

TWITTER SENTIMENT ANALYSIS

US Presidential Elections 2020



Project: Text Analytics

Members: Brett | Jocelyne | Mac | Priyanka | Shruti Kapur

8:30 am

Problem Statement



Opinions about the upcoming **US presidential election** are divided, both over what outcome is likely and over what outcome is desirable. One thing we can all agree on, however, is that there are profound differences in both style and substance between the contenders. Twitter has become the first place voters go to seek **information**, read **breaking news**, and **share their opinions** about political candidates during the presidential campaign.

If tweet sentiments are a proxy for the expected electoral turnout and result, what should our Presidential candidates be tweeting this election season?

Data Sources and Preprocessing

Data Sources: Twitter

Dates of Extraction: 10/08/2020-10/09/2020

Tweets Pulled: 35,290 total → 12,588 with states in profile location

Data extracted: Tweet text, date and time, user profile location

Search Keywords:

- **SCRAPER:** Debate, Biden, Trump, Pence, Harris, plus six key issues discussed in debate: COVID, economy, race, integrity, supreme court, Russia
- **ANALYSIS:** (1) Biden vs Trump (2) Battleground states (3) Issues of importance (derived by topic modeling)

Pre-Processing:

1. Extract states from location
2. No RTs pulled
3. Remove irrelevant frequent words (for eg, fly during VP debate)
4. Remove websites
5. Remove emoticons and unrelated/ambiguous symbols
6. Group words into a bucket: For eg realdonaldtrump, trump, pence, trumpence are all replaced by Trump
7. Removed stop words

Frequency of Important words:

Trump	13290
Biden	10892
China	1615
COVID	1539
vote	1163
election	1142
healthcare	1013
Russia	946
climate	834
BLM	745

Analysis

Biden vs Trump

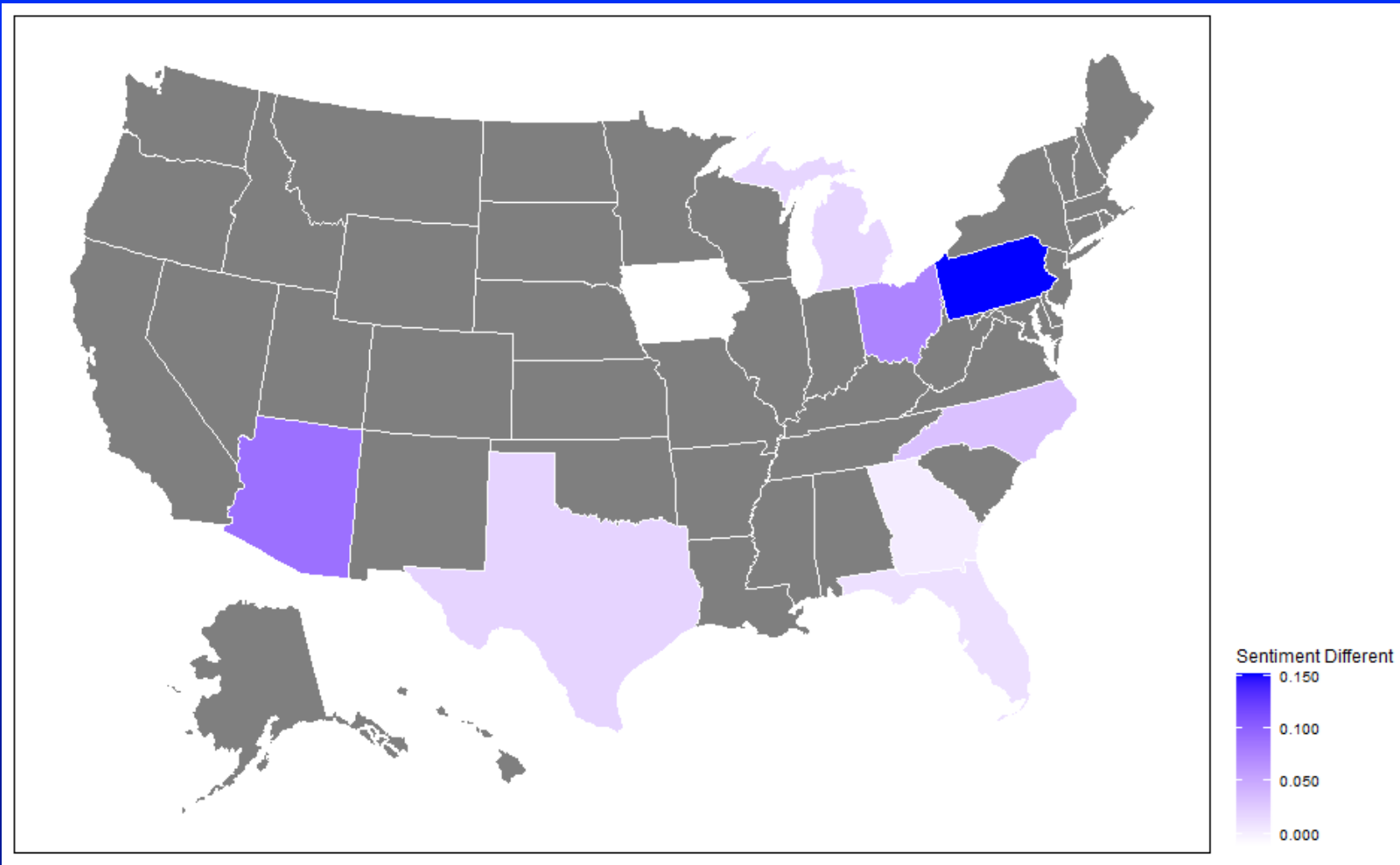
SENTIMENT ANALYSIS

- **Vader:** +/-5 words and sent_tokenize used around our keyword (Biden/Trump) and group by the states of importance
- We did not use the tweets which used both the words in a single sentence as it gave dissimilar results versus the overall sentiment for the candidate in the state
- We did an overall sentiment analysis and went further into state specific sentiments, focusing on the battleground states

Location	Biden	Trump	Difference
Overall	0.019	-0.017	0.036
Nevada	0.146	-0.084	0.230
Connecticut	0.124	-0.091	0.215
Pennsylvania	0.074	-0.078	0.152
Florida	0.017	0.006	0.011
Texas	-0.001	-0.021	0.019
Kansas	-0.056	0.100	-0.156
Missouri	0.000	0.173	-0.173

Results

Biden vs Trump in Battleground States



- **Pennsylvania, Arizona, and Ohio** tend to have more positive sentiment toward **Biden**
- **Iowa, Georgia, and Florida** have more positive sentiment toward **Trump**
- **Texas, Michigan, and North Carolina** average to have near 0 sentiment, indicating a toss-up

Analysis

Battleground States

TOPIC MODELING

- Round 1: Use LDA model to read each tweet and found the most relevant topic
- Round 2: Ensure that the most frequently occurring topics were relevant and actionable, keeping the model integrity intact (*coherence score*)>0.5
- Iterate multiple times to find the model that gave the most robust coherence score
- The exercise gave us a set of keywords that were later used for sentiment analysis within each battleground state

Battleground State (electoral votes)	Topics
Texas (38)	Health, Race, COVID, jobs, economy, pandemic, women, supreme court, tax, science, climate, integrity
Florida (29)	Integrity, pandemic, tax, economy, race, healthcare, business, supreme court, leadership, women, gun, order
Pennsylvania (20)	Pandemic, race, economy, tax, court, women, integrity, climate change, corrupt, COVID, healthcare, supreme court
Ohio (18)	Women, climate change, job, economy, integrity, fracking, supreme court, COVID, scandal, healthcare, racism, tax, terrorist
Michigan (16)	Supreme court, healthcare, abortion, healthcare, job, COVID, climate change, market, racism, economy, lie
Georgia (16)	Race, healthcare, COVID, women, record, business, tax, climate change, convincing, presidential, lie, criminal
North Carolina (15)	Healthcare, COVID, race, riot, women, economy, lie, court, tax, science, war, job
Iowa (6)	Science, covid, women, foreign policy, race, leadership, border, climate change, trust

Results

Issues of Importance

SENTIMENT ANALYSIS

- **Vader:** +/-5 words used around our keyword (issue of importance) and group by the states
- Analyze topic sentiment in state and between states (common topics)
- Analyze the overall sentiment of the Twitter users within battleground states

Battleground State (electoral votes)	Sentiment Analysis
Texas (38)	Court: 0.1666 Race: -0.2743
Florida (29)	Healthcare: 0.1527 Gun: -0.34
Pennsylvania (20)	COVID: 0.0471 Economy: -0.3950
Ohio (18)	Economy: 0.3767 COVID: -0.1571
Michigan (16)	Job: 0.6249 Race: -0.3675
Georgia (16)	Race: 0.0877 COVID: -0.2454
North Carolina (15)	Economy: 0.2817 Riot: -0.7964
Iowa (6)	Women: 0.4019 COVID: 0.1591

Overall Results

CANDIDATE SENTIMENT

TX	
FL	
PA	
OH	
MI	
GA	
NC	
IA	

Biden	Trump
-0.001	-0.021
0.017	0.006
0.074	-0.078
0.067	-0.008
0.010	-0.008
-0.006	-0.008
0.022	-0.010
0.073	0.084

ISSUES OF IMPORTANCE SENTIMENT

Battleground State	Sentiment Analysis
Texas (38)	Court: 0.1666 Race: -0.2743
Florida (29)	Healthcare: 0.1527 Gun: -0.34
Pennsylvania (20)	COVID: 0.0471 Economy: -0.3950
Ohio (18)	Economy: 0.3767 COVID: -0.1571
Michigan (16)	Job: 0.6249 Race: -0.3675
Georgia (16)	Race: 0.0877 COVID: -0.2454
North Carolina (15)	Economy: 0.2817 Riot: -0.7964
Iowa (6)	Women: 0.4019 COVID: 0.1591

Insights and Recommendations

INSIGHTS

- States characterized as **battleground** also tend to have **mixed** twitter sentiment
- **Biden's** states with the lowest sentiment relative to Trump are **IA, GA, FL, and MI**, while **Trump's** are **PA, AZ, OH, and NC**
- There is no single important issue but a multitude of battleground issues like **COVID, economy, race, and healthcare** which indicate that candidates have to walk a tight line in campaigns

RECOMMENDATIONS

- Candidates should try to sway votes based on the issues of importance in the final leg of campaigns for battleground states
- Candidates should have more focus on states where their sentiments are slightly positive v/s the opponent or the race is a toss-up to get the decider votes
- The leading candidate should identify what states his opponent might pour money into and try to washout his attempt to maintain status quo

CAVEATS

- User profile location may not be the user's voting location
- Is the Twitter crowd representative of the voter base?
- Twitter attracts more criticism/trolls than praises, bound to show more negative sentiments
- The Silent Majority might not be on Twitter
- Creates ambiguity when tweets refer to multiple debates and news events