

Pre- and Post-Debate Democratic Primary Data: Twitter, Google Trends, and Polls

Fundamentals of Computing and Data
Display, Fall 2019

Zoe Padgett and Fatou Thiam

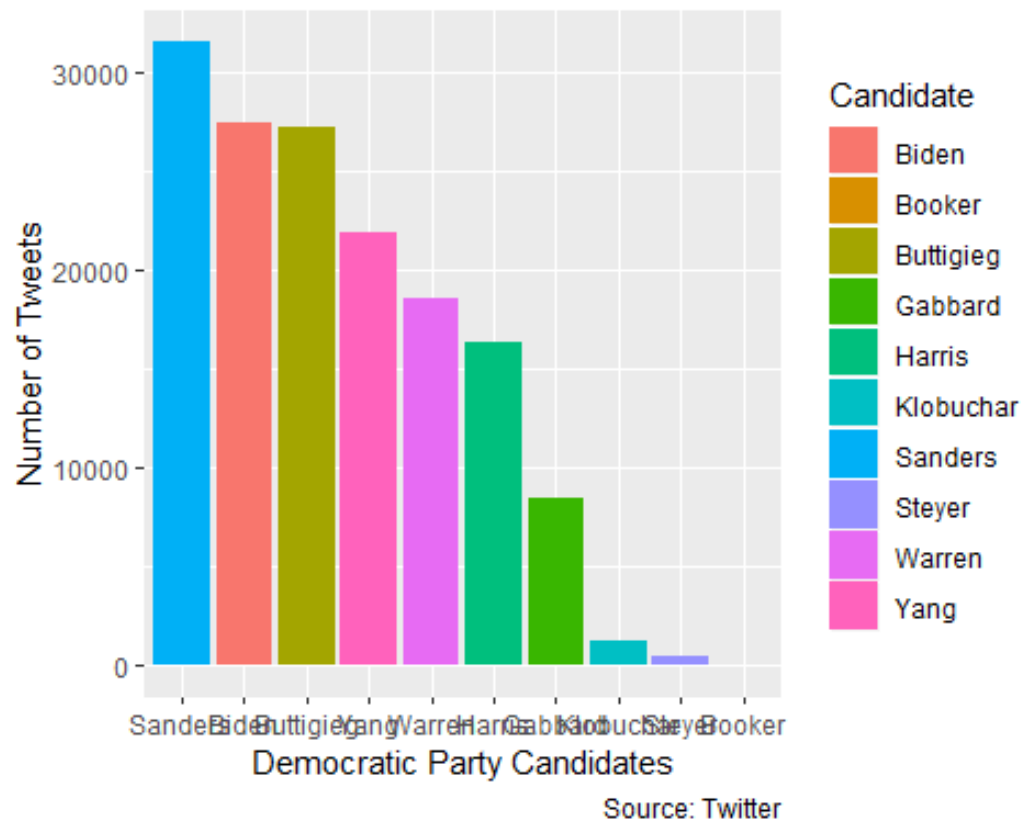
Overview

- Can Twitter data and Google trends data be used to supplement polling results, by providing real-time information to candidates while they wait for polling data to come in?
- Exploratory analysis to see if pre- and post-debate polling data during the 2020 U.S. Democratic primary election aligns with these real-time sources

Twitter Data

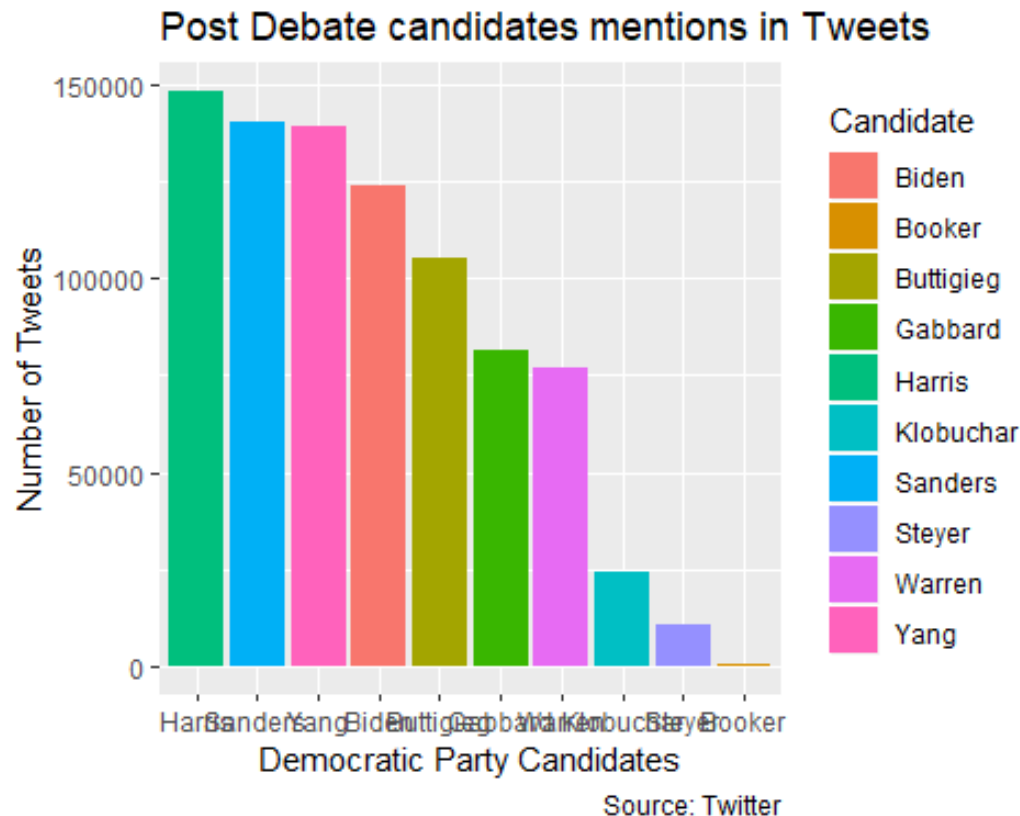
- Used Rtweet package and search_tweet function with selected keywords to gather tweets between November 20th and November 21st
- Separated the 1.3 million tweets after removing irrelevant tweets in 2 groups: pre and post debate
- Generated maps by number of tweets per candidate and ran sentiment analysis for 5 candidates

Pre-debate number of tweets



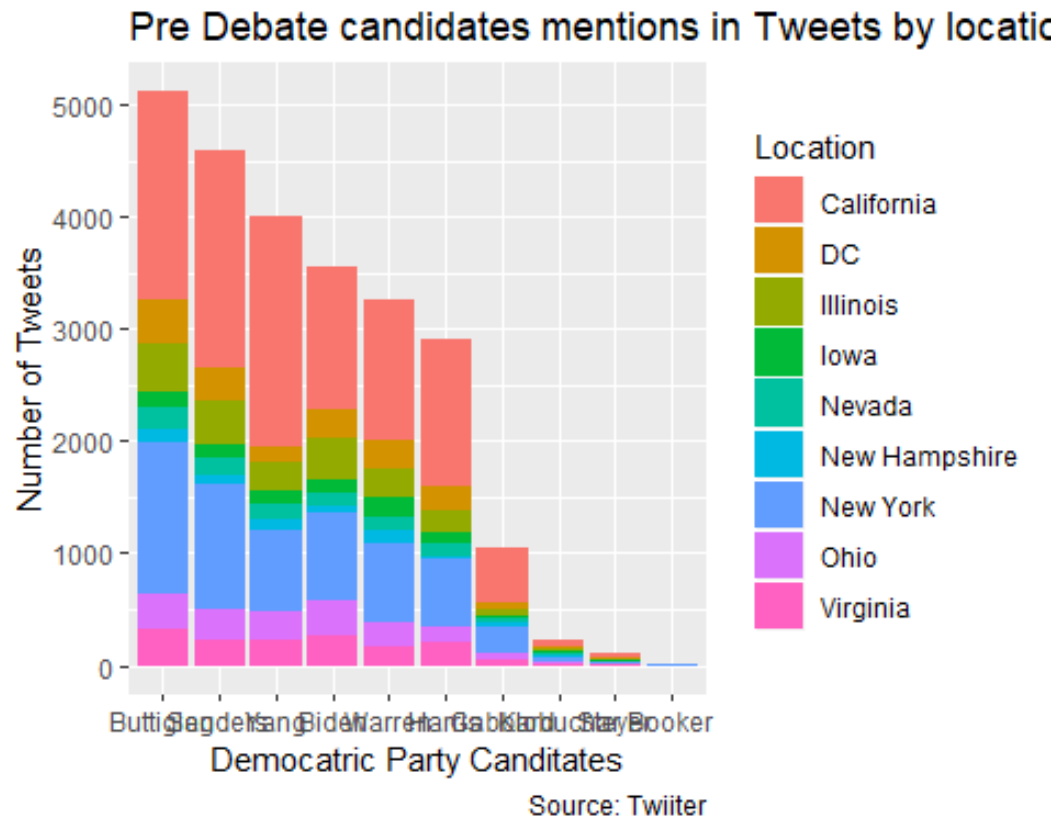
Pre-debate Mentions in Tweets, by Candidate

Post-debate number of tweets



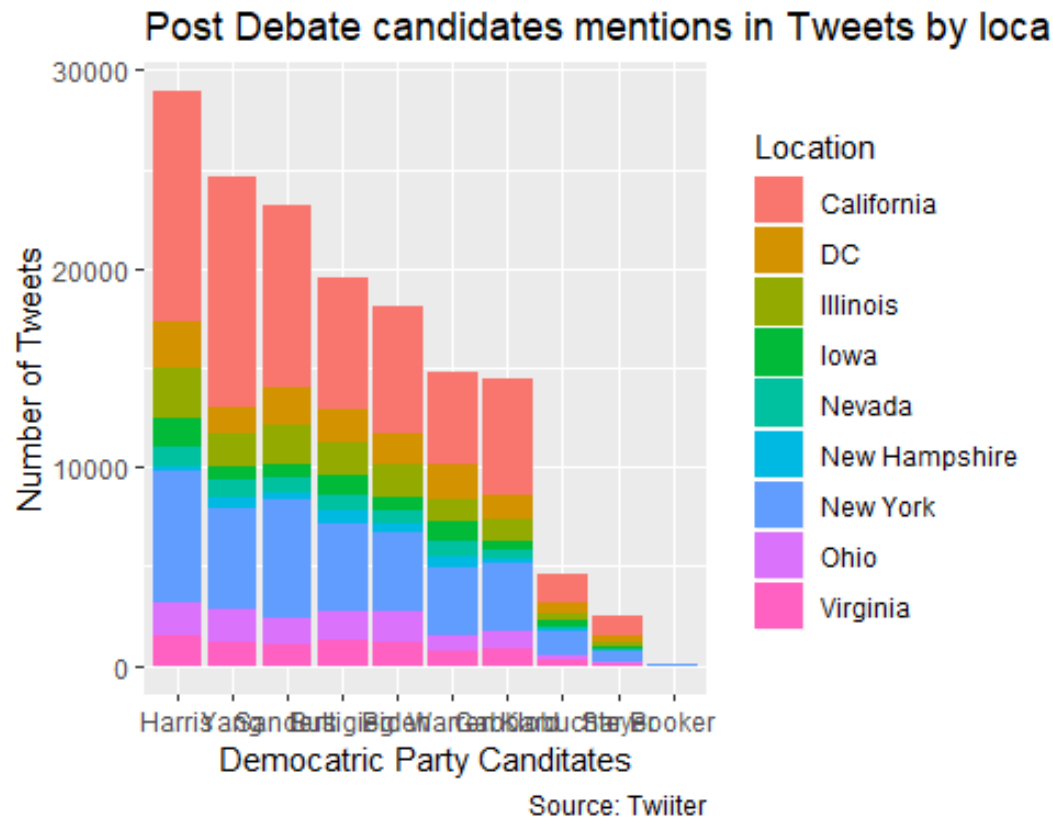
Post-debate Mentions in Tweets, by Candidate

Tweets location pre-debate



Pre-debate Mentions in Tweets, by Candidate & location

Tweets location post-debate

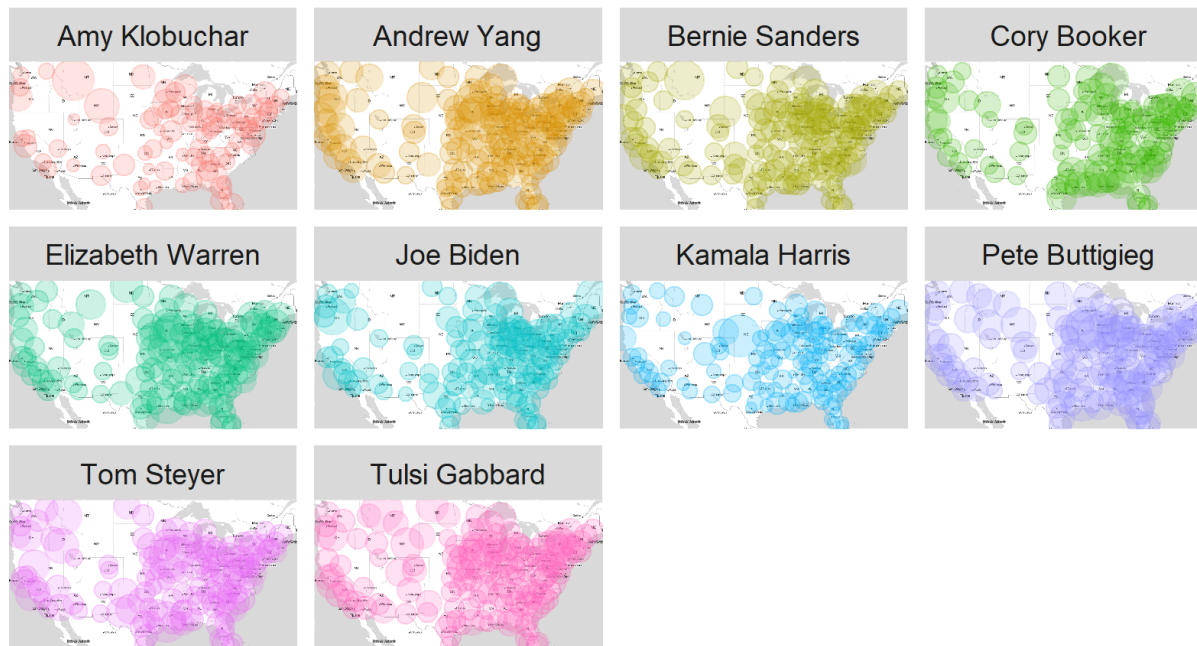


Post-debate Mentions in Tweets, by Candidate & location

Google Trends Data

- Used gtrends package to pull Google Trends data for each candidate before and after the debate
- Generated maps for each candidate to see the geographic distribution of searches
- Graphed the search popularity of each candidate pre- and post-debate

Google Trends Map



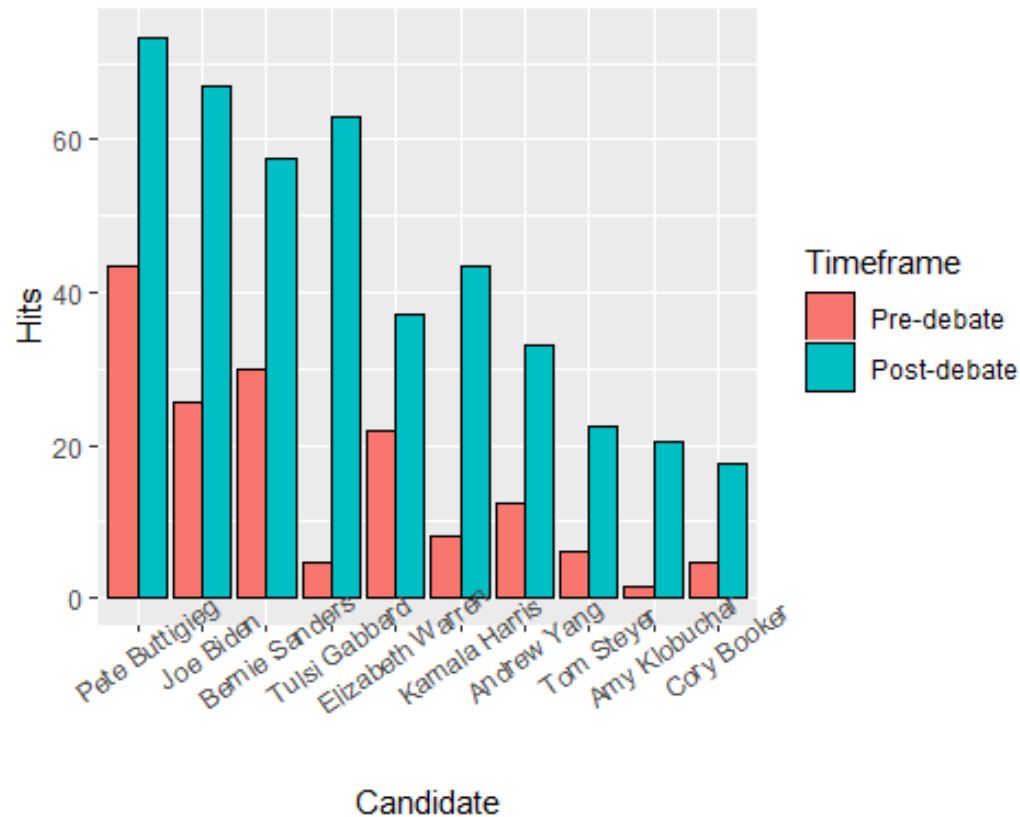
Candidate

●	Amy Klobuchar
●	Andrew Yang
●	Bernie Sanders
●	Cory Booker
●	Elizabeth Warren
●	Joe Biden
●	Kamala Harris
●	Pete Buttigieg
●	Tom Steyer
●	Tulsi Gabbard

Hits

●	25
●	50
●	75
●	100

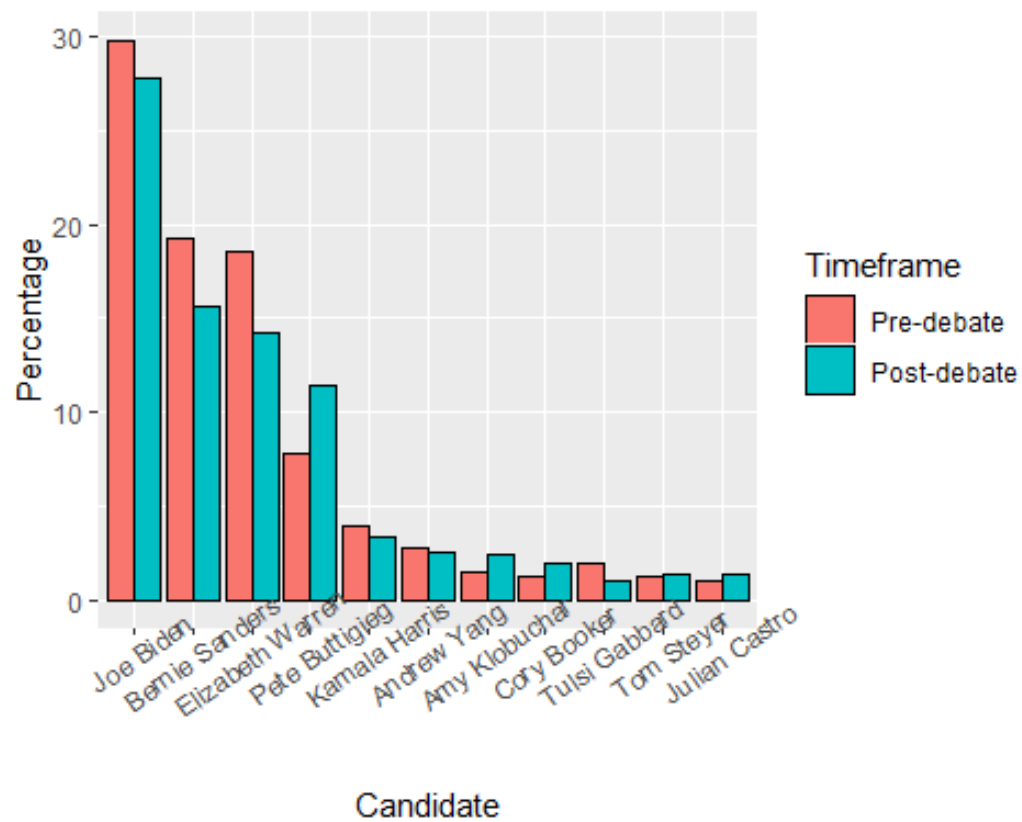
Candidate Search Popularity



Pre- and Post-debate Google Trends Data, by Candidate

Polling Data

- Data from RealClearPolitics, average polling results pre- and post-debate



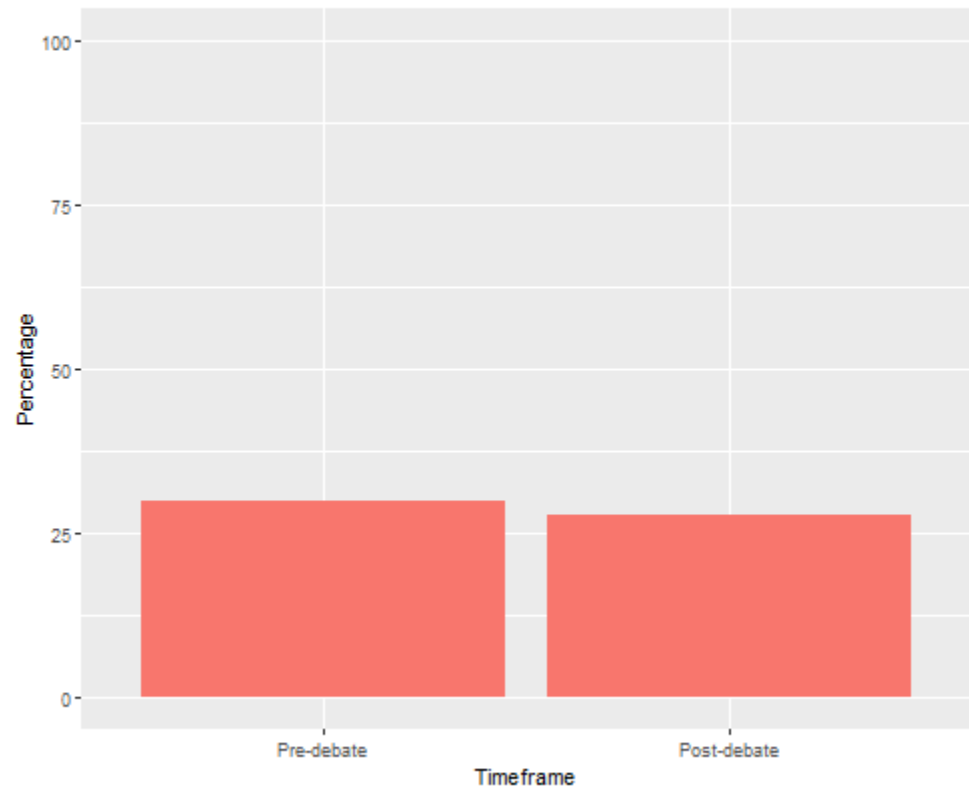
Pre- and Post-debate Polling Data, by Candidate

Results

2020 Democratic Primary Candidate Data

Candidate
Joe Biden ▼

Data Type
Polling ▼



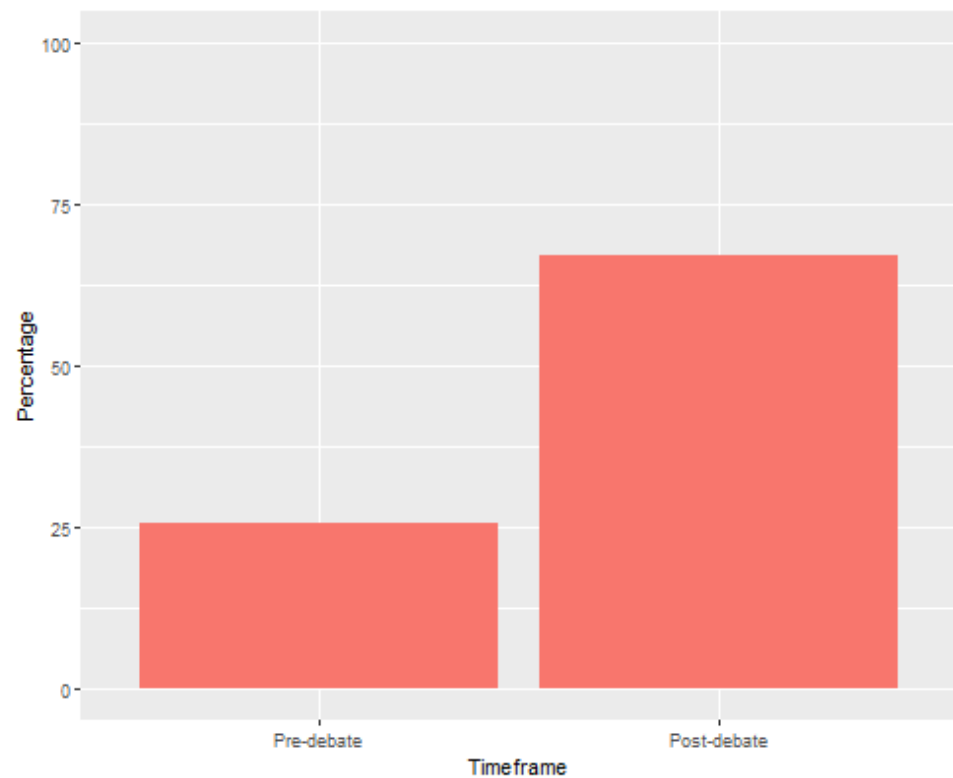
2020 Democratic Primary Candidate Data

Candidate

Joe Biden ▼

Data Type

GTrends ▼



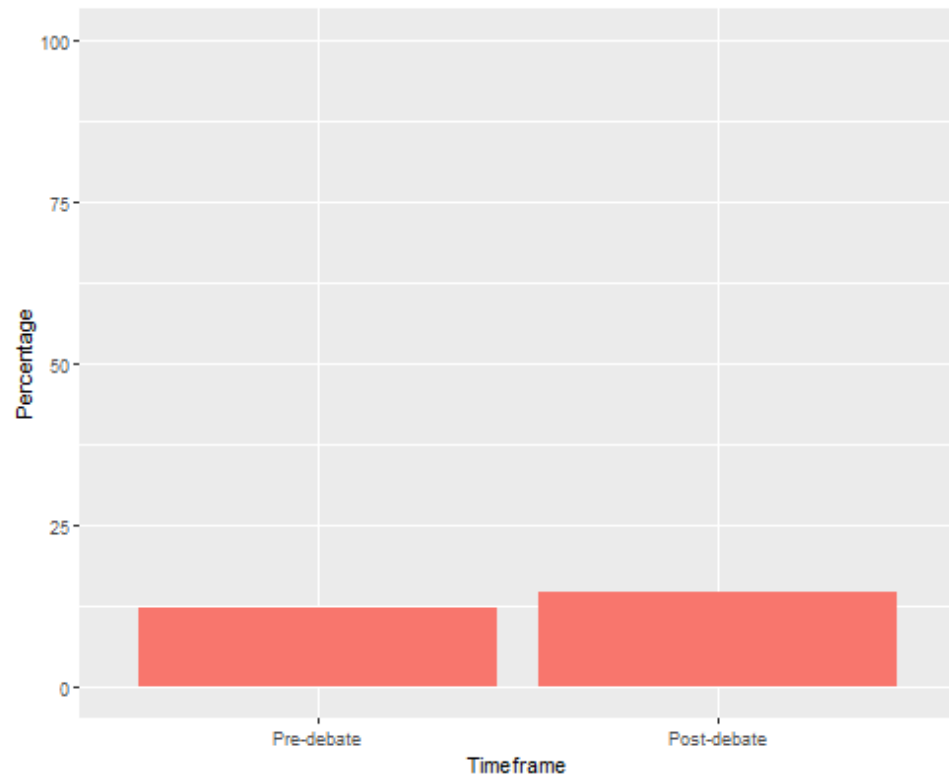
2020 Democratic Primary Candidate Data

Candidate

Joe Biden ▼

Data Type

Twitter ▼



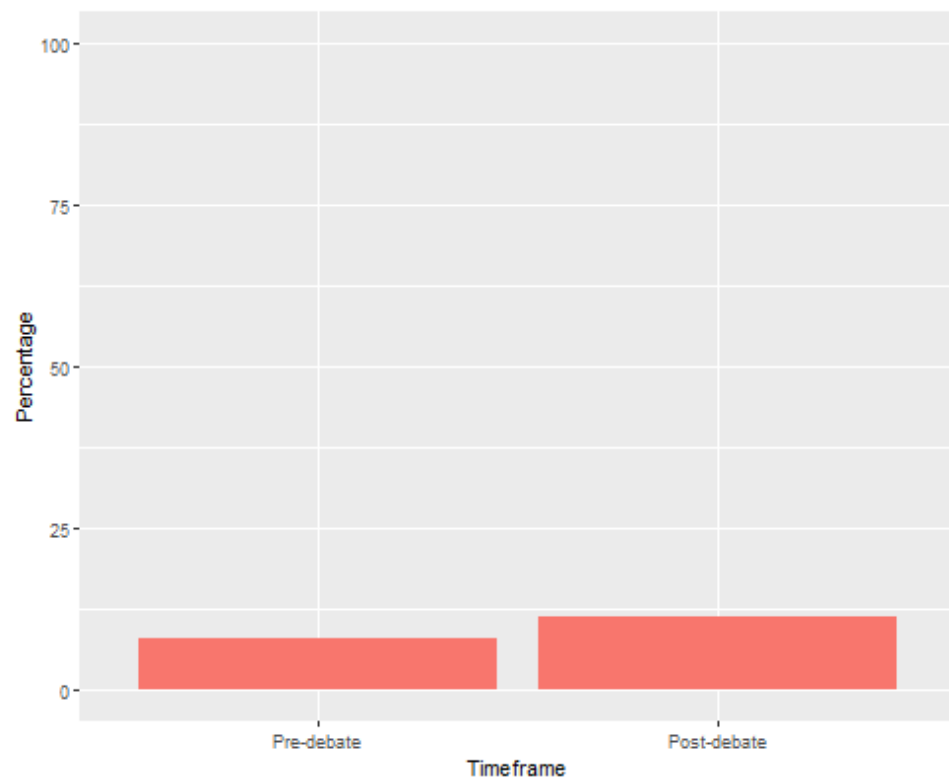
2020 Democratic Primary Candidate Data

Candidate

Pete Buttigieg ▼

Data Type

Polling ▼



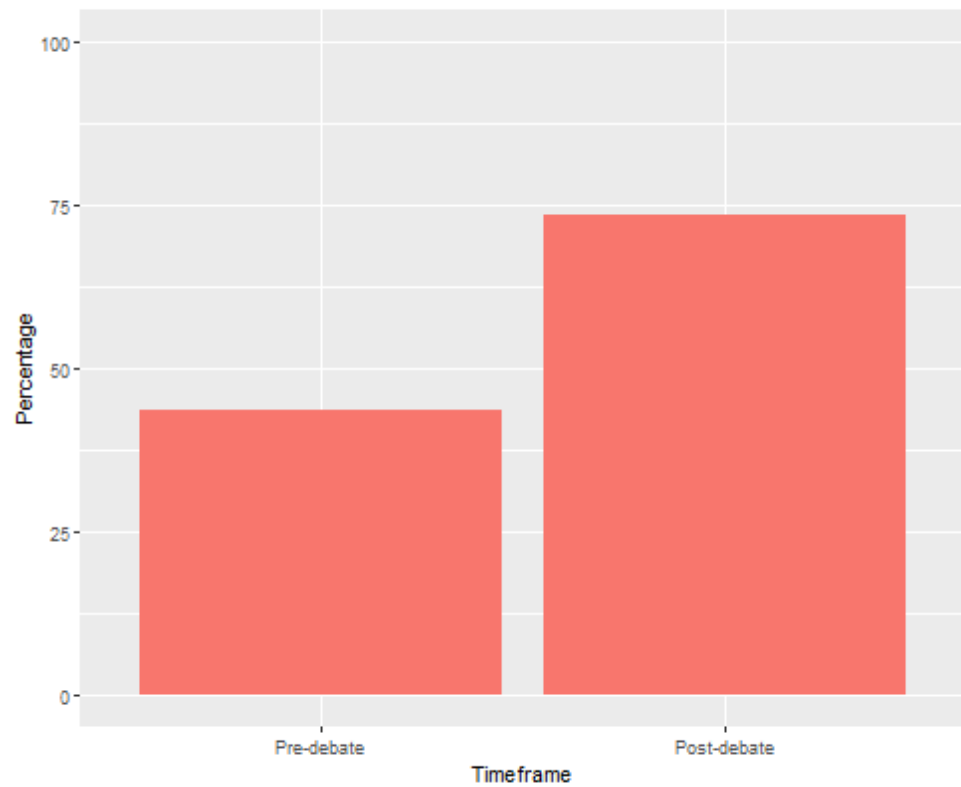
2020 Democratic Primary Candidate Data

Candidate

Pete Buttigieg ▼

Data Type

GTrends ▼



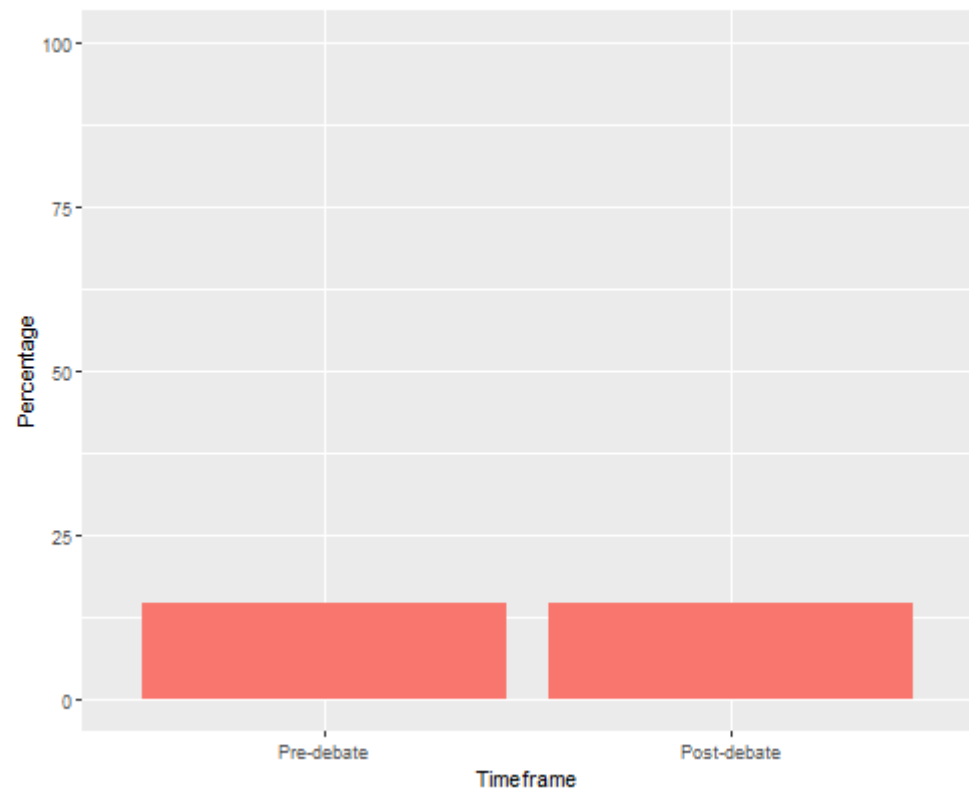
2020 Democratic Primary Candidate Data

Candidate

Pete Buttigieg ▼

Data Type

Twitter ▼



Conclusion

- Neither Twitter nor Google Trends data align with polling data
- These data are not representative of US voters, lack of demographic data for adjustment
- Variable-to-variable comparison between two data sources may not be realistic (e.g., do number of searches and positivity of tweets actually mean more people will vote for a candidate?)
- Many difficulties in practical usage of this data (e.g., location fields)
- Overall, we don't think this data should be used in place of survey data
- However, this data could still be useful for candidates in understanding how social media users perceive them before and after a debate, or in understanding their search popularity