Neutral
Negative


#7 On January 20, 2017 Bernie Sanders Will Be Sworn In as America's 45th President
Number of Tweets
120
100
80
60
40
20
0
Neutral
Positive


#8 Poll: Trump dominates GOP field
Number of Tweets
100
80
60
40
20
0
Neutral
Positive
Negative


#9 Bernie Sanders Is a Once in a Lifetime Presidential Candidate. The Time Is Now.
Number of Tweets
100
80
60
40
20
0
Neutral
Positive
Negative



Single-Candidate Followers (in This Subset)

Number of Clinton-only followers: 1858
Total tweets: 6432

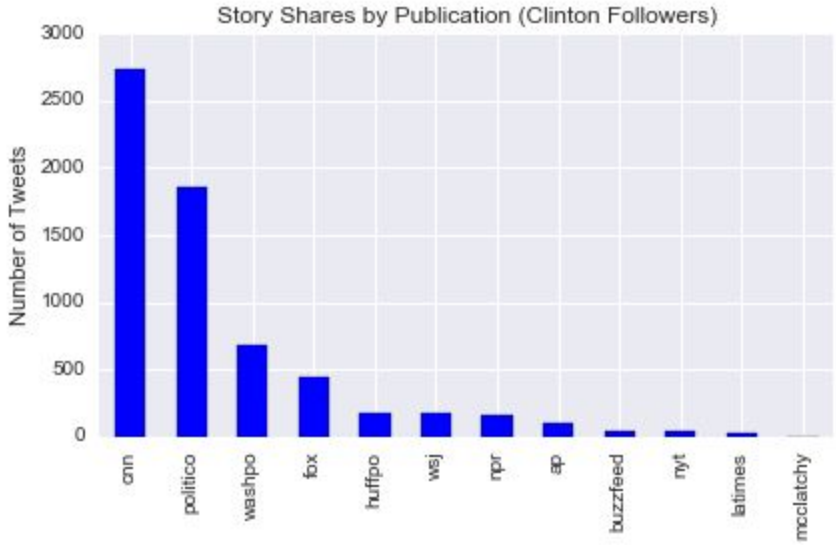
Number of Trump-only followers: 2018
Total tweets: 8815

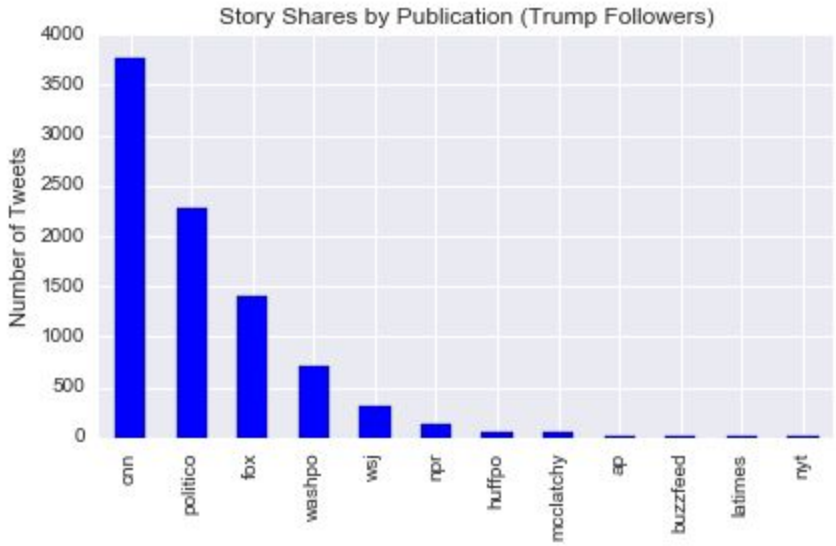
Number of Sanders-only followers: 1616
Total tweets: 4725

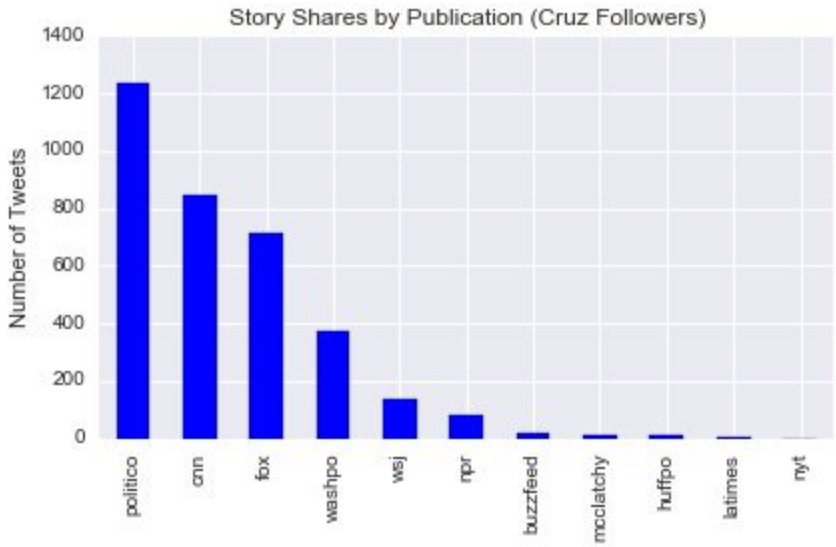
Number of Cruz-only followers: 916
Total tweets: 3441

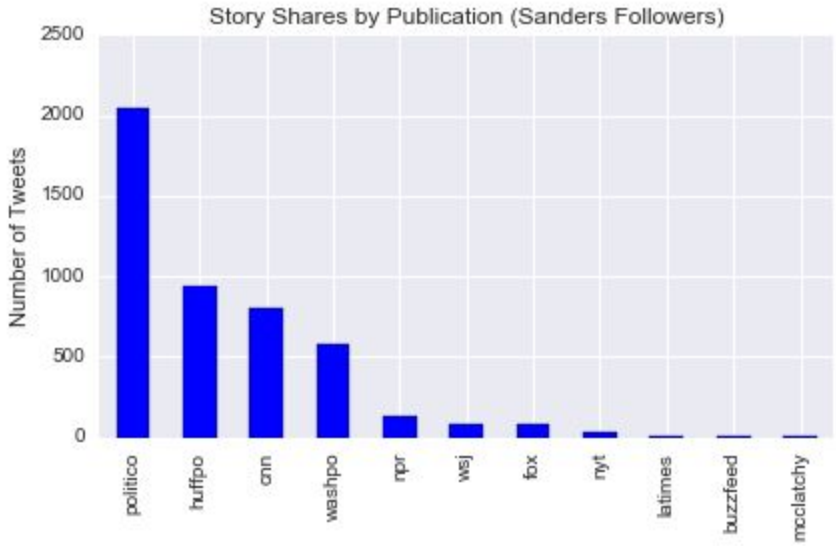
Number of Kasich-only followers: 159
Total tweets: 1048

Top Shared Stories and Publications for Each Camp


Story Shares by Publication (Clinton Followers)
Number of Tweets
3000
2500
2000
1500
1000
500
0
cnn
politico
washpo
fox
huffpo
wsj
npr
ap
buzzfeed
nyt
latimes
mcclatchy


Story Shares by Publication (Trump Followers)
Number of Tweets
4000
3500
3000
2500
2000
1500
1000
500
0
cnn
politico
fox
washpo
wsj
npr
huffpo
mcclatchy
ap
buzzfeed
latimes
nyt



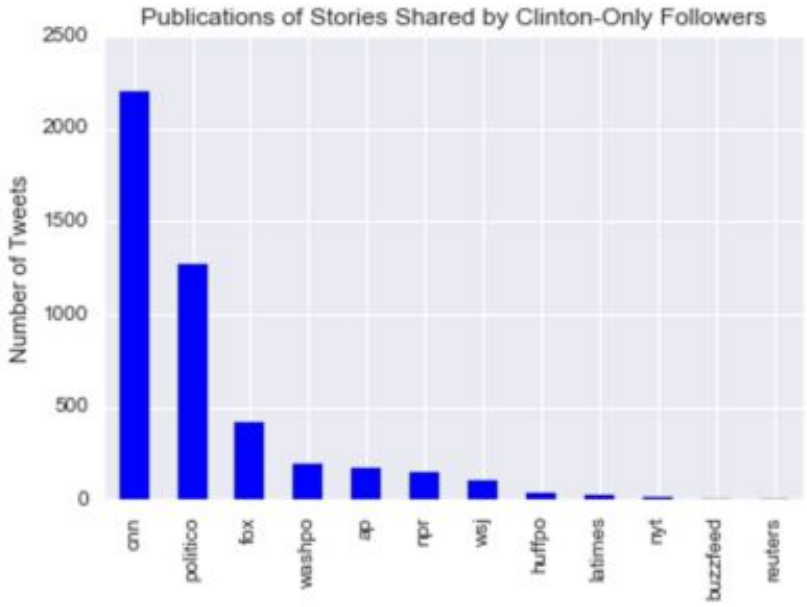


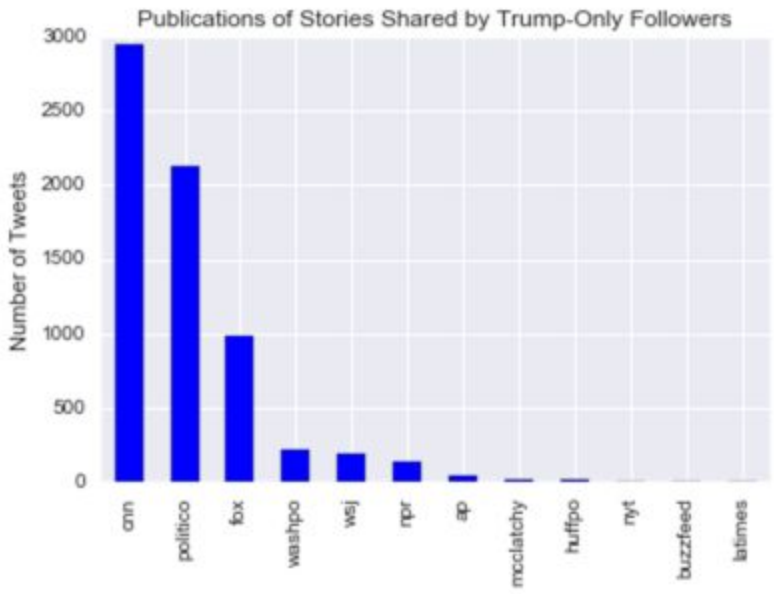
Go in tomorrow with the most <u>interesting</u> findings
Also make sure to do the top/10/bottom 10 emotional content sanity checks
Bench the twitter user attributes for now (save the data)
Correlations please
I think this finding X is the most interesting
BECAUSE it connects to existing theory T AND
Adds to it
Refutes it

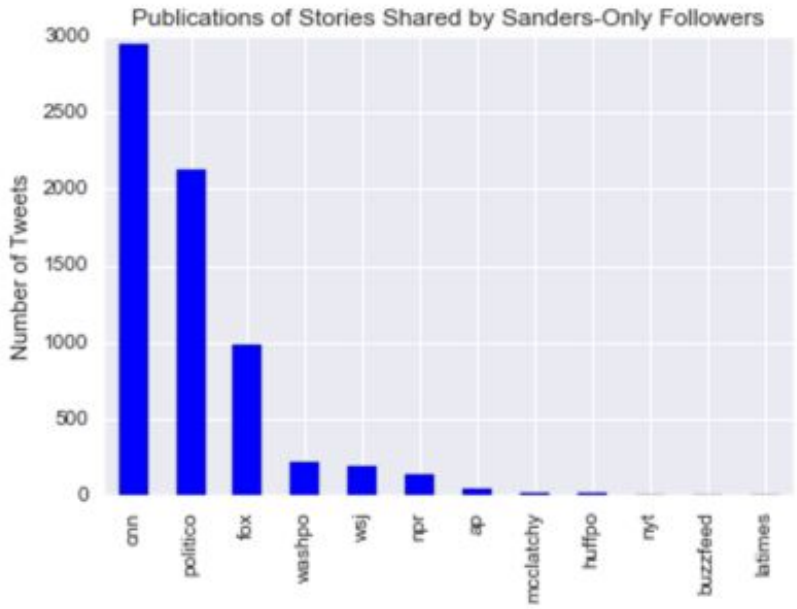
Section: Behavior of Single-Candid Tweeters

Do people who only follow hillary… share stories about only hillary…. From which org… neg or pos?

Frequency of Tweeting: Most Authors just Tweet Once







Top 10 Shared Stories by Clinton-Only Followers

Judge dismisses lawsuits over Clinton's emails	29
Ben Carson just called his new plan a 'genuine flat tax.' He's wrong.	16

GOP senators: If Cruz wins, we lose 13
White House won't commit to supporting Sanders over guns 13
Dont be fooled — Donald Trump's 'silent majority' doesn't exist 12
GOP's Trump revolt: Too little too late? 12
Republican elders ask 'who's worse' for the GOP brand: Trump or Cruz 12
Pope vs. Trump: 'Not Christian' to only build border walls 12
I Get Sanders' Appeal. But He's Not a Credible President. 11
Diplomats, national security officials blast Sanders on Iran normalization, ISIL 11

Top 10 Shared Stories by Trump-Only Followers

Trump basks in his spotlight 43
FOX NEWS POLL: Trump, Cruz lead Republican field nationally 32
MEDIA BUZZ: Conservative elite, sounding a bit panicked, says Trump will destroy GOP - Trump says Cruz should resolve citizenship issue - Carson prays with Clemson coach ahead of title game - CAMPAIGN 2016 LIVE BLOG 31
Terrorists use Trump's 'Muslim ban' speech in recruitment video 30
Why I'm voting for Trump 29
Cruz: Hillary Clinton's backers are 'very eager to support' Donald Trump 26
Donald Trump Is Shocking, Vulgar and Right 25
Campaigns lower expectations in Iowa, brace for Trump win 25
Ted Cruz's misleading attack that Donald Trump 'convinced' the government to 'bulldoze' a widow's home 24
Poll: Evangelicals flocking to Trump 23

Top 10 Shared Stories by Sanders-Only Followers

State Department asks for deadline extension on Hillary Clinton emails 70
Bernie Sanders Will Win Iowa, New Hampshire, Nevada and South Carolina. Here's Why. 50
Hillary's Corporate Democrats Taking Down Bernie Sanders 44
Bernie Sanders Is a Once in a Lifetime Presidential Candidate. The Time Is Now. 25
Bernie Sanders Is Now the 'Inevitable' Democratic Nominee and Presidential Winner 22
The Nation endorses Bernie Sanders 21
Sanders cracks Clinton's Nevada firewall 18
On January 20, 2017 Bernie Sanders Will Be Sworn In as America's 45th President 18
5 Reasons Bernie Sanders Will Win the Presidency in a Landslide 17
Donald Trump Is Shocking, Vulgar and Right 15

User Features

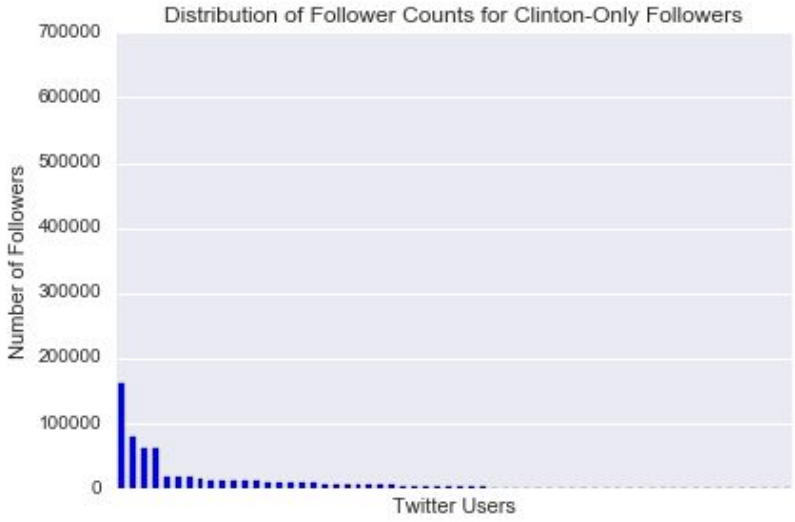
Number of users with deleted/suspended accounts:
59 out of 1,339 (1282 remaining --- off by 2?)

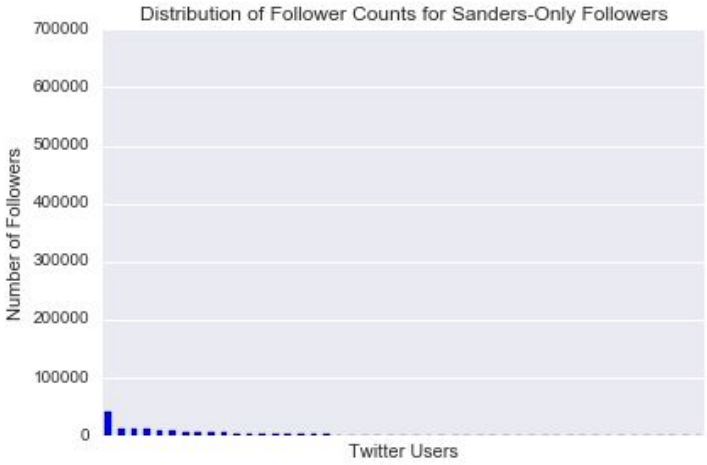
Features we have:

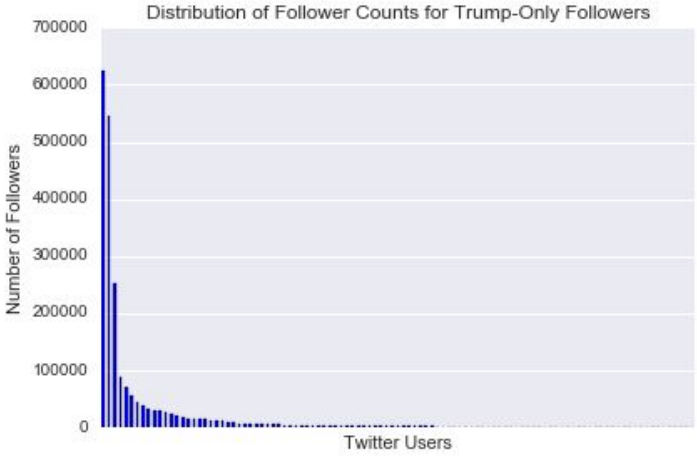
- Follower count
- Friend count
- Favorites count (participation)
- Verified
- Candidates
- Sentiment of Tweet

How loud is your megaphone (Follower Count)?

For each camp of single-candidate followers:







List of all Features so Far:

Story Context

- **Org**
- **Candidate**

Story Context

- **Emotional Content**
- **Reading Level**
- **Topic**

User

- **Credibility (verified?)**
- **Role (# followers : # followees -- Soroush)**
- **Camp (What candidate)**

Y axis:

- **Volume of tweets**
- **Controversiality (via Tweets / Soroush thesis) :**
    - **(p + n) ^ min(p,n)**

Features currently up in the air:

- Gender: Prashanth's model not yet ready
- Replies: talked to Martin + Soroush (impossible with API, potentially possible by crawling page -- talking to Prashanth)

NOTES

Comparison to Milkman's Corpus:
6,956 articles total, timespan Aug 30-Nov 30 (2008) [**Us:** 6,496]
Subset of 2,566 manually coded

Quality Checks:

- I sampled 100 random unmatched articles to look for any clear patterns-- do not see anything sticking out (I have attached those stories in a CSV).
- Will include caveats of this pipeline in methodology section.

Questions:

- Should we include Kasich followers?
- Should we look at those who follow more than 1 candidate?