Election Forecasting with Google

Grant Wilson

Agenda

A. Problem
    a. Polling error
B. Background
    a. Seth Stephens-Davidowitz
C. Goal
D. Google Trends
E. Election Simulation
    a. Methodology
F. Poll Weighting
    a. Recency
    b. Pollster Rating

https://www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark/

Problem

Pew Research Article

- Anti-institutional feelings of Trump supporters possibly made them less likely to respond to polls
- Social Desirability

> “Some have also suggested that many of those who were polled simply were not honest about whom they intended to vote for. The idea of so-called “shy Trumpers” suggests that support for Trump was socially undesirable, and that his supporters were unwilling to admit their support to pollsters.” - PewResearchCenter

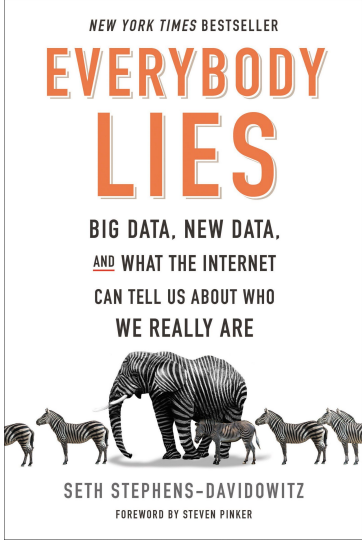
- Relevant Issue
  - **Polls in 2016 were inaccurate and underrepresented support for a “socially undesirable” candidate**

Background

- Seth Stephens-Davidowitz (author of Everybody Lies)
  - Effect of racial animus on the 2012 Presidential Election
  - Google Trends as an indicator of voting behavior

    “Barack Obama won 52.9 percent of the popular vote in 2008 and 365 electoral votes, 95 more than he needed. Many naturally concluded that prejudice was not a major factor against a black presidential candidate in modern America. My research, a comparison of Americans’ Google searches and their voting patterns, found otherwise. If my results are correct, racial animus cost Mr. Obama many more votes than we may have realized.”

- **Do Google Trends searches vary across typical “Democratic” and “Republican” voting states?**
  - **Yes**

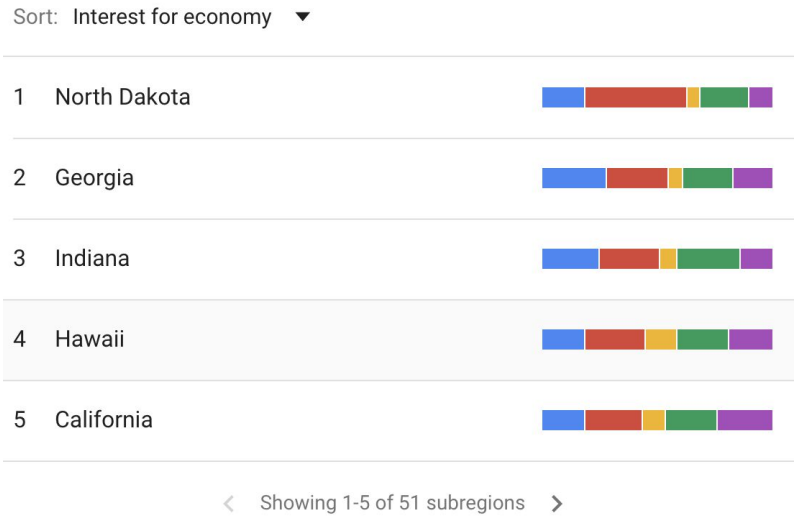


Goal

1. Correct for polling inaccuracies with another indicator less susceptible to biases
2. Create a model to predict outcome for a state based on Google Search Trend Data

Google Trends

- Google will track the popularity of search terms on geographic scales
- Relative popularity
  - States with different political leanings will have different proportions of search terms
- Google Trends predictions
  - Weighted by recency
    - Current search trends
    - Account for past trends

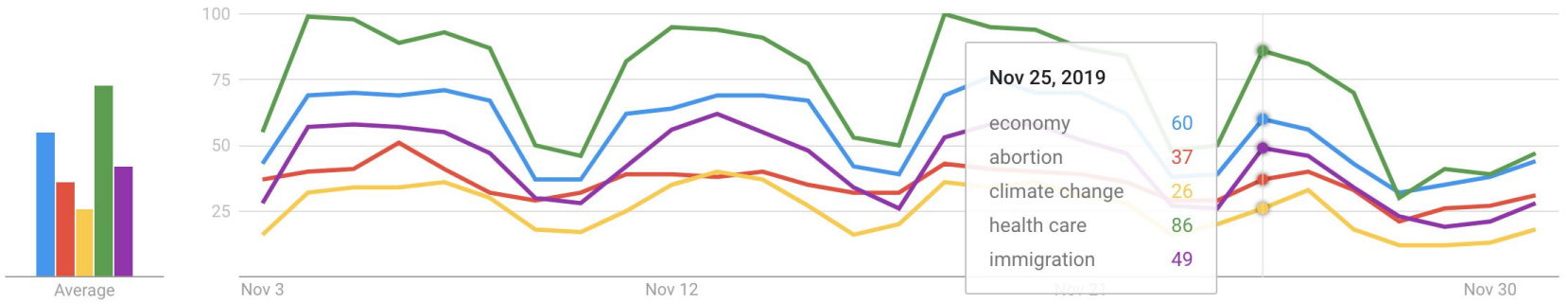


Google Trends | Compare

- economy — Search term
- abortion — Search term
- climate change — Search term
- health care — Search term
- immigration — Search term

United States | Past 30 days | All categories | Web Search

Interest over time



Election Simulation

1. **Single Election**
    a. For states without polls
        i. Assigned to Nominee of the party voted with in the majority of the last 7 elections
    b. For states with polls
        i. 100 simulations of an election in that state
            1. Polling plus Google
    c. Count electoral votes
2. **Multiple Elections**
    a. 1000 simulations of an election

https://fivethirtyeight.com/features/the-polls-are-all-right/

2016's presidential polls were about as accurate as average

Weighted-average error in polls in final 21 days of the campaign

CYCLE	NATIONAL	STATE	COMBINED
2016	**3.1**	**5.2**	**4.8**
2012	3.3	3.7	3.6
2008	2.3	3.9	3.6
2004	2.2	3.5	3.2
2000	3.9	4.6	4.4
1996	6.4	4.8	5.3
1992	4.6	5.2	5.1
1988	3.5	5.0	4.6
1984	5.4	4.5	4.7
1980	8.9	8.6	8.6
1976	2.5	3.8	3.4
1972	2.6	4.6	4.3
Average	**4.1**	**4.8**	**4.6**

Pollsters that are banned by FiveThirtyEight because we know or suspect that they faked their data are not included in the averages. Averages are weighted based on the number of polls a particular firm conducted.

Election Simulation

- Election
  - Simulated at the state level
  - Ignored national polling in favor of statewide
    - Electoral College
- Randomness
  - Add random error term to win likelihood
- Google Trend prediction
  - Add a term to account for Google Trends prediction for the state

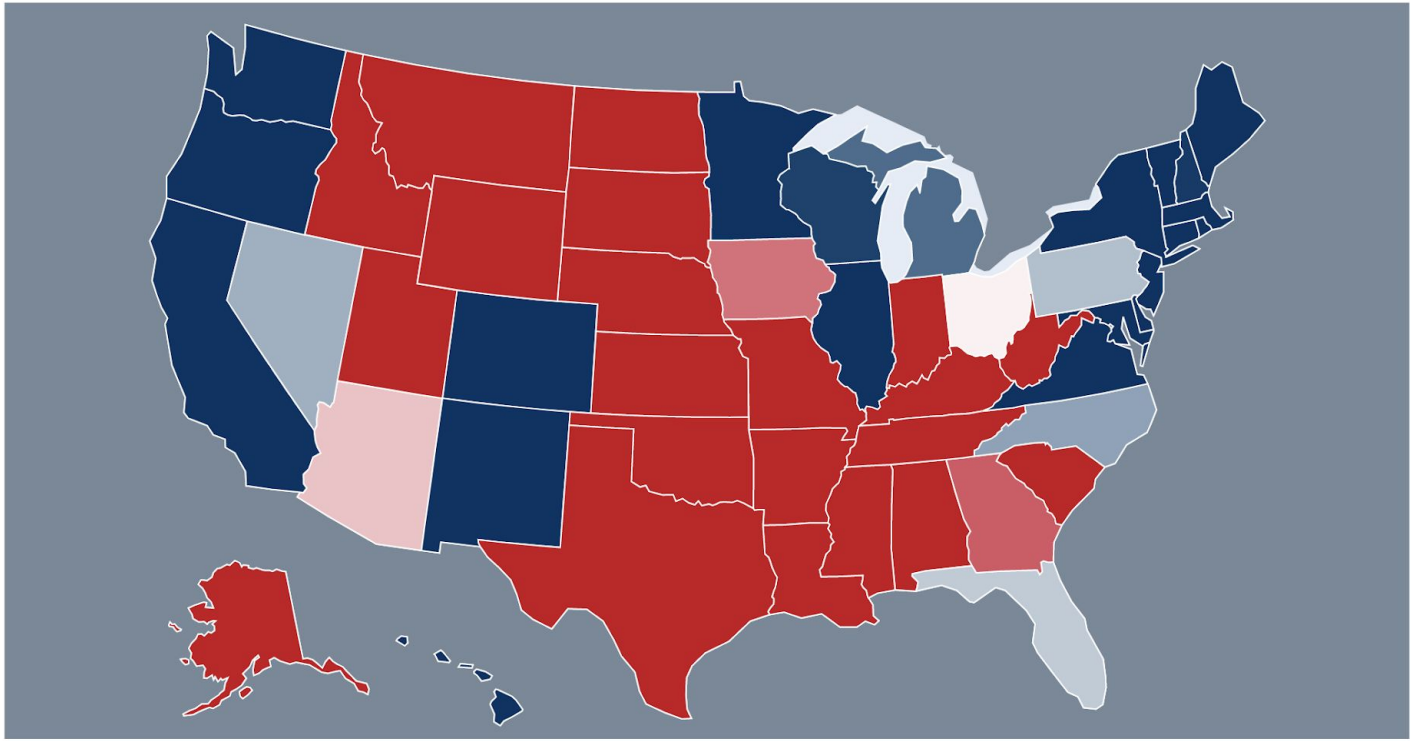
https://fivethirtyeight.com/features/how-fivethirtyeight-calculates-pollster-ratings/

Weighting

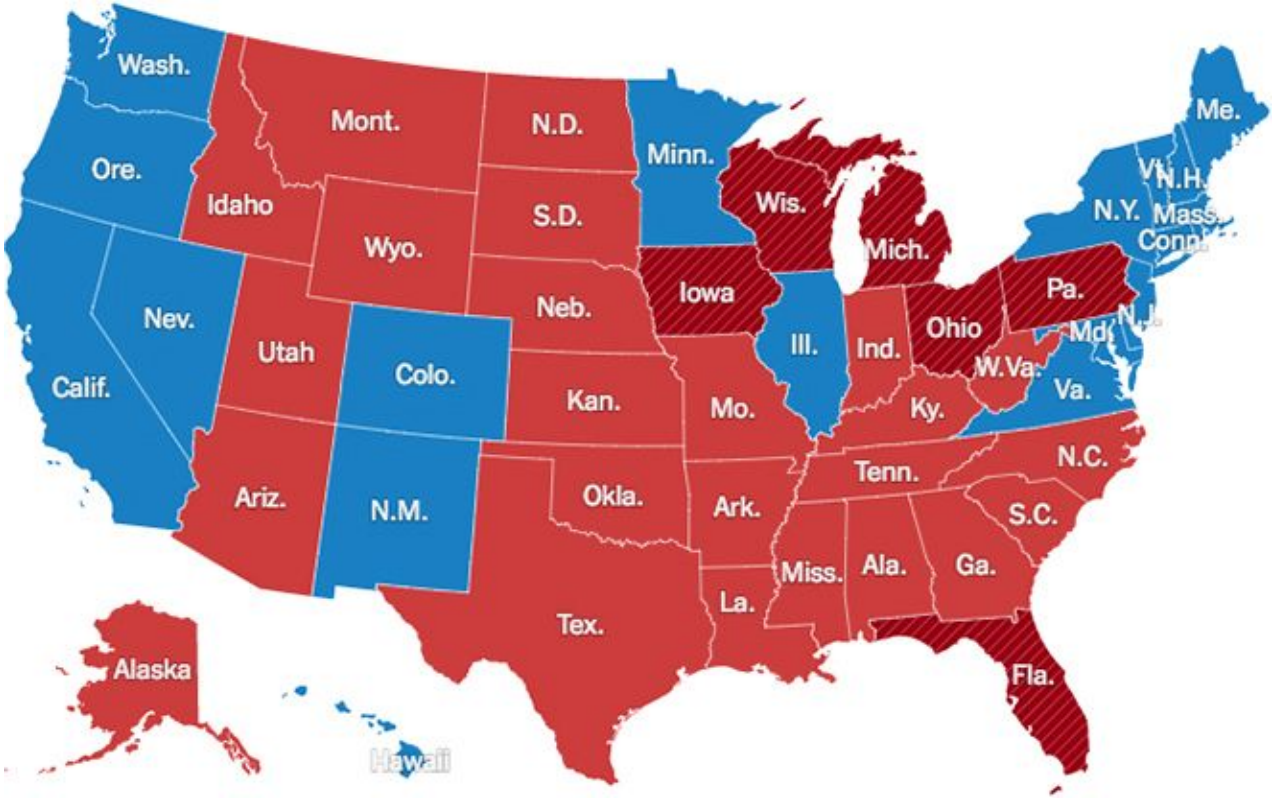
- FiveThirtyEight was more successful in 2016 in part due to their weighting of their collected polls
  - Pollster Rating
    - Type of Poll
    - Accuracy
    - Methodology
  - Recency
    - All forecasting models should account for past polling data, but should focus on recent polls

https://www.gwilson.org/2020-election-forecasting-1

2020 Donald Trump vs. Elizabeth Warren Simulation Results
(Hover for breakdown)



https://www.nytimes.com/elections/2016/results/president



Conclusion

1. Model trained on 2016 Google Trends Data alone successfully predicted 47 of 51 “states” in the 2016 Presidential Election.
    a. Includes D.C.
2. Google Trends data can add another dimension to polling and highlight battleground states not identified by polling alone.

Data Sources

- Pollster Rankings
  - FiveThirtyEight's Pollster Rankings
- Presidential Election Polls
  - FiveThirtyEight Polls: Presidential
- Google Trends
  - Google Trends Home Page
- Google Trends Unofficial API
  - General Mills Python Google Trends API
- Historic Voting Trends 1976-2016
  - Harvard Database
- Ballotpedia
  - Ballotpedia Presidential Election Accuracy