


538 Election Data for the 2023/2024 Election Cycle



Kaitlyn Zhang

Background:

- The results of the 2016 election were surprising for many people.
- Despite many credible pollsters predicting a relatively easy win for Democratic nominee Hillary Clinton, the Republican nominee Donald Trump won the electoral college by a comfortable margin.
- 538 predicted a 71.4% chance of Clinton winning and Trump was only given a 28.6% chance (Silver 2016).
- Despite this, Nate Silver, the founder of 538, says that they “...think our model did a good job in 2016” (Silver 2020).
- He argues that although they predicted Clinton would win, their forecast for Trump was much higher than other organizations. In fact, on October 31, 2016, 8 days before the 2016 election, Nate Silver (2016) pointed out that there is an increased chance of the popular vote splitting with the electoral vote.

Chance of winning

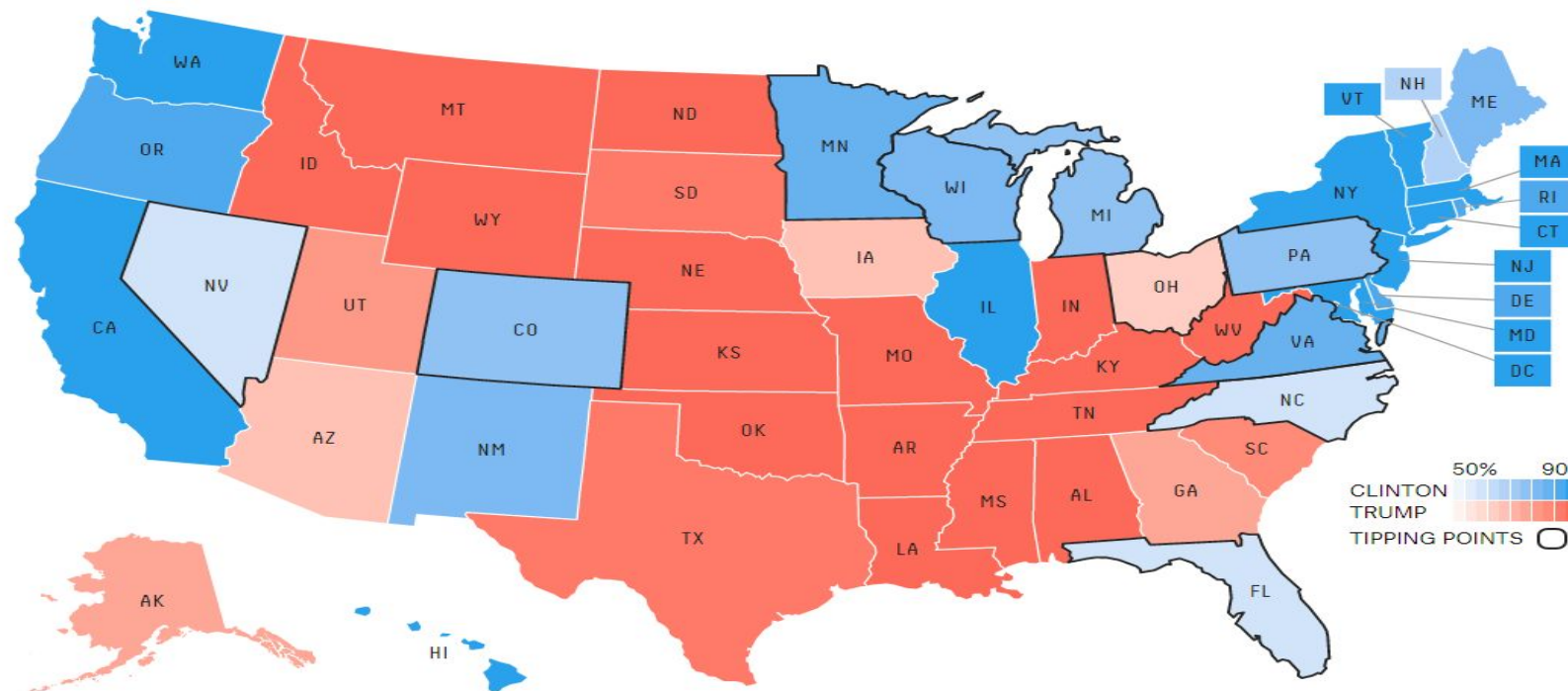


Hillary Clinton

71.4%

Donald Trump

28.6%



Background

- I wanted to explore the data that 538 provides
- Scaled up this database can be used to accurately predict elections

general_polls

poll_id
pollster_id
pollster
start_date
end_date
sample_size
cycle
office_type
stage
dem
rep

- General Election polling, looking to see if voters would vote Democrat or Republican
- Date Range: 11/18/22-4/20/23
- Election Cycle: 2024
- 24 rows, 11 columns
- https://zhangkaitlyn1.shinyapps.io/general_polls

gov_polls

poll_id
pollster_id
pollster
state
start_date
end_date
sample_size
cycle
office_type
stage
party
answer
pct

- Polled for voters' Governor preferences across states
- Date range: 3/2/2023-4/20/2023
- Polled for both general elections and jungle primary elections
- Jungle primary elections are when all candidates are in the same primary elections regardless of political party
- 2023/2024 cycle
- 19 rows, 13 columns
- https://zhangkaitlyn1.shinyapps.io/gov_polls/

house_polls

poll_id
pollster_id
pollster
methodology
state
start_date
end_date
sample_size
cycle
office_type
seat_name
stage
answer
pct

- Polled for voters' preferred house representative in different states
- Date Range: 3/5/2023-4/2/2023
- Election cycle: 2024
- 6 rows, 14 columns
- https://zhangkaitlyn1.shinyapps.io/house_polls/

primary_polls

poll_id
pollster_id
pollster
methodology
start_date
end_date
sample_size
cycle
office_type
party
answer
pct

- Polled for voters' preferences on Presidential candidates at the primary level
- Date range: 4/21/2023-4/25/2023
- Election cycle: 2024
- 19 rows, 12 columns
- https://zhangkaitlyn1.shinyapps.io/gov_polls/

president_polls

poll_id
pollster_id
pollster
methodology
start_date
end_date
sample_size
cycle
office_type
stage
party
answer
pct

- Polled for voters' presidential candidate preference at the general election level
- Date Range 4/24/2023-4/25/2023
- Election cycle; 2024
- 16 rows, 13 columns
- https://zhangkaitlyn1.shinyapps.io/president_polls/

Limitations

- Smaller amount of data, there is much more data available (like, a lot) so I will possibly scale this database up in the future
- No primary keys, election polls often about more than one election/election cycle, so pollster, poll_id, and pollster_id are not unique to each row
- I had to change the data I was using several times due to it being too large/not enough tables

Data Source:

- [Our Data | FiveThirtyEight](#)
- Updates frequently as new polls come out
- Includes recent data as well as historical data
- Scaling the database up to include historical data is a possibility in the future