Predicting Election Trends By Web Data

BigSurv 2020
Utrecht, Netherlands

Deji Suolang (presenter)
Kaan Cem Ketenci

Three primary data sources:

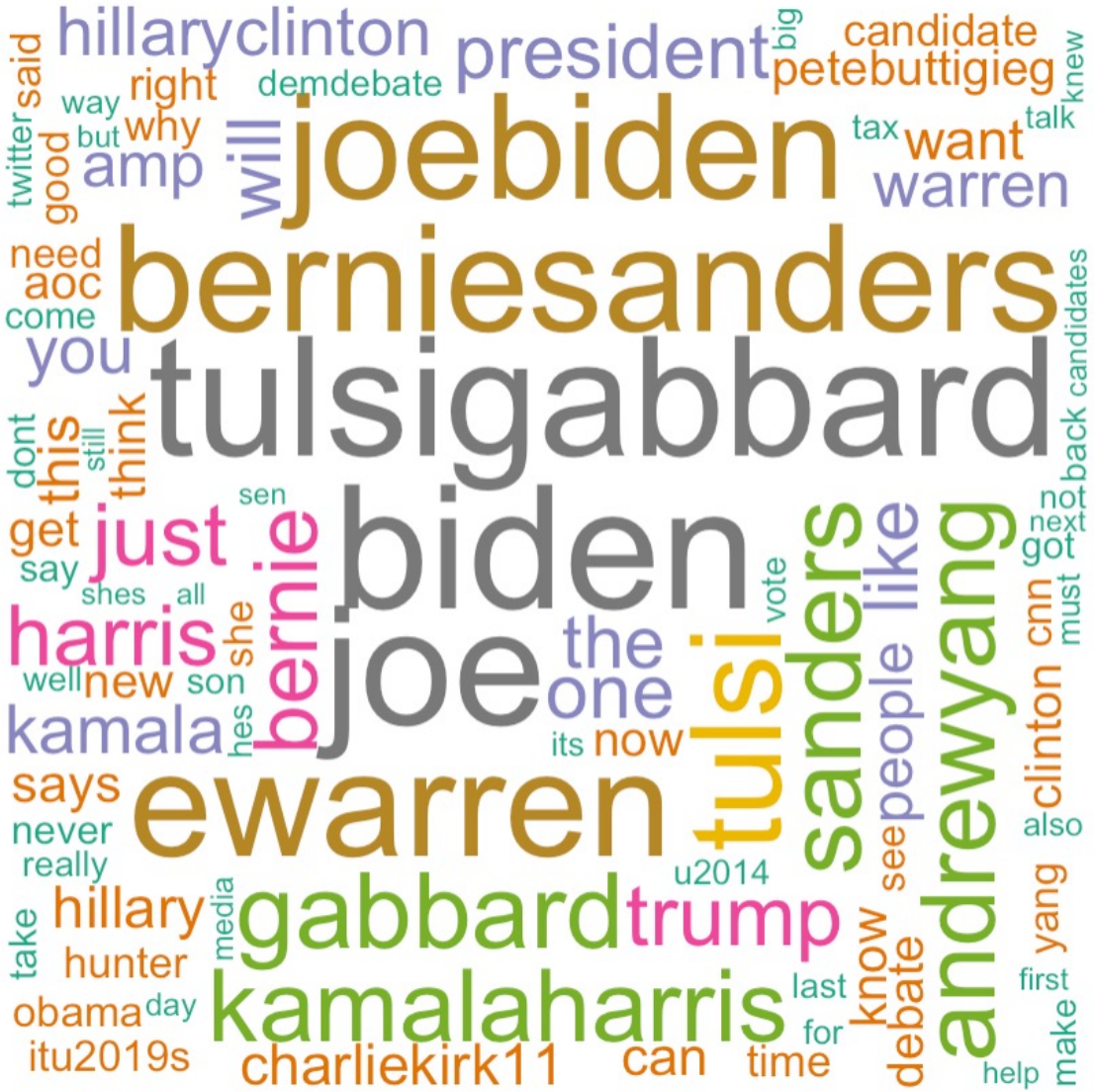
1. **Twitter streaming API**

2. **Odd checkers Website**

3. **Google Trends**

What did tweets say?

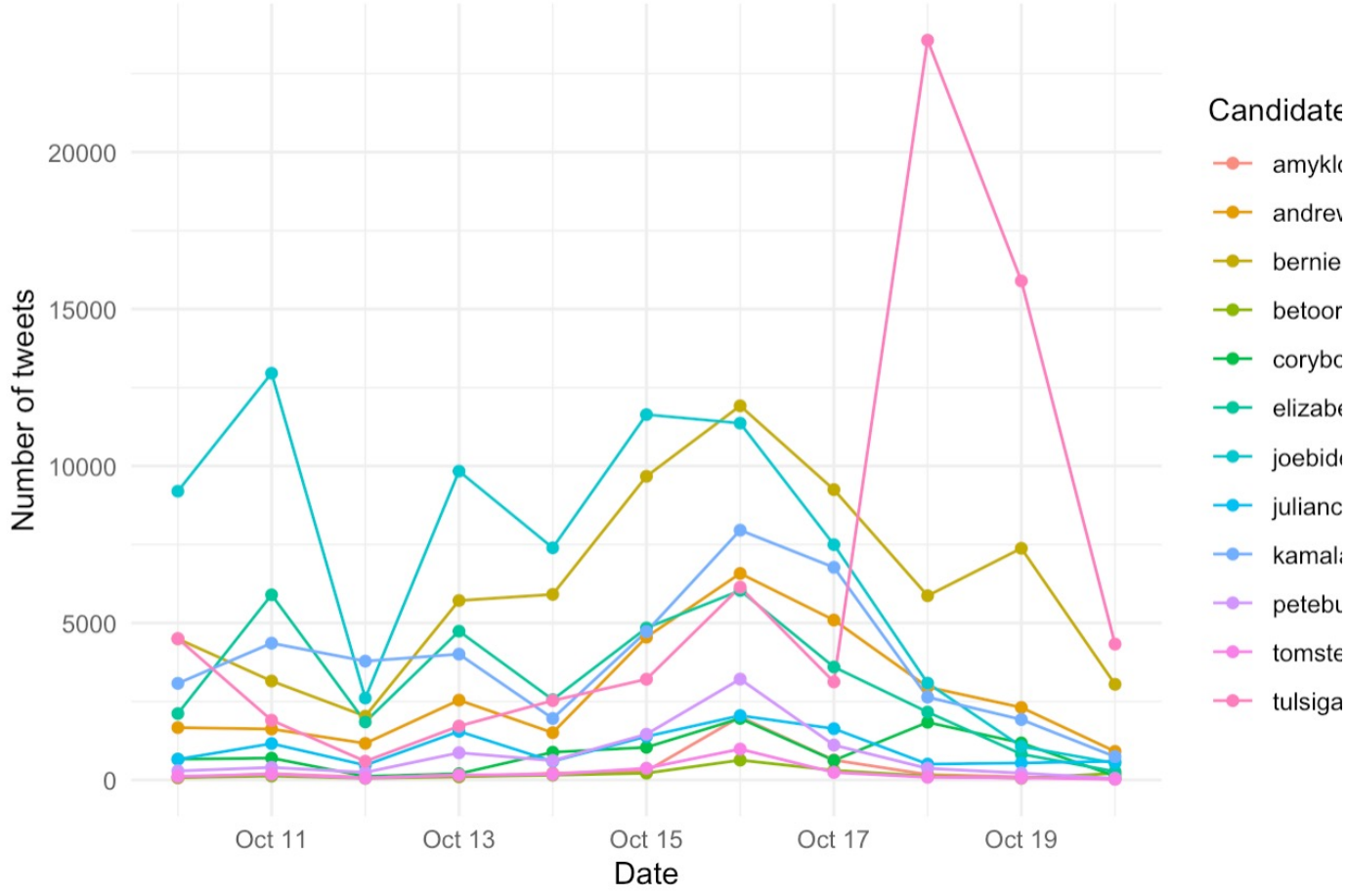
- The contents we collected are highly correlated with our research topic.
- Tulsi Gabbard, Bernie Sanders and Joe Biden are top three in terms of tweets volume.



Numbers of Tweets Change Over Time

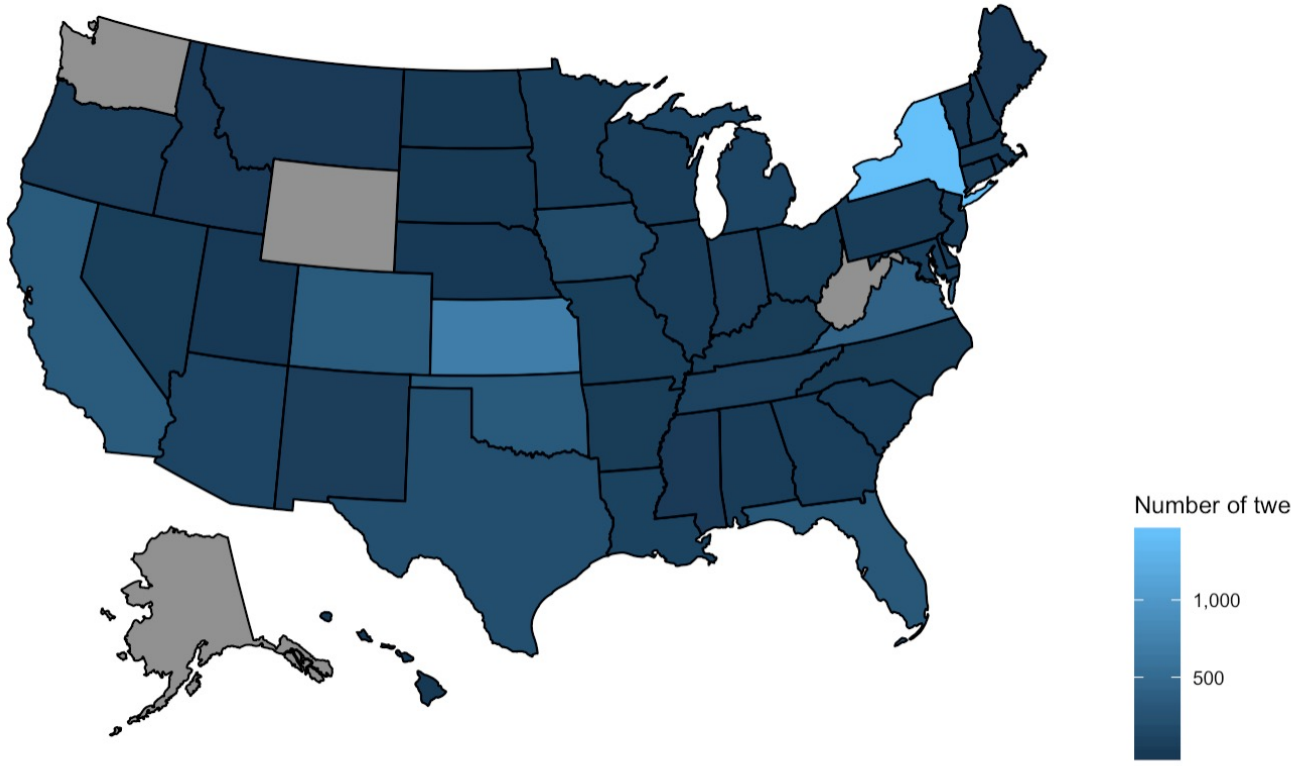
Changes in the number of tweets and the public support for candidates may be correlated.

There are more tweets related to Joe Biden, Bernie Sanders, Kamala Harris, Cory Booker, Amy Klobuchar, and Andrew Yang, compared to the tweets related to another half of the candidates.

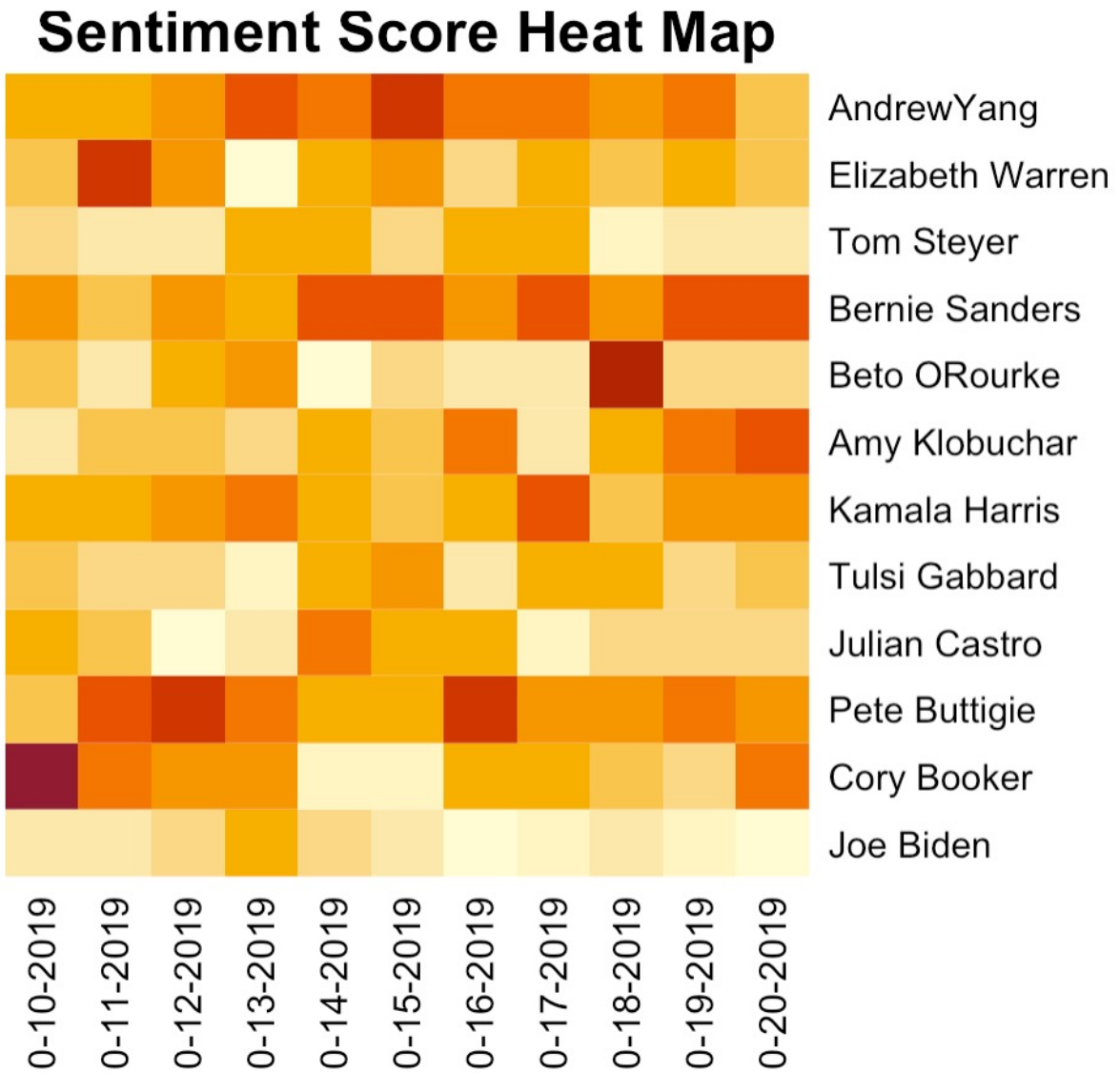


Where did the tweets come from?

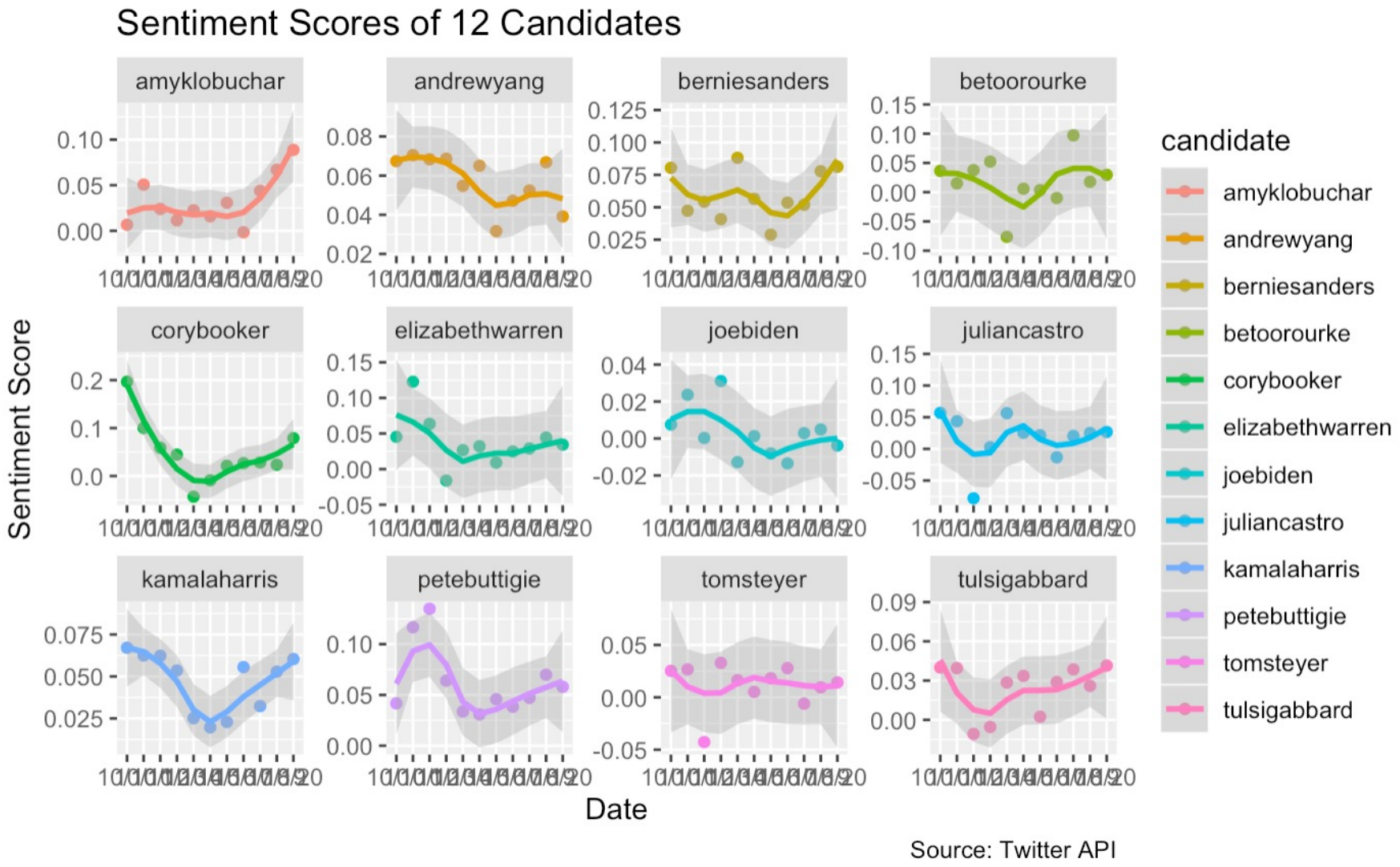
The map indicates the number of tweets that came from each state. The Twitter data we collected covers most of the states in the U.S. and can be comparable to a nationally representative sample of US likely voters.



Sentiment Analysis



Sentiment Score for 12 candidates

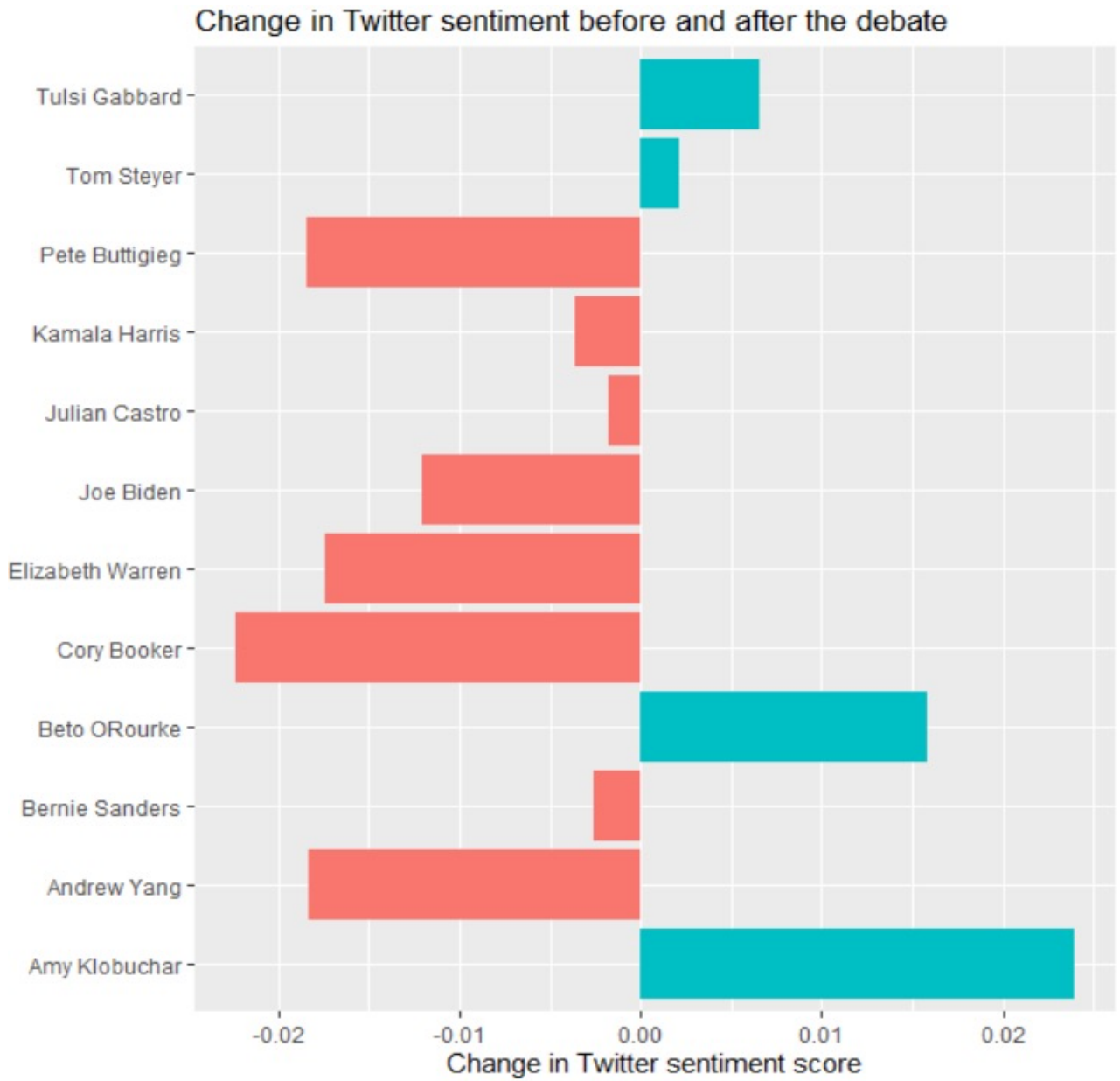


Sentiment Score for 12 candidates(cont’d)

- Positive Trend: Amy Klobuchar, Bernie Sanders and Tulsi Gabbard
- Constant : Elizabeth Warren, Joe Biden, Julian Castro, and Tom Steyer
- U shape Trend: Beto O’rouke, Cory Booker, Kamala Harris, and Pete Buttigieg
- No one has significant negative trend

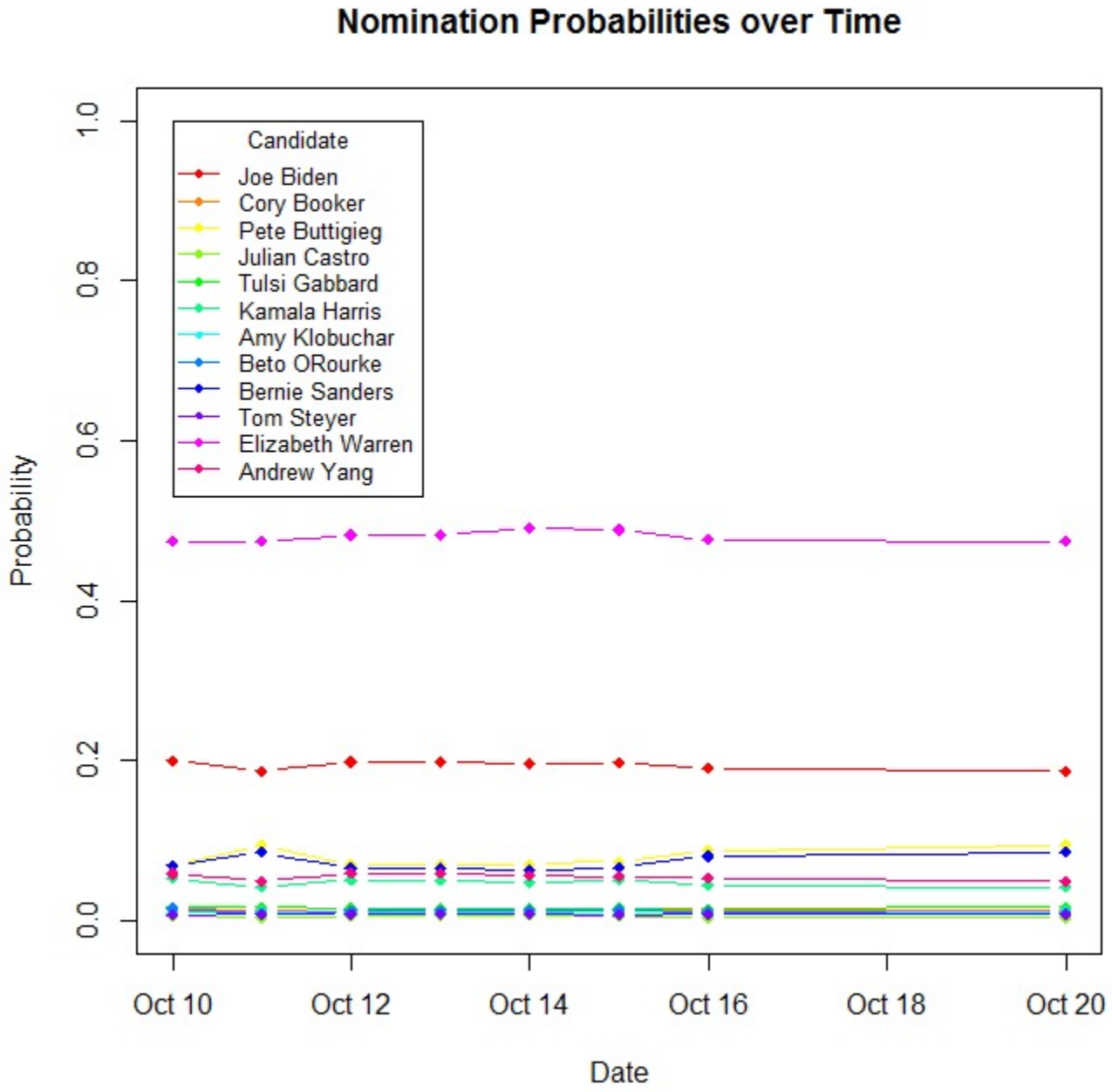
How the sentiment changed before and after the debate?

- Compute the mean sentiment score of the last 6 days before the debate, and the 5 days following the debate.
- Compute the mean difference
- See whether the change is positive(green) or negative(red).



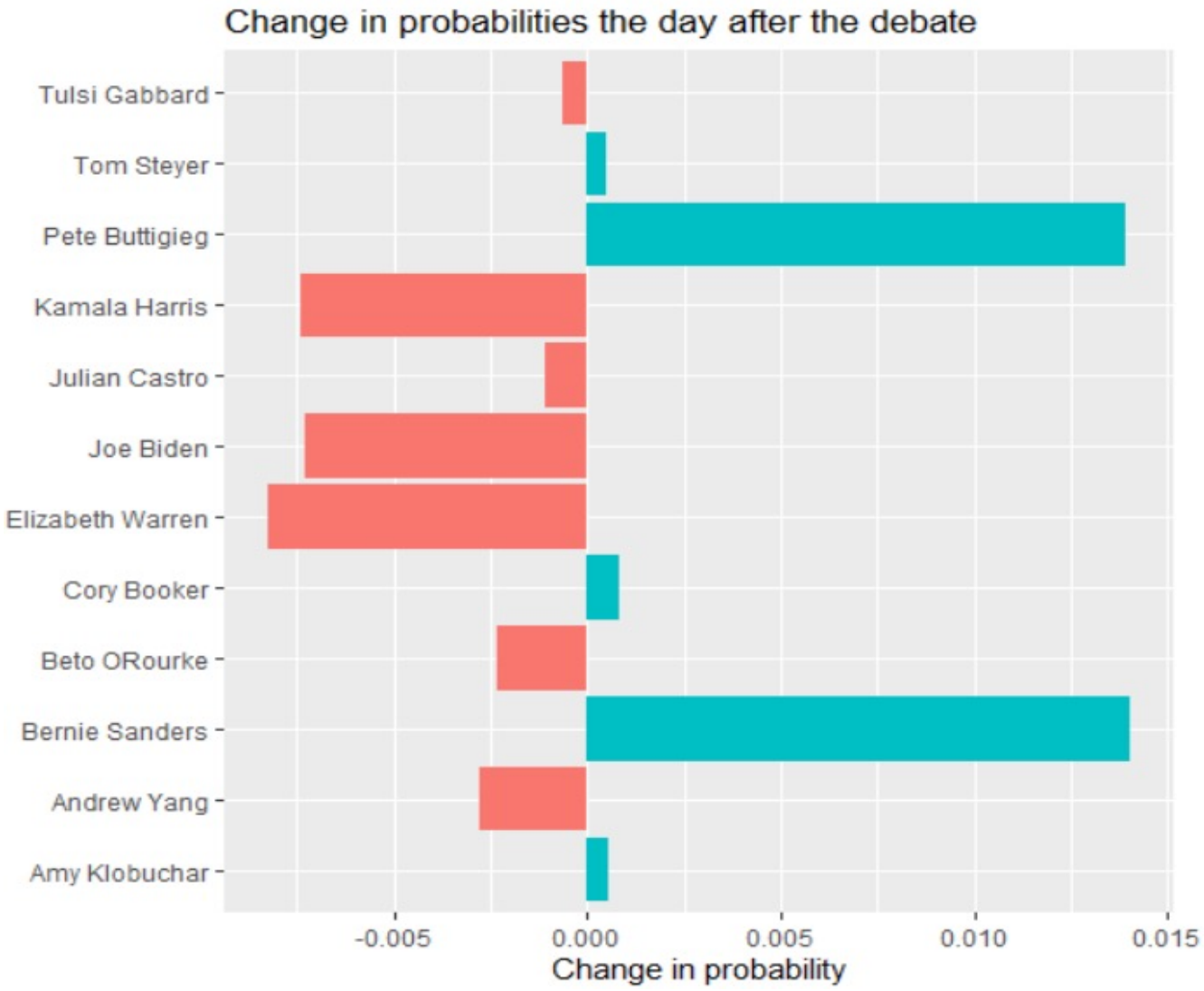
Odds Checker Probabilities before and after the debate

These increases and decreases in implied probabilities of winning the nomination matched almost perfectly with political analyses by mainstream news organizations of debate performances of each candidate. This justifies that the odd schecker data can be a very good **real-time indicator** of dynamic shifts in voter support **before any opinion poll result is released.**



Odds Checker Data – Implied Probabilities

Comparison of odds checker data from 8pm on October 15th (the start of the CNN debate) and 7pm on October 16th (the evening following the debate night) is likely to reflect perceived debate performances of respondents. The implied probabilities for Elizabeth Warren, Joe Biden and Kamala Harris have each dropped by close to 1%, whereas the probabilities for Pete Buttigieg and Bernie Sanders have both increased by more than 1%.

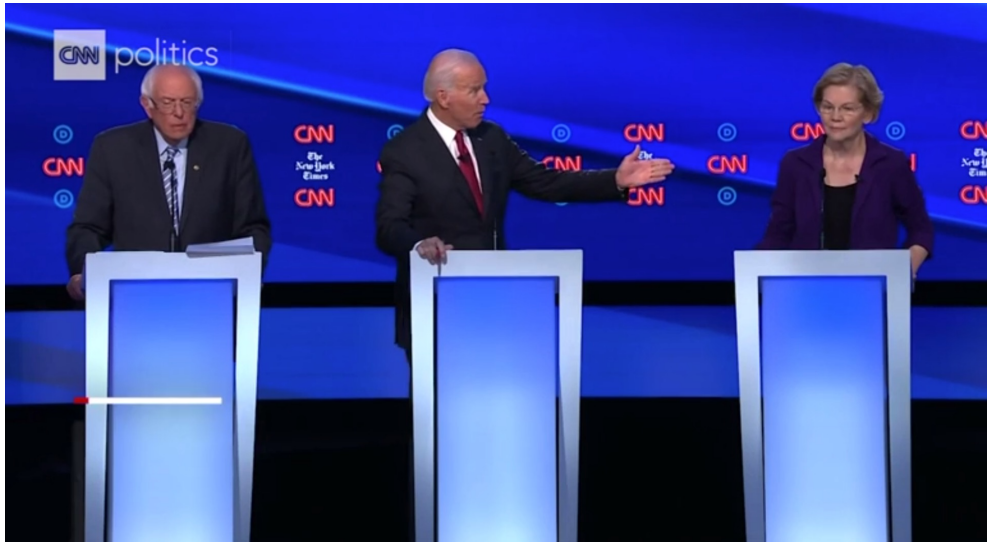


Winners and Losers of the Debate based on Oddschecker vs CNN Political Analysts

- Oddschecker Results:

- Winners: Pete Buttigieg, Bernie Sanders

- Losers: Joe Biden, Kamala Harris, Elizabeth Warren



- CNN Editor Analysis:

- Winners: Pete Buttigieg, Amy Klobuchar, Bernie Sanders, Andrew Yang

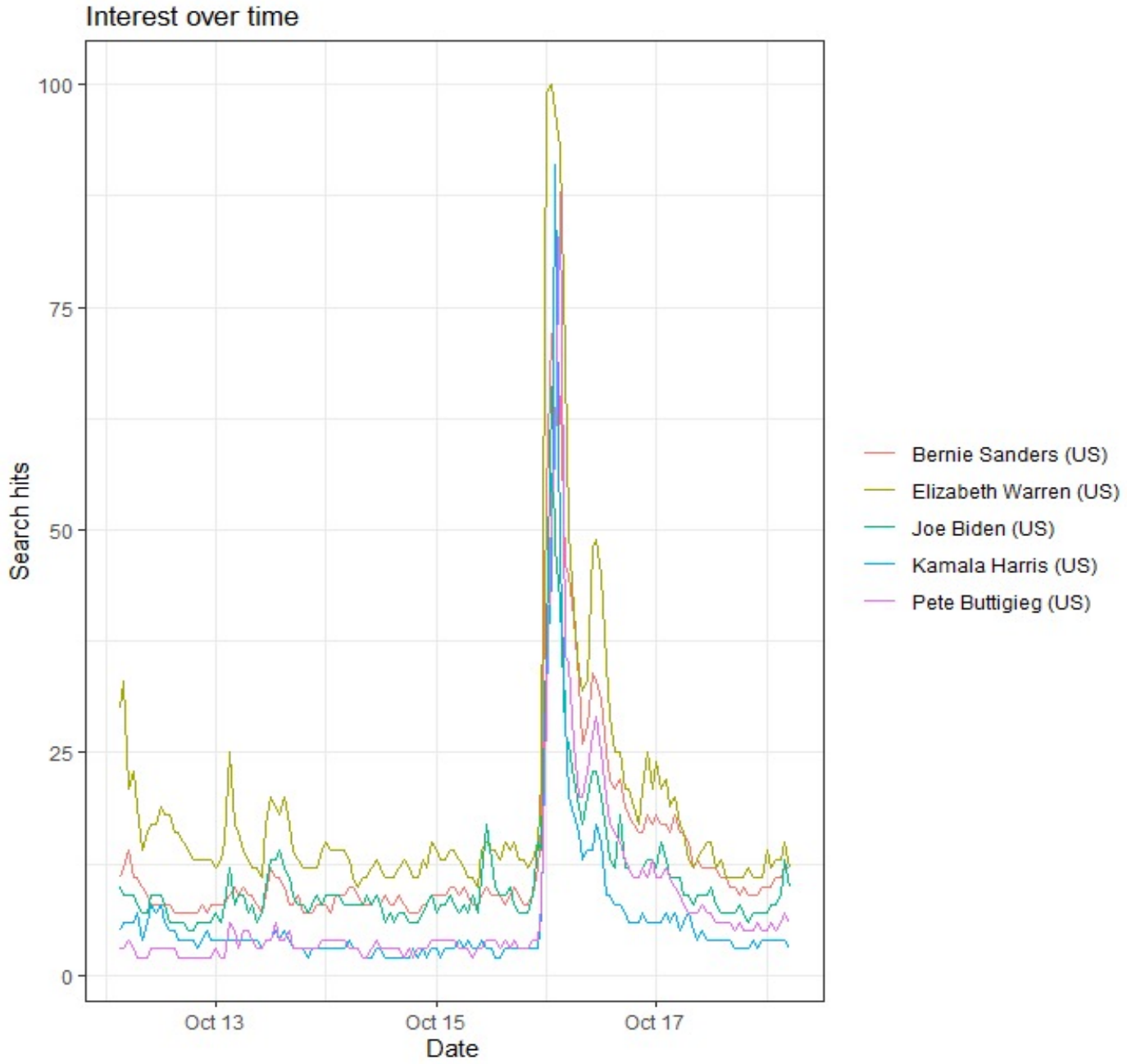
- Losers: Joe Biden, Kamala Harris, Elizabeth Warren, Tom Steyer

Analysis by Chris Cillizza, CNN Editor-at-large

https://www.cnn.com/2019/10/15/politics/who-won-the-democratic-debate/index.html

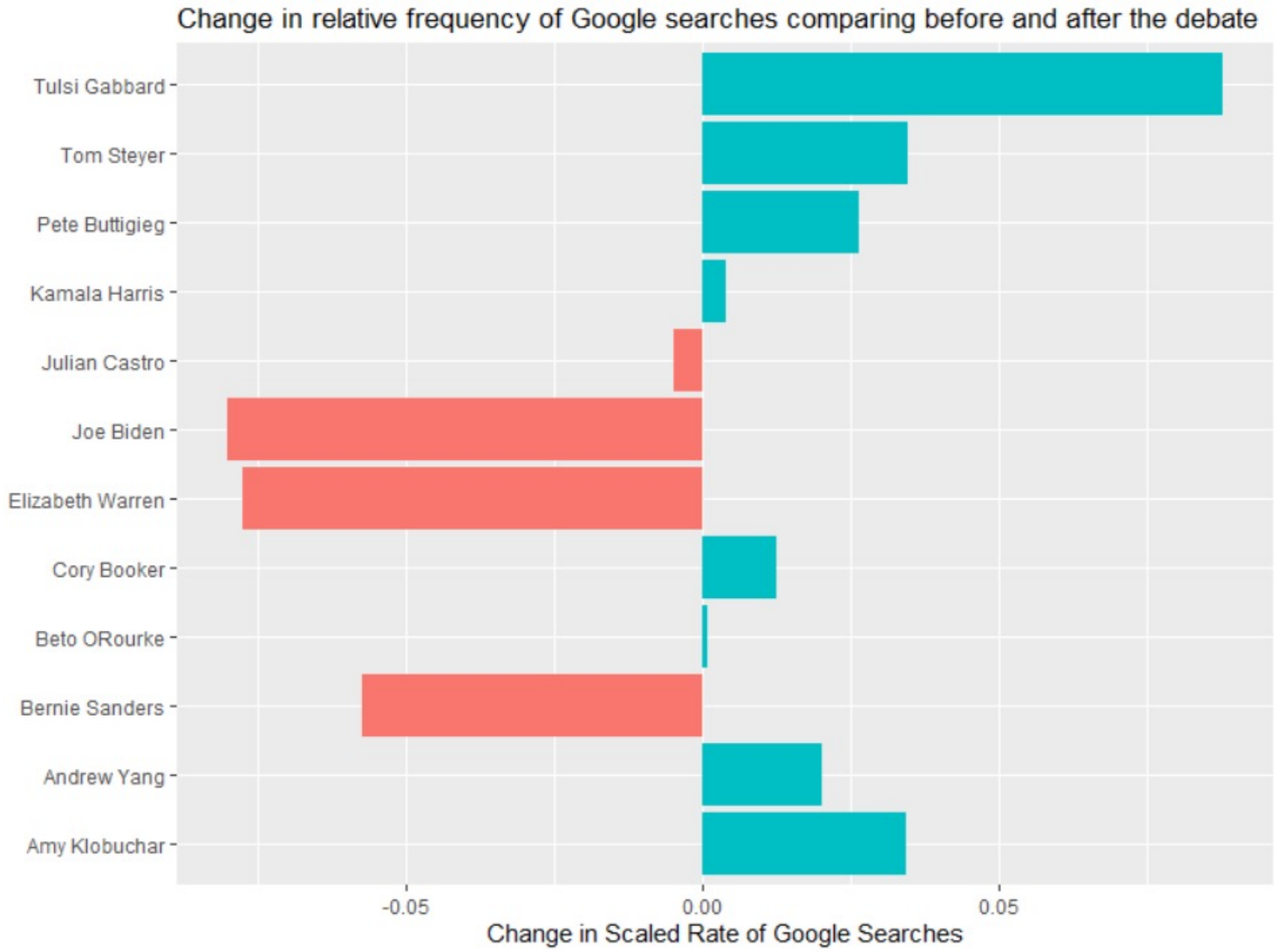
Google Trends over Time

Google search statistics between October 10 and October 20 have been gathered for each candidate. There is a clear spike in Google searches on the night of October 15 (during the debate) and the next day. High relative volume of searches for a candidate may correlate with increasing voter support for that candidate.

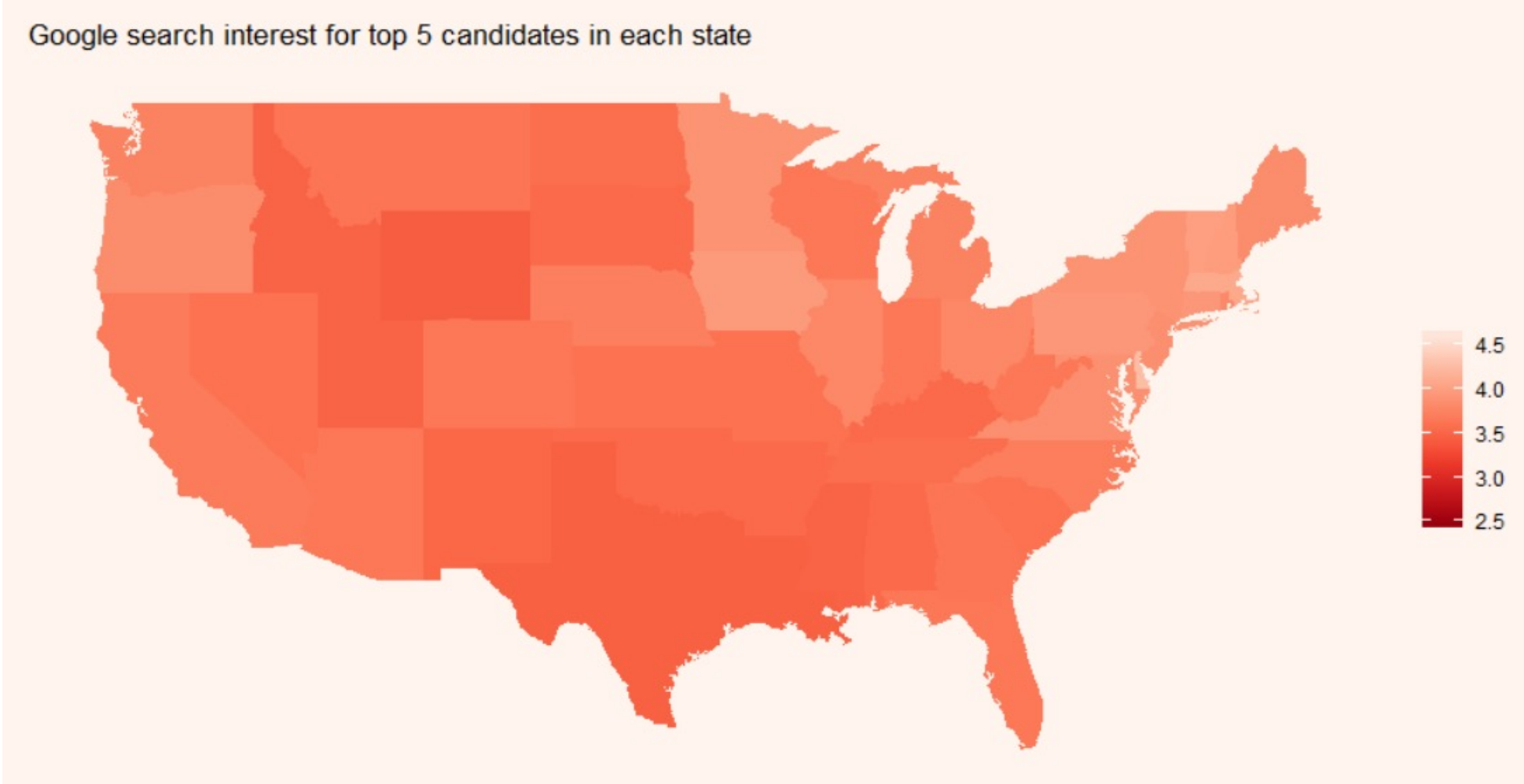


Comparison of Google Searches Before and After the Debate

Relative shares of Google searches of the top 3 candidates Joe Biden, Elizabeth Warren and Bernie Sanders have all dropped, whereas the relative shares by Tulsi Gabbard, Tom Steyer, Amy Klobuchar and Pete Buttigieg have increased after the debate.



Google search trends of candidates' names across states



Linear Model of Opinion Poll Estimates Jointly Using Google, Twitter and Oddschecker Data

```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              -0.001591   0.002359  -0.675    0.519
Change_in_Twitter_Sentiment              -0.189540   0.173958  -1.090    0.308
Change_in_Probabilities_by_Oddschecker    0.157429   0.329425   0.478    0.646
Change_in_Scaled_Google_Searches          0.122598   0.051648   2.374    0.045 *
```

- Changes in the relative number of Google searches of candidates' names is a statistically significant predictor of changes in voter support indicated by nationally representative surveys.
- Increasing implied probabilities of nomination of a candidate is an early predictor of increasing polling numbers of that candidate. More data is needed to justify statistical significance.
- Further work is needed to use Twitter sentiment scores as a predictive tool to forecast changes in voter support in national surveys.

Thank you for listening!

https://www.indiewire.com/feature/democratic-presidential-debate-1202181759/