The Accuracy and Predictiveness of State-Level Presidential Polls

Brittany Alexander

June 15, 2020

Outline

Introduction to Presidential Election Polling

- Presidential Election polls aim to predict both the winner and the proportion of votes for candidates nationwide and in individual states
- Predicting the winners of states is important because a candidate must win the most electors which is largely determined who wins a state
- The number of electors per states is the number of respresentives in Congress plus two.
- We define a state as competitive if it's margin on election day is between -5 and 5 points. Competitive states are of more interest.

Challenges of Using Polls to Predict Elections

- Election polls are taken before the election and voters change their minds
- Polls typically include undecided voters
- It is common for polls to ignore minor candidates

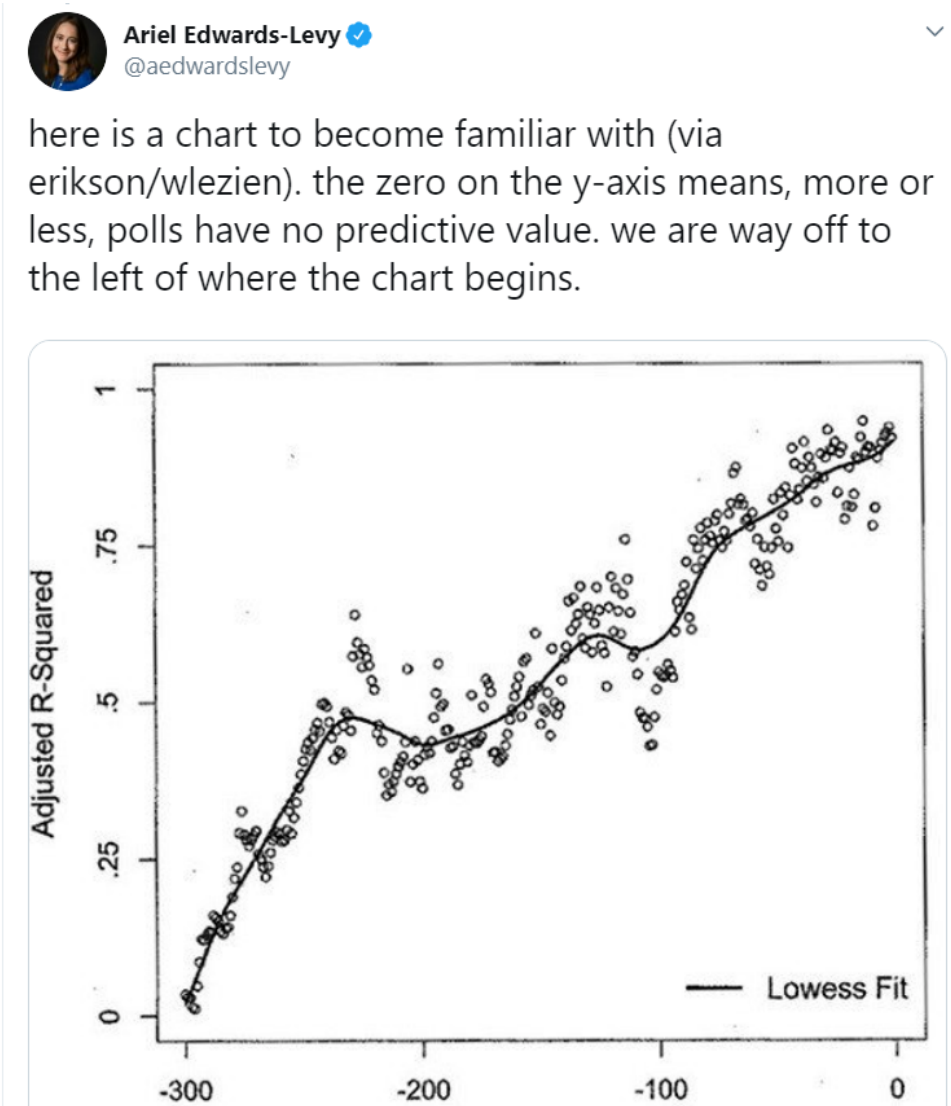
Data Source

Huffington Post Pollster has polling data for 2012-2016. There was a site to get 2008 data before Pollster merged with Huffington Post, but that link is broken. There are 5756 state level polls across the three elections. Most states have multiple polls for each election year. This data set also contains new variables to model polling across years. New variables includes:

- The election results as both the margin and the two party vote
- Days until the election at the start, end, and middle of a poll
- Various versions of polling error

Inspiration

Twitter Thread



Literature Review

- Hillygus, D. S. (2011) describes a history of election polling but doesn't include much on state level polls
- Bon et.al 2019 focused on the effects of undecided voters and polling bias
- Shirani-Mehr (2018) built a model to decompose bias and variance in polls but focused on the last two weeks of the election
- Alexander (2019) built a model and looks at the accuracy of averaging the polls
- None of these studies focuses on individual polls during a broad range of time

Accuracy

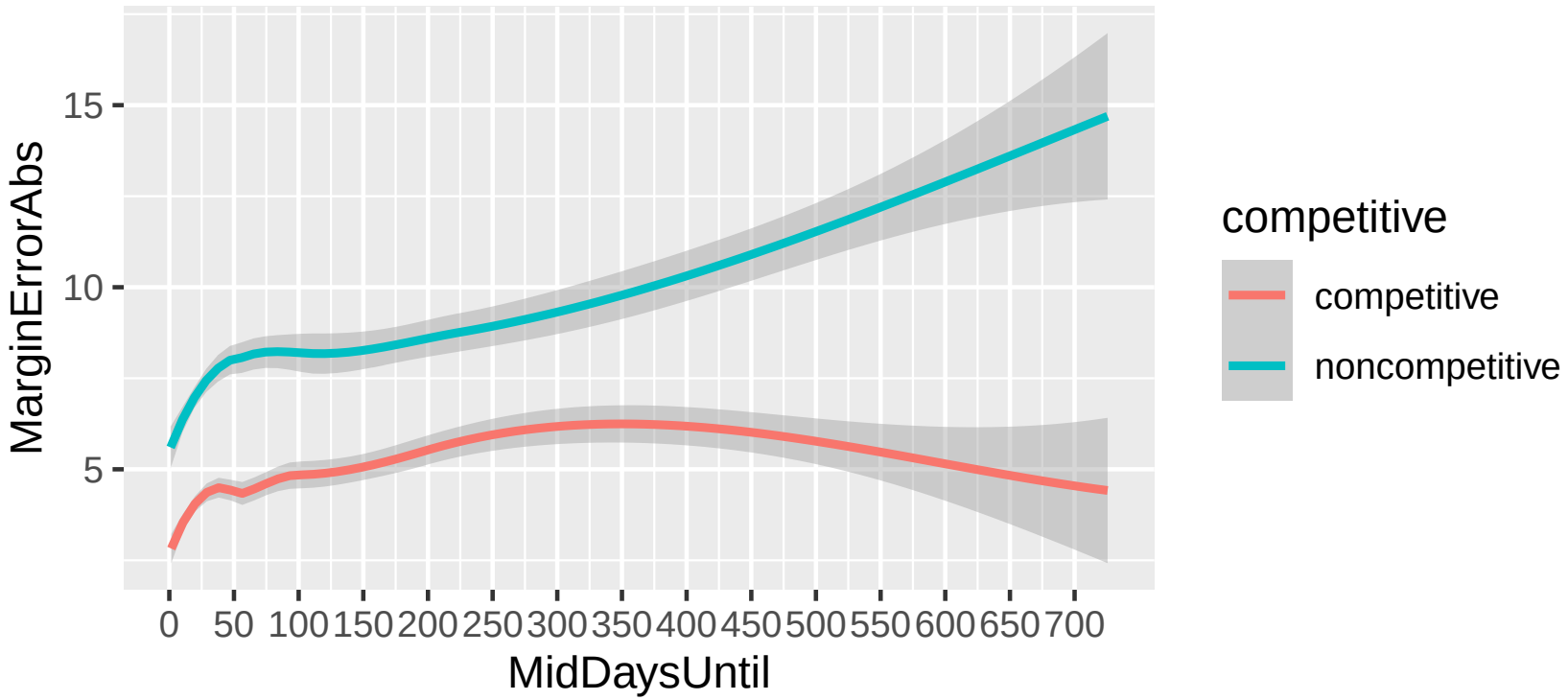
Accuracy in polling has two components: percent called correctly, and distance between a polls results and what happens on election day. Additionally accuracy can be viewed in terms of margin, and in terms of vote. Margin is defined as the difference between the Democratic vote (or poll support) and the Republician vote (or poll support). Accuracy in terms of vote is measured by first applying the formula: $d_{new} = \frac{d}{d+r}, r_{new} = \frac{r}{d+r}$ to polls and vote results so that the Republican and Democratic support sums to 1. This standardizes results to deal with different levels of undecideds, and the inclusion of minor candidates.

Predictiveness

Predictiveness is defined by the strength of a correlation between two variables and R^2 of fit regressions. Predictiveness is matters because it tells us if and when we can consider polls to have predictive value in the election.

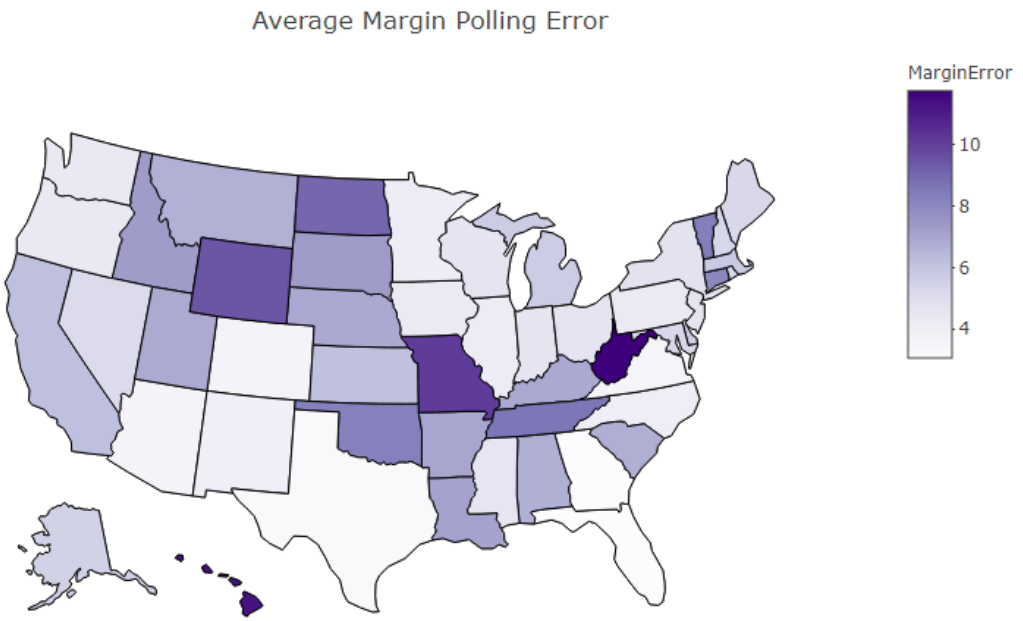
Accuracy Over Time

One key question is how accurate polls are far out from election day. 200 days would be approximately in late January (start of the primary), and 100 days before is the end of August. This uses lowess smoothing.



Average Error By State

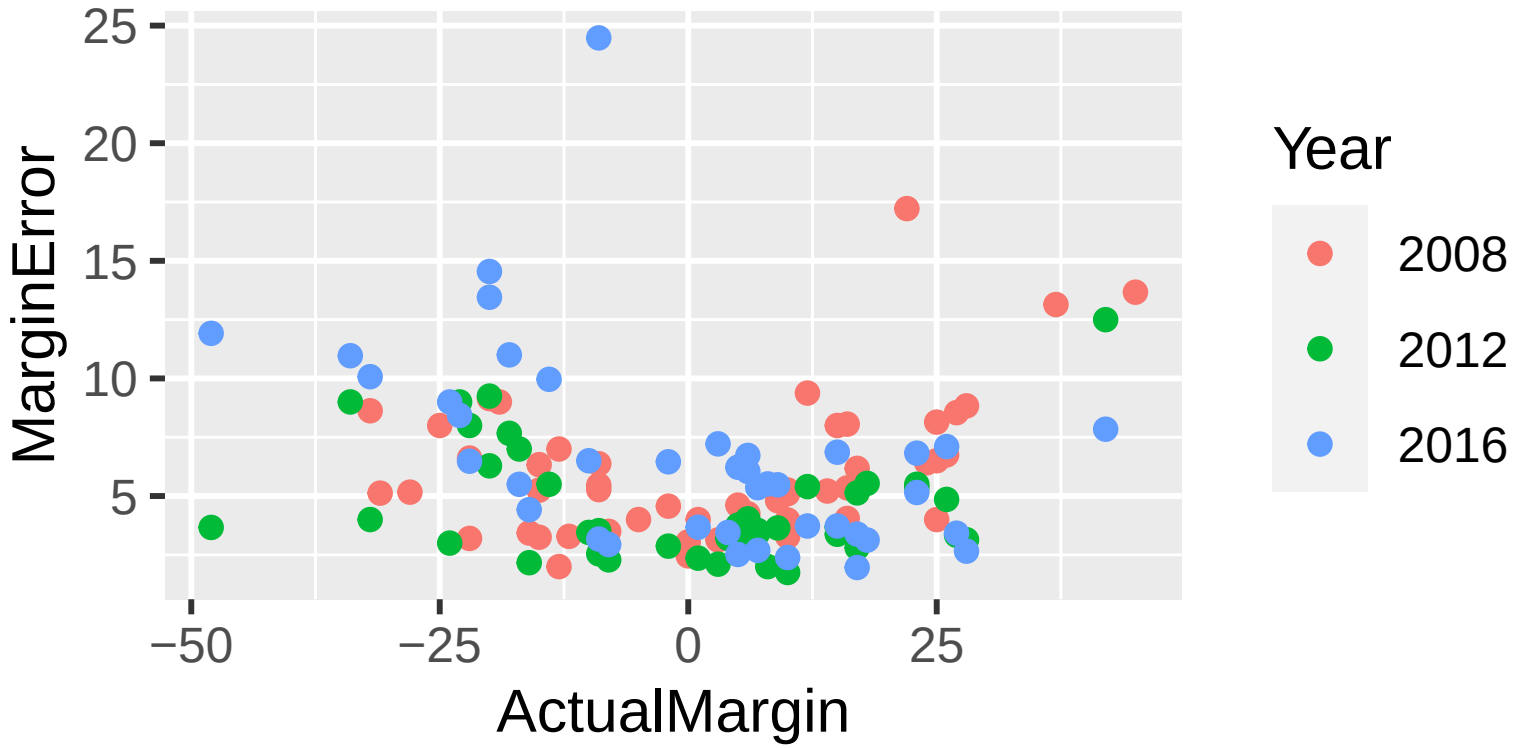
Below is a Choropleth of the Average Polling Error in the last 60 days until the election.



Accuracy By Partisanship

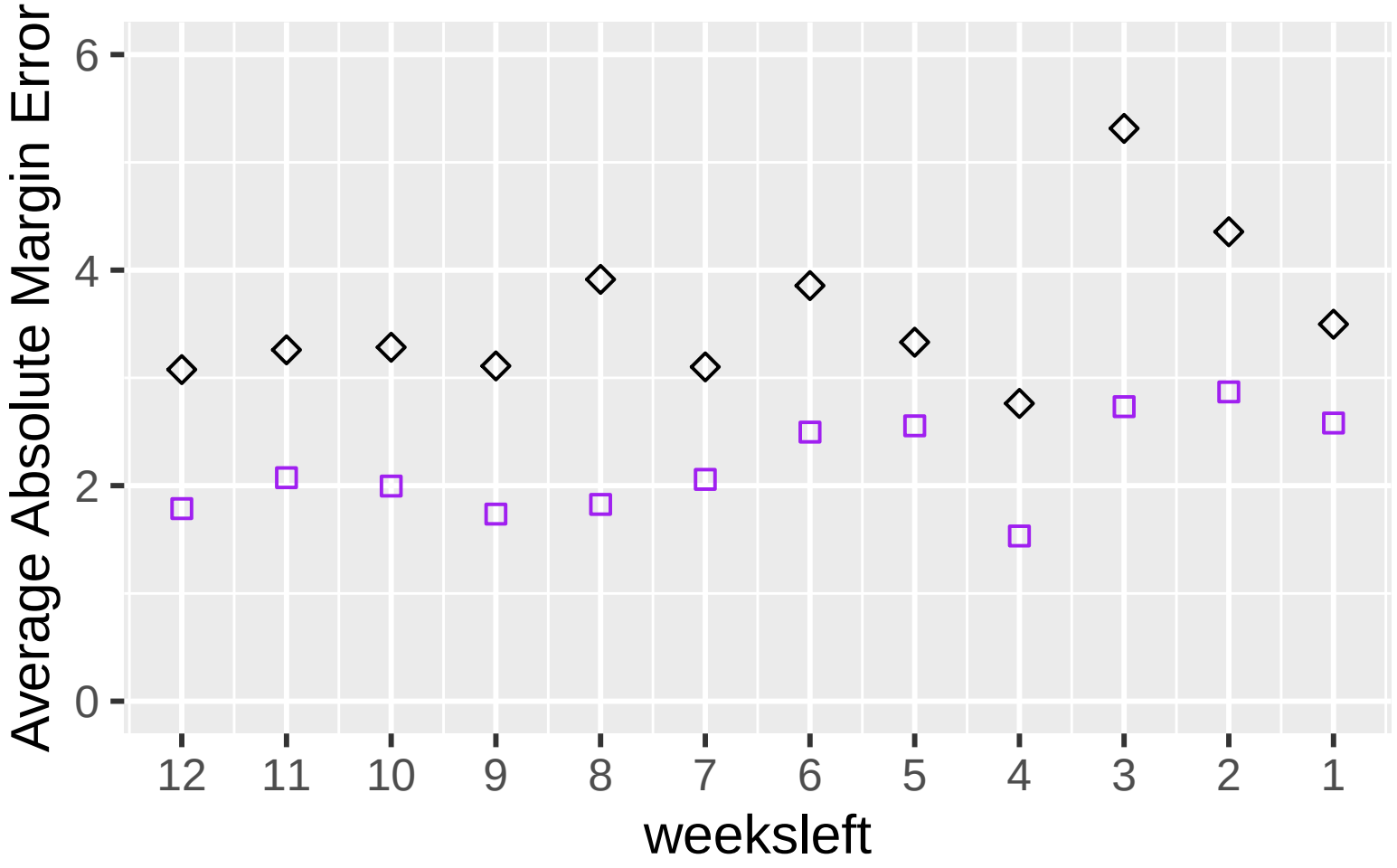
Below is a plot of the Average Margin Error of polls from the last 60 days for a state by it's actual margin on election day.

- Recall that proportions are most variable when they are closest to .5 suggesting that competitive states should have more sampling error.
- Possible explaination: Nonsampling factors such as poll quality, frequency polled, etc. explain this phenomon.



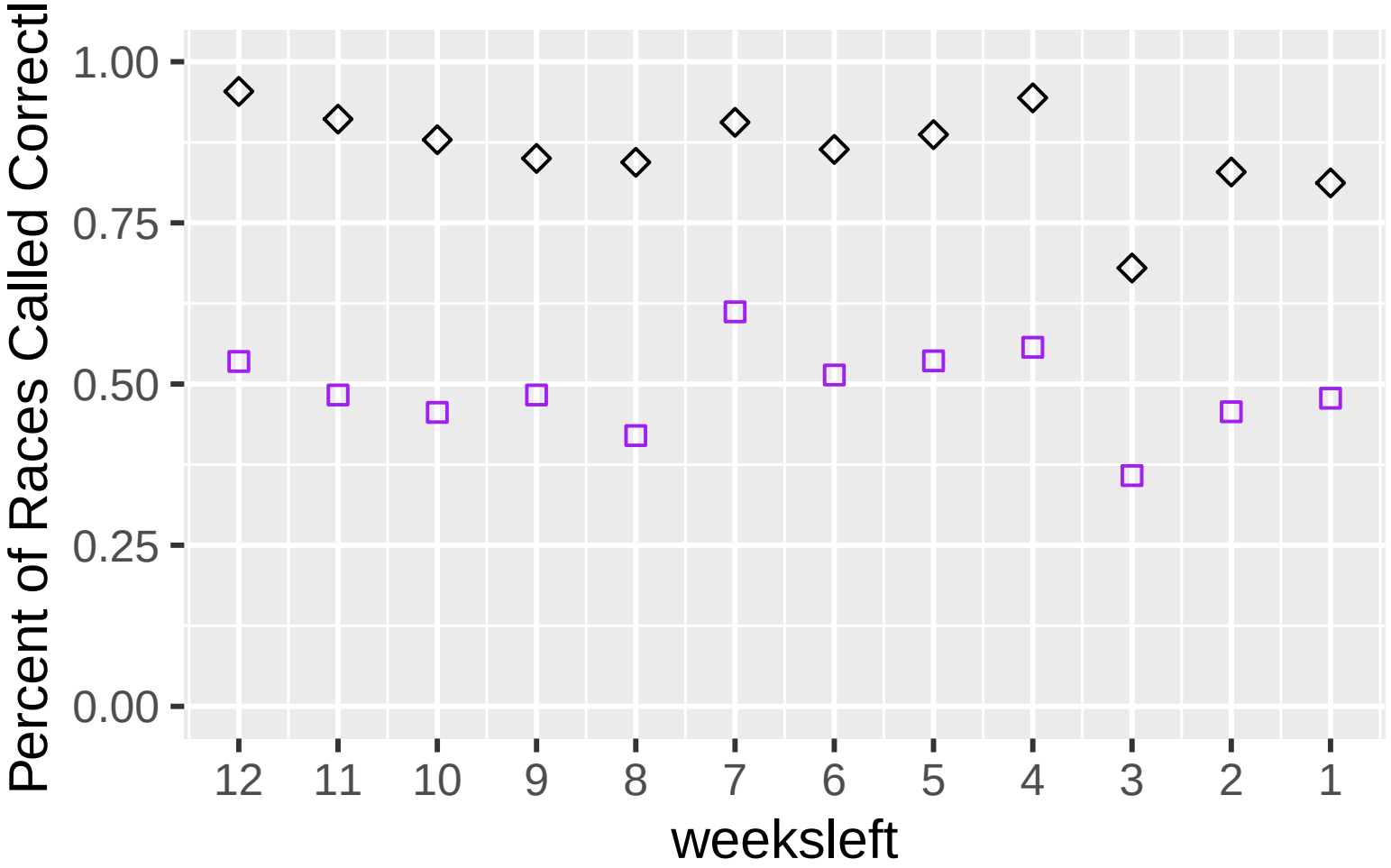
Margin Error in Competitive and Non-Competitive States

Below is a plot showing the average Margin error in competitive and noncompetitive states, broken up by week for the last 12 weeks of the election. Purple Square is competitive states, and Black Rhomus is noncompetitive states.



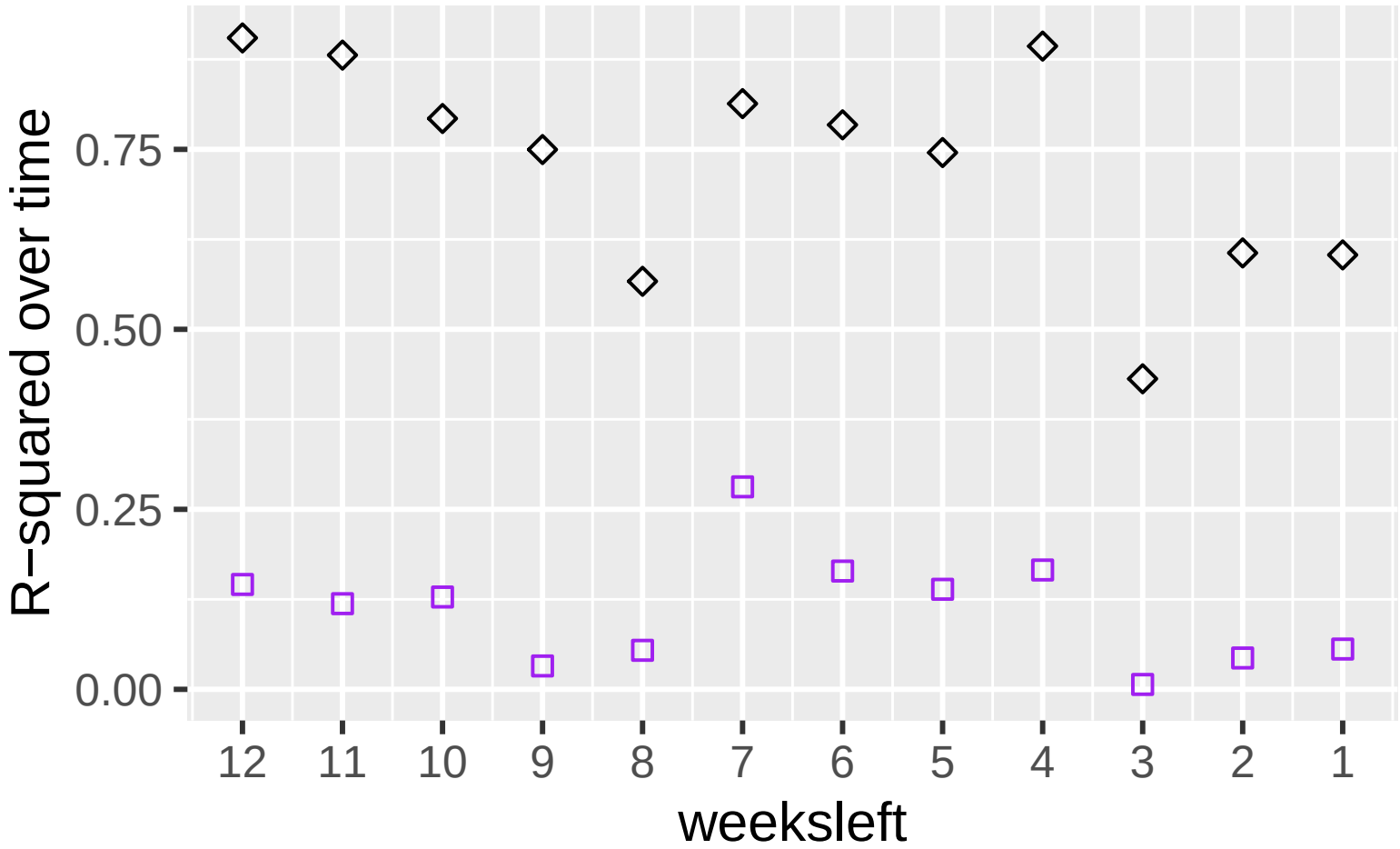
Percent of Races Called Correctly

Below is a plot of the percent of races called correctly. Purple Square is competitive states, and Black Rhomus is noncompetitive states.



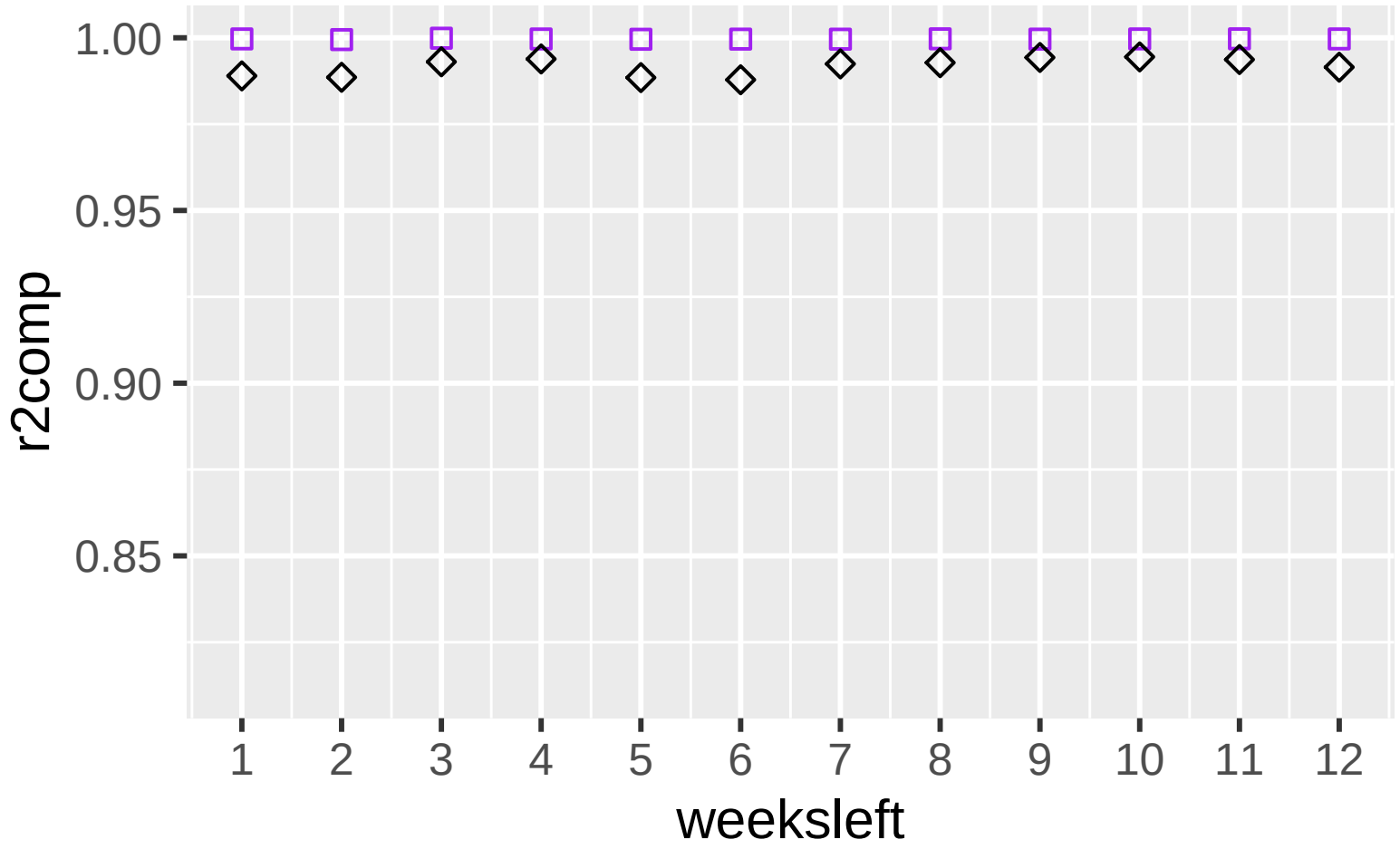
Predictiveness Attempt 1

- Below are estimated R^2 to predict the margin on election day given the polling margin for competitive and noncompetitive states. Purple Square is competitive states, and Black Rhomus is noncompetitive states.



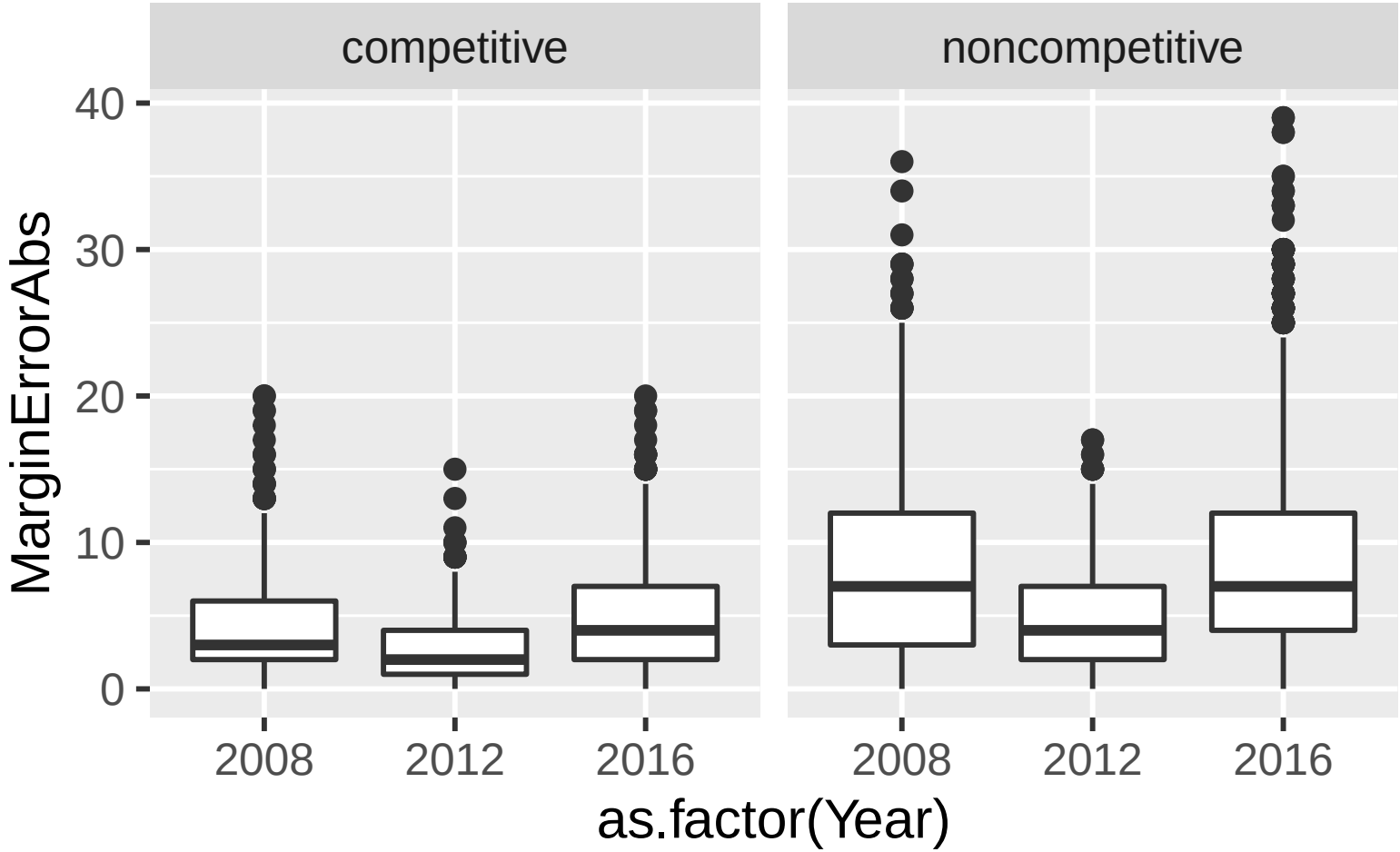
Predictiveness Attempt 2

- A mixed model including random effects for Year, State, and Year interacted with state is now fit. We plot the psuedo R^2 for the mixed model. Purple Square is competitive states, and Black Rhomus is noncompetitive states.



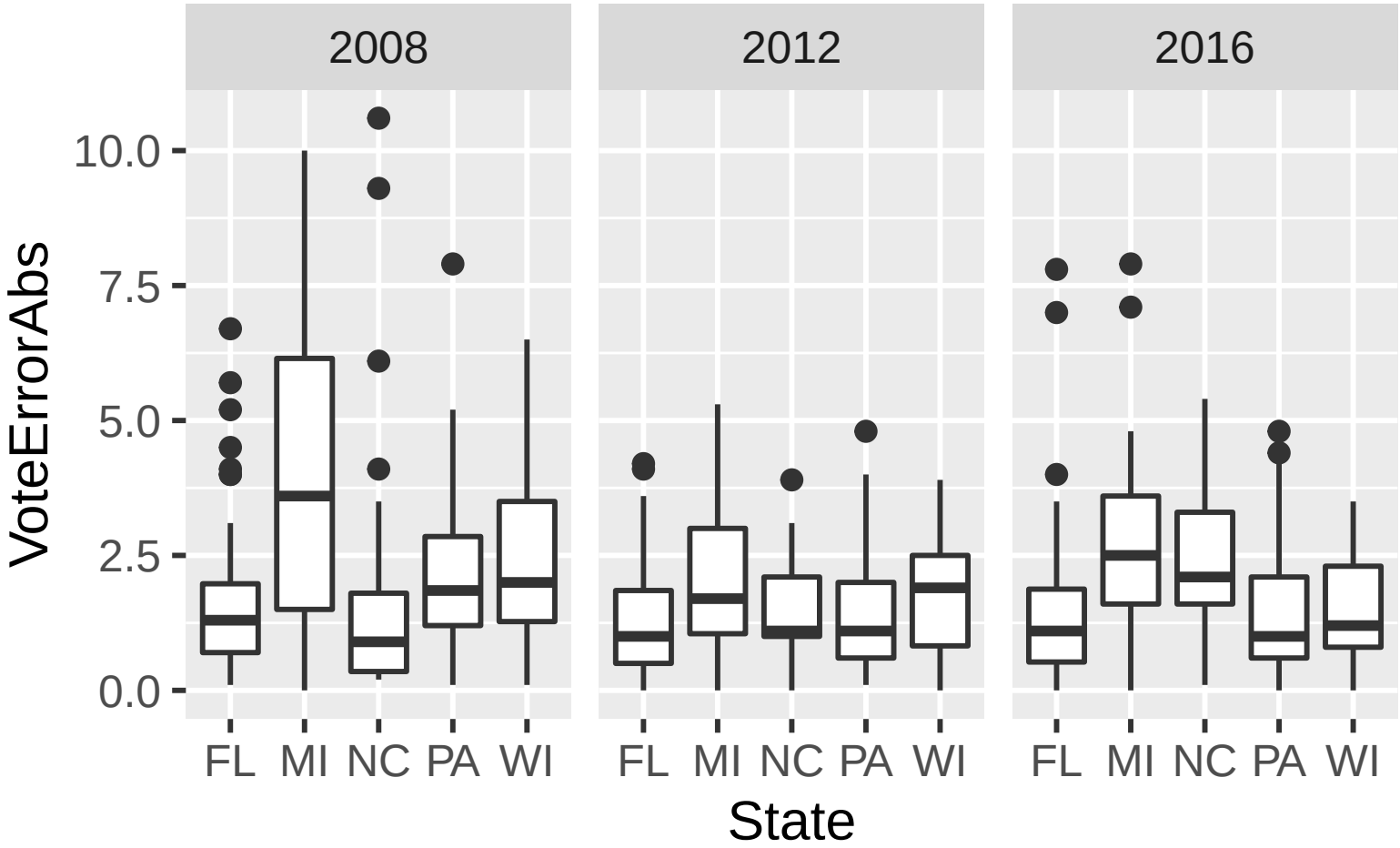
2016 Errors were Not Abnormally Large

- Polling Errors in 2008 were highly similar in absolute value to 2008. 2008 and 2012 broadly underestimated Democratic support, but in 2016 Republican support was underestimated.



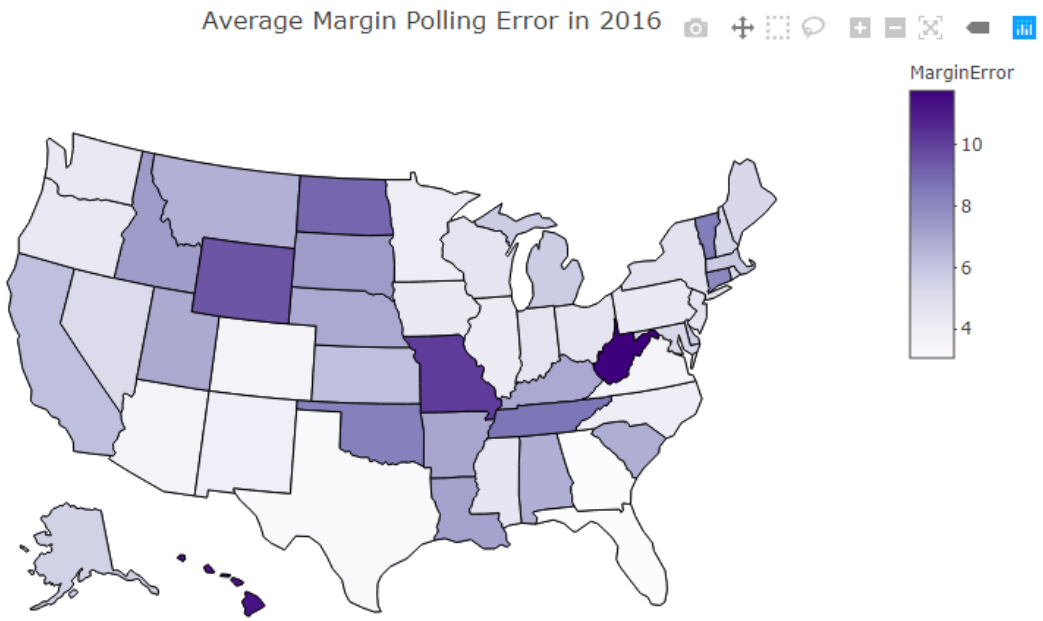
Polling Errors in FL, MI, NC, PA, WI

- Polling errors were not abnormally larger in 2016 in FL, MI, NC, PA, WI in terms of error, but the races were not called correctly.



Polling Errors in FL, MI, NC, PA, WI Part Two

Below is a plot of polling errors in 2016, and we see that FL, MI, NC, PA, WI are not unsual compared to other states.



Conclusion

Polls have small predictive value throughout the election process due to inconsistent patterns across states and years. Polling Accuracy is relatively stable starting three weeks out. A more controlled in depth analysis and model is needed to adjust for the differences in states, polling quality, and polling volume. However, it is clear that while polling error was larger in 2016 than 2012, it is similar to 2008. The focus on FL, MI, NC, PA, WI being wrong ignores that the direction of sampling error is random and can not be controlled. Overall polls were reliable in those states compared to previous years.

References

- Alexander, Brittany (2019), "A Bayesian Model for the Prediction of United States Presidential Elections," SIAM Undergraduate Research Online, **12**.
- Bon, J. J., Ballard, T., & Baffour, B. (2019), "Polling bias and undecided voter allocations: US presidential elections, 2004–2016," Journal of the Royal Statistical Society: Series A (Statistics in Society), **182(2)**, 467-493.
- Hillygus, D. S. (2011). "The evolution of election polling in the United States," *Public opinion quarterly*, **75(5)**, 962-981
- Shirani-Mehr, H., Rothschild, D., Goel, S., & Gelman, A. (2018), "Disentangling bias and variance in election polls," Journal of the American Statistical Association, **113(522)**, 607-614.