# STATS 15 Final - Polling Error Analysis in the 2020 Presidential Election

Winston Li, Ryan Chu, and Nathan Tjong

## Part 1: Introduction

### 1.1: Research Question

Every election cycle in America, the anticipation is palpable. There are the usual proceedings — debates, speeches, ads, and viral moments. But there is one element that has been a common fixture of elections — political polling.

It makes sense why. Before an election, people are fixated with "predicting the results" and figuring out who's ahead and who's behind. Polling is also incredibly important to judge the public's opinion on certain politicians and issues. And while on a surface level, taking random samples of the overall electorate should produce a representative sample of the results, there have been many instances where polls have been off.

In this paper, we will be looking at how close polls were in terms of predicting the outcome of the 2020 election. How accurate were polls overall in 2020? Were there differences between states? Were polls closer to the election more accurate than those farther away? Were there differences between pollsters? These are some of the questions we will be taking a look at in our exploratory graphics and analysis model.

That leads us to our primary research question for this paper: **in the 2020 Presidential election, how accurate were pollsters' polls?**

### 1.2: About the 2020 Election

The 2020 general Presidential election was held on November 3rd, 2020 [1]. The two major party candidates were former Vice President Joseph R. Biden from the Democratic Party and incumbent President Donald J. Trump from the Republican Party. There were some minor third party candidates, including those from the Libertarian and Green parties, but they comprised very small parts of the overall electorate.

In the United States, the election process occurs in two parts — the primary and general election. The primary election is when each political party (such as the Democratic and Republican party) choose among themselves through smaller elections and caucuses about who should be the nominee for their party.

President Donald J. Trump was considered the leader of the Republican Party at the time, so he was easily chosen as their 2020 presidential nominee. However, there were primary challengers, and data has shown that every modern President in American history who has been challenged by a candidate in the primary — even if that challenger didn't win — has ended up losing in the Electoral College [2]. Regardless, President Trump had strong support within the Republican Party, but there was a small group of Republicans who had become disillusioned with him.

In the Democratic primary, there was a contentious primary process with over 20 candidates. After numerous primaries, former Vice President Joseph R. Biden was chosen as the nominee. He was considered a moderate in the race, consolidating support from other center-left candidates before the race ended. However, he did not garner the support of many progressives, who chose to vote for Sen. Bernie Sanders or Sen. Elizabeth Warren instead. While a small minority of those voters ended up not voting or voting for third parties (such as the Green Party), the vast majority of progressives consolidated their support around Joe Biden. However, the main challenge for Joe Biden was to consolidate support from all across the political spectrum — from disaffected Republicans to progressive Democrats — after a bruising Democratic loss in 2016.

---

[1] https://ballotpedia.org/Presidential_election,_2020
[2] https://time.com/5682760/incumbent-presidents-primary-challenges/

Who wins the election in the United States is not determined by a national popular vote, but the Electoral College. In this system, each state is assigned a number of "electors", which is determined by adding the number of Senate seats and House of Representative seats a state has. Each state has two Senate seats, but the number of House seats varies based on population. For instance, in 2020, California had 2 Senators and 53 representatives, leading them to have 55 electoral votes. In Montana, a state with a much smaller population, they had 2 Senators and 1 representative, leading them to have 3 electoral votes.

**This is important because this changes the campaigning of each candidate, and will inform our analysis in the following parts.** There are a total of 538 electoral votes, meaning that to win the Presidency, a candidate generally must win just over half — or 270 electoral votes. Presidential candidates will end up focusing their campaigning on what are known as swing states — states that contain a significant number of electoral votes but are competitive and can go either way.

If a candidate wins a state, with the exception of Maine and Nebraska, they win ALL of their electoral votes. States like California and Montana are not important to candidates because they are so heavily Democratic and Republican respectively, but there are states like Nevada, Arizona, Pennsylvania, and Florida that are competitive and are where both presidential candidates AND pollsters focus their attention. Because of this, our analysis will be limited to polls from swing states.

The primary process for both parties occurred in the beginning of 2020, and by the summer of 2020, both parties officially nominated their candidates. Both the Democratic and Republican conventions were held in August 2020. We'll discuss our reasoning more later, but this is partially why we will be focusing our analysis on polls conducted after August 2020. From August to November, there were some high-profile incidents, such as President Trump contracting COVID-19 and the two presidential debates, but as we'll see later, those events had a much more muted effect compared to the events that occurred in 2019 and 2020 (such as the first impeachment, the start of the COVID-19 pandemic, the finalization of the presidential primaries, & the death of George Floyd and a reckoning over racial inequality).

Both candidates gathered support from different coalitions and people — leading them to acquire support from different demographics and areas of the country. Some of the major issues differentiating the candidates included their approach to COVID-19 (Biden supported mask mandates & restrictions, while Trump did not), the economy (Biden supported higher taxes on the wealthy, while Trump did not), environment (Trump rolled back environmental regulations, Biden would reimpose them), and health care (Trump attempted to repeal the Affordable Care Act, while Biden stated that he would expand it).

Trump's support primarily came from rural voters, the elderly, evangelicals, those without a college education, and male, white Americans, while Biden's support came from young, urban voters, racial and religious minorities, those with college educations and beyond, and many female Americans. Those categories are not perfect; for instance, compared to 2016, Trump gathered more support of Hispanic (especially Cuban) voters, and Biden gathered more religious (such as Catholic) support, but this provides a bit of a gauge of the national environment at the time.

In the end, Joe Biden won the election, gathering 306 electoral votes to Trump's 232. As classified by news pundits, the key competitive states & districts Biden won were in (sorted by region):

**West:** Nevada (6 votes, hold from 2016), Arizona (11 votes, flip from 2016)

**Midwest:** Minnesota (10 votes, hold), Wisconsin (10 votes, flip), Michigan (16 votes, flip), Pennsylvania (20 votes, flip), Nebraska's 2nd district (1 vote, flip from 2016)

**East:** New Hampshire (4 votes, hold), Virginia (13 votes, hold)

**South:** Georgia (16 votes, flip)

The competitive states & districts Trump won were in:

**Midwest:** Iowa (6 votes, hold), Ohio (18 votes, hold)

**East:** Maine's 2nd district (1 vote, hold)

**South:** North Carolina (15 votes, hold), Texas (38 votes, hold), Florida (29 votes, hold)

In the later parts of this report, we'll look into this data further to confirm our initial assumptions, and then answer our research questions.

## 1.3: About Political Polling

It is impossible to gauge the state of a race by asking every single voter. Instead, pollsters collect random samples of Americans — who are either registered to vote or likely to vote — to see who they are planning on voting for. If every pollster was able to perfectly sample each chosen voter, and if every voter who was contacted answered truthfully, polling would be extremely accurate. However, in practice, that's often not the case.

Many news organizations, polling firms, and even universities have their own polling firms, and they each use different methods for collecting samples. For instance, CNN and FOX News randomly call voters, CBS News and Politico have online polls available to anyone, and the Associated Press and Pew Research Center have online polls only available to people who have been recruited offline [3].

As one might imagine, no single method is perfect. Some people might not pick up their phone. Others might not spend enough time on the internet to encounter an online poll. This is where polling errors might arise; if certain groups are underrepresented in polling, that could skew the results one way or another. Generally, pollsters try to combat this by ensuring their sample is "representative" — such as making sure that the demographics of the sample match the demographics of the overall voting population, and that they target "high-propensity" voters (meaning voters who are likely to vote). However, those calculations can become skewed if there is a sudden event that changes the demographic makeup.

For instance, there was speculation in 2020 that, due to COVID-19, Democrats tended to stay home more than Republicans. Multiple post-election analyses noted that Democrats were overrepresented in phone polls, possibly indicating that Democrats were more able to pick up the phone and answer the poll. Again, while the pollster may have anticipated this to happen and adjusted their sample, this sudden change caught them by surprise and likely skewed the results beyond their internal corrections [4].

The second part of polling is questioning. Most pollsters ask the same set of questions — who are you planning on voting for President, how sure are you of that decision, which political party do you prefer to control Congress, etc. For our analysis, since we are focused directly on the 2020 Presidential election, we will only focus on the answer to that first question — who does each voter plan on voting for President. After each pollster collects data, they compile their results and publish them to news organizations, social media, or other platforms.

Most polls are commissioned by independent news organizations or universities — such as the Associated Press, CNN, and FOX News — whose main goal is to report accurately on the state of the race. But as one might imagine, polling is also a valuable tool for campaigns and partisan interests. They will often commission their own polls to gauge the state of the race. Sometimes, they are interested in the actual results. But often, those results may be unintentionally or intentionally skewed to be used for campaign fundraising or PR purposes. Other times, campaigns commission "push polls", where they are not interested in the actual result, but instead they word the questions in such a way to either 1) gather more information about a voter so they can keep sending them ads or 2) insinuate positive qualities about themselves or negative qualities about their opponent(s). We will be focused on polls that were officially published to the public, not internal campaign polling.

There are some instances where polling missed the mark, and some speculations by pollsters about why:

**1948 Presidential Election** [5]

The two major candidates running in the 1948 presidential election were incumbent Democrat Harry Truman and Republican Thomas Dewey. Polling in this election was conducted much the same way as it was in previous elections, so pollsters were confident about their predicted result — that Dewey would soundly defeat Truman.

Some newspapers were so confident that they announced the results a little bit too early. For example, the Chicago Daily Tribune, a newspaper whose editorial board had a strong hatred for Truman, started printing newspapers stating that "DEWEY DEFEATS TRUMAN" before the winner was actually announced on Election Night. The next morning, it was clear Truman won, but it was too late as the newspapers had already gone out — leading to a famous photo where Truman is holding up the Chicago Daily Tribune newspaper projecting his loss.

While this was not the first time polling had made a relatively severe error in predicting the outcome of the election, it was a major blow to the credibility of the polling industry as a whole. There was no single cause, but instead a multitude of factors including:

---

[3]https://www.pewresearch.org/fact-tank/2020/08/05/key-things-to-know-about-election-polling-in-the-united-states/
[4]https://www.politico.com/news/2021/04/13/dems-polling-failure-481044
[5]http://mathcenter.oxford.emory.edu/site/math117/historicalBlunders/

- Pollsters used a method called "quota sampling", where pollsters choose respondents from a given set of demographic proportions. The issue is that demographics can change rapidly from one election to the next, and so the polling industry now has changed to random sampling.

- Polling at the time stopped sometimes weeks from Election Day, making their predictions out of date.

- Polling incorrectly overpredicted the turnout of Republicans, leading to a self-fulfilling prophecy; many Republicans did not turn out because they thought the election was a "guaranteed" victory.

- There were also some smaller reasons speculated, such as that the questions were not worded fairly and some pollsters only targeted those with phone numbers — which, at the time, skewed toward higher income, likely Republican demographics.

**2000 Presidential Election**

While the election between Democrat Al Gore and Republican George Bush didn't necessarily suffer from a high level of polling error, it does represent an instance where polls can still miss the mark because of how close an election was.

This election was also one of the unique instances where a candidate (Al Gore) won the popular vote, yet still lost the Electoral College — showing why national popular vote polls are not that helpful when it comes to predicting the outcome of an election. The race ended up coming down to Florida, where George Bush won by 0.009%, which is well within most polls' margins of error. There were some notable errors when it came to projections though, mostly because news organizations were not provided accurate counts of how many votes remained in Florida. The race was called for Al Gore in some organizations, George Bush by others, then retracted by others, before the race was ultimately called for Bush after a recount and key Supreme Court decision. Again, while this instance doesn't highlight a specific error in polling, it does showcase how polls might not be able to do much to evade errors when it comes to elections with such small margins.

**2016 Presidential Election** [6]

The two major party candidates in 2016 were Democrat Hillary Clinton and Republican Donald Trump. Before election day, most polls predicted Clinton leading in key swing states; however, on Election Day, Trump ended up winning key races in the South (Florida, North Carolina) and Midwest (Michigan, Pennsylvania, and Wisconsin), solidifying his victory. And while the specific errors in polling were not exactly the same as previous elections, they still followed the two major themes — they didn't get responses from the right people, and they didn't correctly project who would turn out and who would stay home.

Here were some plausible explanations: [7]

- One of the top reasons why polls (in general) can be incorrect is that they didn't hear back from a representative share of the population. Some Trump supporters may have been "shy" and less likely to admit they supported Trump due to potential social consequences. Another factor might be their distrust of media institutions, often perpetuated by Trump himself. If people don't trust the media, they might not want to participate in a poll run by "CNN" and the "New York Times".

  - For instance, one pollster who predicted Trump would win — Trafalgar Group — ran some of their ads online on conservative news / opinion websites. A Trump voter may be more willing to answer a poll from an organization they trust. In addition, Trafalgar has a "higher" reputation from conservatives compared to other news sites such as CNN, NBC, or ABC.

- Another reason could be that undecideds simply broke for Trump in a way that pollsters could not expect. Clinton and Trump were two incredibly unpopular candidates, and there were two major October Surprises that could have swayed the results — the release of the Trump Access Hollywood tape and the re-opening of the investigation into Clinton's email server.

  - The last event was so close to Election Day that it didn't give pollsters a chance to catch up, and many voters likely made their decision by voting early / by mail before the FBI director announced they found no wrongdoing a day before the election.

---

[6]https://www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark/
[7]https://www.nytimes.com/2017/05/31/upshot/a-2016-review-why-key-state-polls-were-wrong-about-trump.html

- Turnout may have been lower from Democrats than expected. Similar to what occurred in 1948, Democrats may have felt that the election was a lock due to high probabilities forecast by pollsters and news organizations.

- One of the most important factors identified by pollsters was the education gap. While some pollsters differentiated based on education level (such as high school education, partial college graduate, college graduate, post-college grad, etc.), others did not. Exit polls from the 2016 election clearly showed that Trump gathered more support from non-college educated voters compared to Clinton, and that group may have been under surveyed by pollsters. **In the 2020 election, nearly all pollsters accounted for education in their polling.**'

Overall, it's clear that, while polling error can be unavoidable, there are often some common factors that can lead to polls that are significantly different from the actual results.

## 1.4 About the Dataset

We are using two data sets:

First, our main data set is a csv titled 'presidential_polls_2020.csv', which was published on FiveThirtyEight's 2020 election forecast GitHub repository. [8]

FiveThirtyEight is a reputable polling forecasting firm that aims to analyze elections, sports games, and other probabilistic events using polls and data. We are confident that the data present in this spreadsheet is correct, and we did some cross-referencing in Part 2 to verify that the polling results in the spreadsheet matched up with the actual data presented by pollsters.

FiveThirtyEight specifically had a 2020 Election Forecast that was based off of polling, experts' opinions, donations, and past election data. For our analysis, since we are only focusing on polling, we will only look at the presidential_polls_2020.csv.

There are numerous variables present in the spreadsheet. We will interpret each of them below, and mark the variables we are using with a star (*).

The csv focuses on polls that compare Biden to Trump — no third party candidates. There are two rows for each single poll — one for Biden's percentage and the other for Trump's. We'll clean this data up in Part 2.

- **candidate_name**\*: name of the candidate for each csv row (JOSEPH R. BIDEN or DONALD J. TRUMP)

- startdate: when answers started to get collected for this specific poll

- **enddate**\*: the last day answers were collected for this specific poll; for the sake of our analysis, we will be classifying each poll by their end date.

- **pollster**\*: name of the pollster conducting the poll

- **samplesize**\*: size of the sample collected for each poll

- **population**\*: a poll can either survey all adults, registered voters (RV), or likely voters (LV). since likely voter polls are considered the most accurate in terms of predicting actual election results, we will be limiting our analysis to LV polls.

- weight: used for FiveThirtyEight's forecast, not useful for our analysis

- influence: used for FiveThirtyEight's forecast, not useful for our analysis

- **pct**\*: the percent of the sample size candidate_name received in this specific poll

- house_adjusted_pct: the pct adjusted for FiveThirtyEight's forecast, not useful for our analysis

- trend_and_house_adjusted_pct: the pct adjusted for FiveThirtyEight's forecast, not useful for our analysis

- tracking: indicates whether a poll is a tracking poll - meaning they use the same sample to measure changes over time. we don't believe differentiating by tracking will make much of a difference in our analysis; as each pollster is able to choose whether they select a new sample for each new poll or not, it may provide perspective on whether that choice by that pollster was the right one in terms of reducing error.

---

[8]https://projects.fivethirtyeight.com/2020-general-data/presidential_polls_2020.csv

- **poll_id**\*: because each poll is split up into 2 rows — one for Biden and Trump's percentage — we need to combine each poll into one row using their common poll_id.

- question_id: each poll was composed of multiple questions, however, this spreadsheet only focuses on the question of who voters would choose if they voted for President at that current moment. because the question is the same for all of the polls in this spreadsheet, we will be disregarding this variable.

Our second data set is a table that includes the results of each U.S. Presidential Election, broken down by state, from 1976 to 2020. It was collected by the MIT Election Lab and published on Harvard's Dataverse, and is entitled `1976-2020-president.tab`. [9]

We are only using 5 variables from the dataset:

- **year:** We will filter for 2020.

- **state:** When we combine this data set with our polling data, we will combine them based on their state so we can compare each statewide poll vs. their actual results.

- **candidate:** We will filter for Biden and Trump.

- **candidatevotes:** The number of votes received by the particular candidate in the named state

- **totalvotes:** The total number of votes cast in the state for President. We will calculate a percentage of victory for Biden and Trump by dividing candidatevotes / totalvotes. We will calculate the percentage margin of victory in the race and compare that with our poll results to determine if there was error present.

- While there are other variables, such as state abbreviations and party, due to the narrow scope of our analysis, we won't need to use those variables.

As a whole, our goal is to compare some of the most prominent pollsters (based on the number of polls published) in some of the most competitive states (based on the margin of victory for either candidate) by their polling error. However, in terms of explaining the reasons behind the error, our speculations will be limited because of the lack of information provided by each pollster. Often, pollsters don't publish the full breakdown of their results — such as how they collected their data, and what are the demographics of their sample size — so we will have to infer based on information they have published. However, we feel like we can still make informative observations surrounding differences in pollsters in key states in the following sections. We will expand on this further in our analysis, conclusion, and areas for further study.

**Our analysis will be broken down into the following steps:**

- First, we will present an overview of our data, including averaged poll errors across all states.

  - We will discuss some general patterns across states and pollsters.

- We will limit our analysis to the FIVE most competitive swing states (by lowest margin of victory for either candidate). We will limit our polls to those conducted after August 2020 — when primaries have finished, and the political environment has (somewhat) settled.

- We will only be focused on the polls that compare the performance between Biden-Trump. While there were third party candidates in the race, they were a relatively insignificant part of the overall results.

  - We will explain our reasons why in our exploratory analysis.

- For each state, we will narrow our analysis to pollsters that published at least three polls in that state from August to November. While there certainly may have been pollsters that published fewer polls, we feel that three is the bare minimum needed to truly see if a pollsters' accuracy (or inaccuracy) to the actual results was due to random chance or something else.

- In each state, based on the top pollsters we have identified, we will run a t-test to determine a 95% confidence interval of the polling error (actual margin of victory in the state - average predicted margin of victory by a pollster).

---

[9]https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX

- We will also determine a t statistic that can help us determine — with 95% confidence — if a pollster's polls were systematically off compared to the actual results, or if they may have been due to random chance.

- Unfortunately, due to our limited samples — and the small number of polls by some pollsters — our answers will NOT be definitive, but we should be able to get a rough picture of the polling landscape from this analysis.

- In addition to state by state analyses, we will also an analysis of all five competitive states together (with pollsters that conducted at least 3 polls in each of those states) to see if any pollsters were consistently more or less accurate than the others.

- Then, we will perform methodology analyses of each pollster to determine any similarities or differences between the most accurate & inaccurate pollsters.

  - Again, many pollsters do not publish their full methodology / crosstabs / sampling methods — many have a business to run, after all. But there is some limited information online, which we will use to try to find potential differences between pollsters.

# Part 2: Loading the Data

When looking at the data from the actual results and polling results of the 2020 election we wanted to focus on one key thing - how much Biden/Trump was predicted to win by in a specific state? To help us find the answer to this question we filtered our data sets accordingly.

## 2.1: Loading and cleaning the data

We loaded the following packages:

- tidyverse - to make graphs, manipulate data, etc

- maps - to make map graphs

- lubridate - general date manipulation

- readr - to read in files

- broom - to tidy data into dataframes

In order to have something to compare the poll results to we imported a dataset of the 2020 election results. We then filtered it to only show the data we were interested in, which was the 2020 election Biden and Trump, the state, and the percentage of votes they each got. In order to determine of the accuracy of polls (will be covered later on) we added a column for the difference of percentage between Biden and Trump. We also included created a vector for the top 5 closest states by this new margin variable which will be used later in the analysis.

Below are the top 10 closest states in the election (ranked by lowest absolute margin of victory for either candidate):

```
raw_results <- read_delim("1976-2020-president.tab",
    delim = "\t", escape_double = FALSE,
    trim_ws = TRUE)
```

```
## Rows: 4287 Columns: 15
## -- Column specification ---------------------------------------------------
## Delimiter: "\t"
## chr (6): state, state_po, office, candidate, party_detailed, party_simplified
## dbl (7): year, state_fips, state_cen, state_ic, candidatevotes, totalvotes, ...
## lgl (2): writein, notes
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
filtered_results <-
  raw_results %>%
  filter(year == 2020, candidate %in% c("TRUMP, DONALD J.", "BIDEN, JOSEPH R. JR")) %>%
  mutate(pct = candidatevotes/totalvotes) %>%
  select(state, candidate, pct) %>%
  pivot_wider(names_from = candidate, values_from = pct) %>%
  mutate(actual_margin = (`BIDEN, JOSEPH R. JR` - `TRUMP, DONALD J.`) * 100) %>%
  select(state, actual_margin)

final_states <- filtered_results %>%
  arrange(abs(actual_margin)) %>%
  select(state) %>%
  head(5) %>%
  pull() %>%
  tolower()

filtered_results %>%
  arrange(abs(actual_margin)) %>%
  head(10)
```

```
## # A tibble: 10 x 2
##    state           actual_margin
##    <chr>                   <dbl>
##  1 GEORGIA                 0.236
##  2 ARIZONA                 0.309
##  3 WISCONSIN               0.627
##  4 PENNSYLVANIA            1.16
##  5 NORTH CAROLINA         -1.35
##  6 NEVADA                  2.39
##  7 MICHIGAN                2.78
##  8 FLORIDA                -3.36
##  9 TEXAS                  -5.58
## 10 MINNESOTA               7.11
```

We will include graphs for the top ten closest states noted above, but we will perform our hypothesis test in Part 4 on the five closest states in our analysis — Georgia, Arizona, Wisconsin, Pennsylvania, and North Carolina. We believe this provides a relatively good sample of swing states in the 2020 election as they cover a wide range of regions in the United States.

South: *Georgia*, *North Carolina*, Florida

Midwest (traditionally Democratic): *Wisconsin*, *Pennsylvania*, Minnesota, Michigan

West (traditionally Republican): *Arizona*, Texas, Nevada

Here we filter the poll data to once again only show us the relevant data. We first import the data from a csv file, then select the state, date, pollster, percentage of votes a candidate received, the poll id, and the candidate date. From this new dataset we create a dataframe for each candidate to contain everything we just selected. Afterwords we create a data frame from the previous 2 by using full_join, which also lets us create a poll_margin variable by once again taking Biden's percentage - Trump's percentage.

```r
final <- read_csv("presidential_polls_2020.csv")
```

```
## Rows: 7786 Columns: 17
## -- Column specification -----------------------------------------------------
## Delimiter: ","
## chr (7): state, modeldate, candidate_name, startdate, enddate, pollster, pop...
## dbl (9): cycle, samplesize, weight, influence, pct, house_adjusted_pct, tren...
## lgl (1): tracking
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

filtered_final <- final %>%
  filter(population == "lv") %>%
  mutate(dates = as.Date(enddate, "%m/%d/%Y")) %>%
  select(state, enddate, pollster, pct, poll_id, candidate_name, dates, samplesize)

Trump <-
  filtered_final %>%
  filter(candidate_name %in% "Donald Trump")
Biden <-
  filtered_final %>%
  filter(candidate_name %in% "Joseph R. Biden Jr.")
compare_both <-
  Biden %>%
  full_join(Trump, by = c("state", "enddate", "pollster", "poll_id", "samplesize")) %>%
  rename(Biden_pct = pct.x, Trump_pct = pct.y) %>%
  mutate(date = as.Date(enddate, "%m/%d/%Y")) %>%
  select(state, pollster, Biden_pct, Trump_pct, date, poll_id, samplesize) %>%
  mutate(poll_margin = Biden_pct - Trump_pct)
compare_both
```

```
## # A tibble: 3,273 x 8
##    state   pollster         Biden~1 Trump~2 date       poll_id sampl~3 poll_~4
##    <chr>   <chr>              <dbl>   <dbl> <date>        <dbl>   <dbl>   <dbl>
##  1 Wyoming SurveyMonkey          33      66 2020-11-01    72663     330     -33
##  2 Wyoming SurveyMonkey          34      65 2020-10-31    72525     361     -31
##  3 Wyoming SurveyMonkey          33      65 2020-10-30    72424     378     -32
##  4 Wyoming SurveyMonkey          33      66 2020-10-29    72345     394     -33
##  5 Wyoming University of Wyo~    31      59 2020-10-28    72334     614     -28
##  6 Wyoming SurveyMonkey          31      68 2020-10-28    72242     739     -37
##  7 Wyoming SurveyMonkey          32      66 2020-10-27    72088     713     -34
##  8 Wyoming SurveyMonkey          31      67 2020-10-25    71978     654     -36
##  9 Wyoming SurveyMonkey          31      67 2020-10-24    71893     649     -36
## 10 Wyoming SurveyMonkey          31      67 2020-10-23    71826     638     -36
## # ... with 3,263 more rows, and abbreviated variable names 1: Biden_pct,
## #   2: Trump_pct, 3: samplesize, 4: poll_margin
```

## 2.2 Additional Cleaning Considerations

Some checks we ran to make sure the numbers seemed correct:

- We checked for unique polls and we found no duplicates. The latest polls were also conducted right before Election Day (Nov 1st, 2020), which makes sense.

```
compare_both %>%
  unique()
```

```
## # A tibble: 3,273 x 8
##    state   pollster         Biden~1 Trump~2 date       poll_id sampl~3 poll_~4
##    <chr>   <chr>              <dbl>   <dbl> <date>        <dbl>   <dbl>   <dbl>
##  1 Wyoming SurveyMonkey          33      66 2020-11-01    72663     330     -33
##  2 Wyoming SurveyMonkey          34      65 2020-10-31    72525     361     -31
##  3 Wyoming SurveyMonkey          33      65 2020-10-30    72424     378     -32
```

```
##  4 Wyoming SurveyMonkey          33     66 2020-10-29  72345    394     -33
##  5 Wyoming University of Wyo~     31     59 2020-10-28  72334    614     -28
##  6 Wyoming SurveyMonkey          31     68 2020-10-28  72242    739     -37
##  7 Wyoming SurveyMonkey          32     66 2020-10-27  72088    713     -34
##  8 Wyoming SurveyMonkey          31     67 2020-10-25  71978    654     -36
##  9 Wyoming SurveyMonkey          31     67 2020-10-24  71893    649     -36
## 10 Wyoming SurveyMonkey          31     67 2020-10-23  71826    638     -36
## # ... with 3,263 more rows, and abbreviated variable names 1: Biden_pct,
## #   2: Trump_pct, 3: samplesize, 4: poll_margin
```

After running the unique() function, we ended up with the same number of rows (3,273) as the actual data frame, which gives us confidence that these polls are all unique.

- We randomly checked polls and, based on the limited information found online, we found that the polls matched up with the results listed.

```
compare_both %>%
  filter(poll_id == 72576)
```

```
## # A tibble: 1 x 8
##   state     pollster Biden_pct Trump_pct date       poll_id samplesize poll_ma~1
##   <chr>     <chr>        <dbl>     <dbl> <date>        <dbl>      <dbl>     <dbl>
## 1 Wisconsin Civiqs          51        47 2020-11-01    72576        789         4
## # ... with abbreviated variable name 1: poll_margin
```

For instance, we were able to verify that this specific poll, conducted by Civiqs in Wisconsin, matched up with the results published on their official Twitter page. [10]

# Part 3: Graphical Analysis of the Data

What does the data look like as a whole?

First, we will take a look at our data set of polls. We will calculate average margins for polls by state, pollster, and other variables to identify any trends.

If we look at the state margin of victory of all 50 states, we can see they fall across a wide range. As we noted in our introduction, only a few states fall within a close enough margin that campaigns choose to campaign within them — known as the swing states.
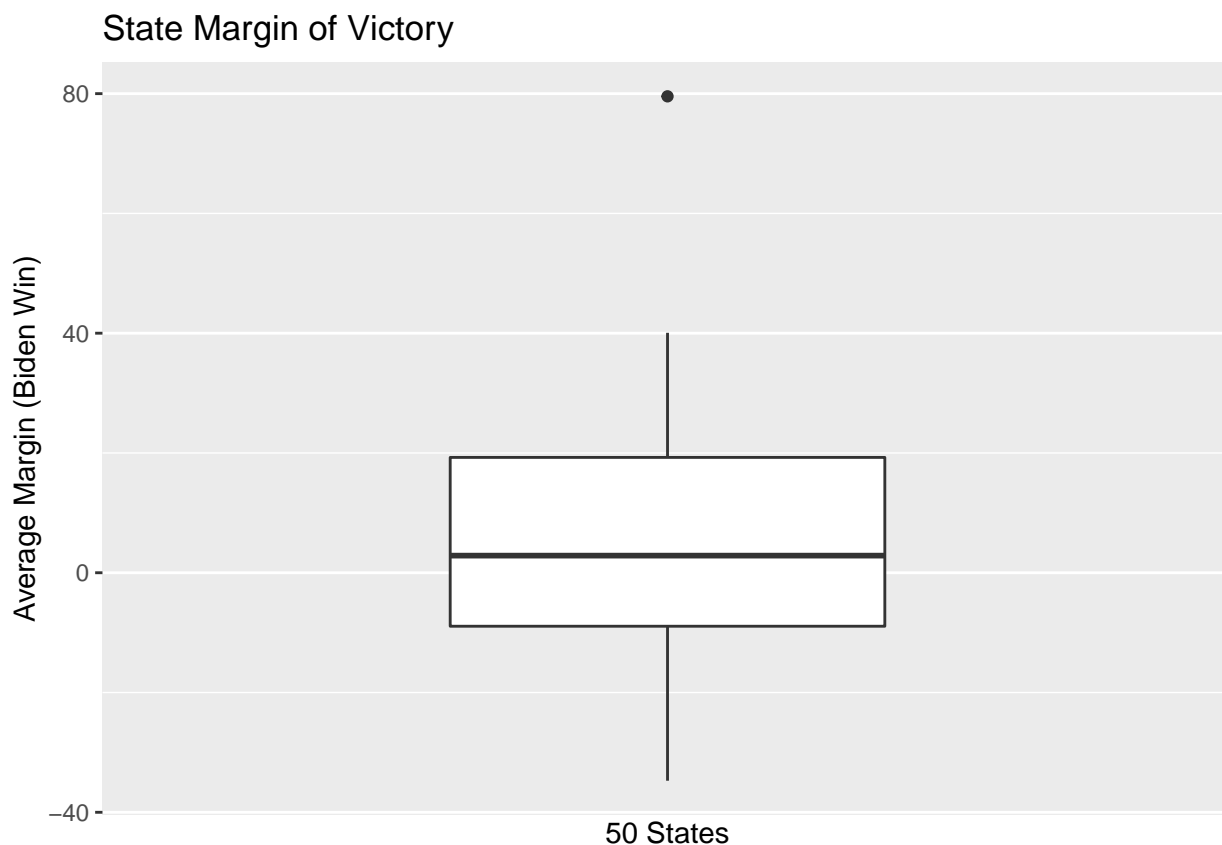
```
compare_both %>%
  group_by(state) %>%
  mutate(avg_margin = mean(poll_margin)) %>%
  mutate(avg_margin = abs(avg_margin)) %>%
  select(state, avg_margin) %>%
  unique() %>%
  arrange(avg_margin)
```

```
## # A tibble: 56 x 2
## # Groups:   state [56]
##   state        avg_margin
##   <chr>             <dbl>
## 1 Georgia           0.318
## 2 Iowa              0.798
## 3 NE-1              2
## 4 ME-2              2.03
```

---

[10]https://twitter.com/civiqs/status/1323071678331539456?lang=ca

```
##  5 North Carolina      2.12
##  6 Ohio                2.13
##  7 Florida             2.40
##  8 Texas               2.43
##  9 Arizona             3.32
## 10 Nevada              3.47
## # ... with 46 more rows
```

```
compare_both %>%
  group_by(state) %>%
  mutate(avg_margin = mean(poll_margin)) %>%
  select(state, avg_margin) %>%
  unique() %>%
  ggplot(aes(x = ,y = avg_margin)) +
  geom_boxplot() +
  scale_x_discrete() +
  labs(title = "State Margin of Victory", x = "50 States", y = "Average Margin (Biden Win)")
```

## State Margin of Victory



Here we can see the general margin of victory falls between -10% (so Trump's victory) and 20%. The outlier is the District of Columbia which is historically very left leaning.

Now let's take a deeper dive into the predicted margins. We'll plot a boxplot that shows the average predicted percentage win for Biden.

```
compare_both %>%
  group_by(state) %>%
  mutate(avg_margin = mean(poll_margin)) %>%
  select(state, avg_margin) %>%
  unique() %>%
  arrange(avg_margin) %>%
#graphing begins
```

```
ggplot(aes(y = avg_margin)) +
geom_boxplot() +
scale_x_discrete() +
labs(title = "Average Predicted % Win for Biden", x = "50 States", y = "Average Margin (Biden Win)")
```

## Average Predicted % Win for Biden
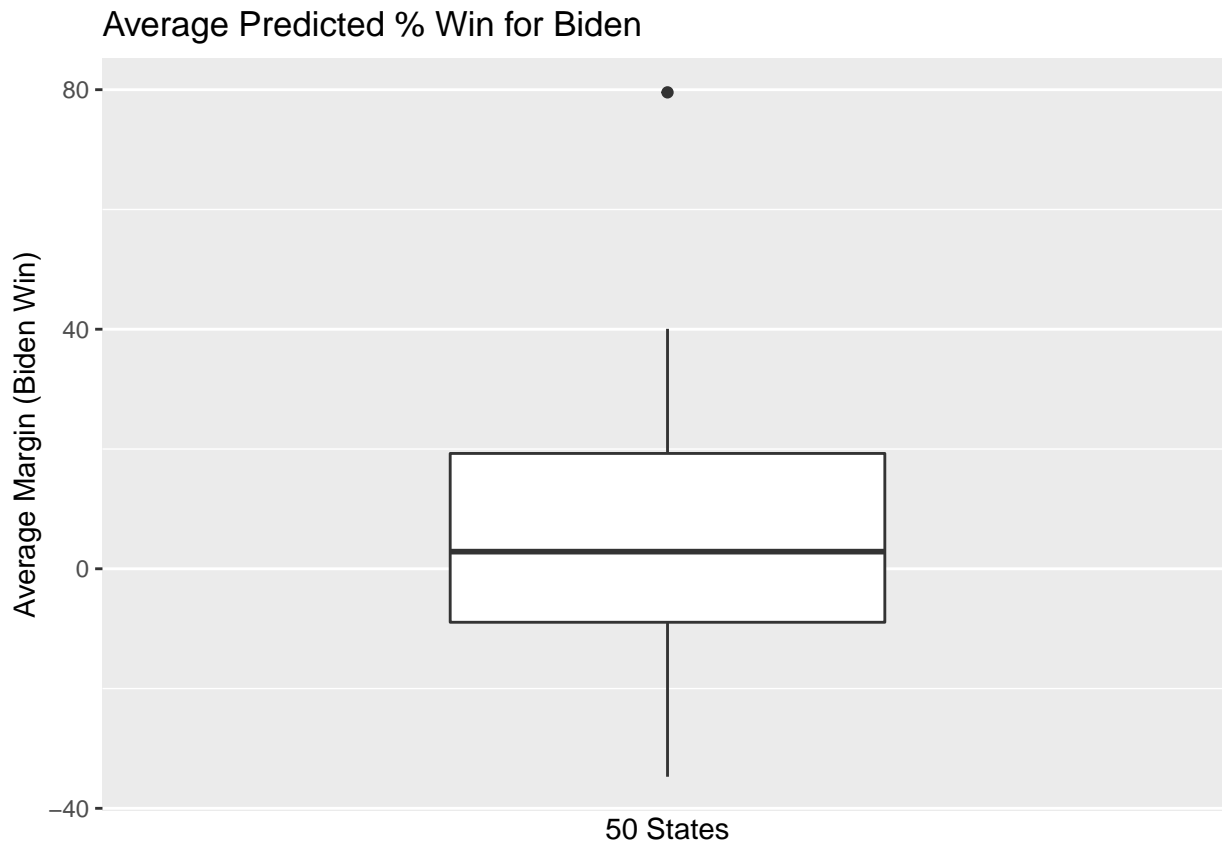


We can see that most of these polls fall somewhere within the actual result. This demonstrates just how important polling errors can be in swing states, where the margins are extremely close.

Now, let's break it down by pollster. If we just dive deep into one particular swing state, Arizona, and look at each pollsters predicted margin of victory, we notice an interesting trend.

```
top_pollsters  <-
compare_both %>%
  group_by(pollster) %>%
  filter(state == "Arizona") %>%
  summarize(count = n()) %>%
  filter(count > 1) %>%
  select(pollster) %>%
  pull()

#avg_margin vs pollsters bar graph
compare_both %>%
  filter(pollster %in% top_pollsters) %>%
  group_by(pollster) %>%
  mutate(avg_margin = mean(poll_margin)) %>%
  select(pollster, avg_margin) %>%
  unique() %>%
  arrange(avg_margin) %>%
#graphing begins
  ggplot(aes(x = reorder(pollster, +avg_margin), y = avg_margin, fill = pollster)) +
  geom_bar(stat = "identity") +
```

```
theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1)) +
labs(title = "Avg Predicted % Win for Biden: AZ", x = "Pollster", y = "Average Margin (Biden Win)")
```

## Avg Predicted % Win for Biden: AZ
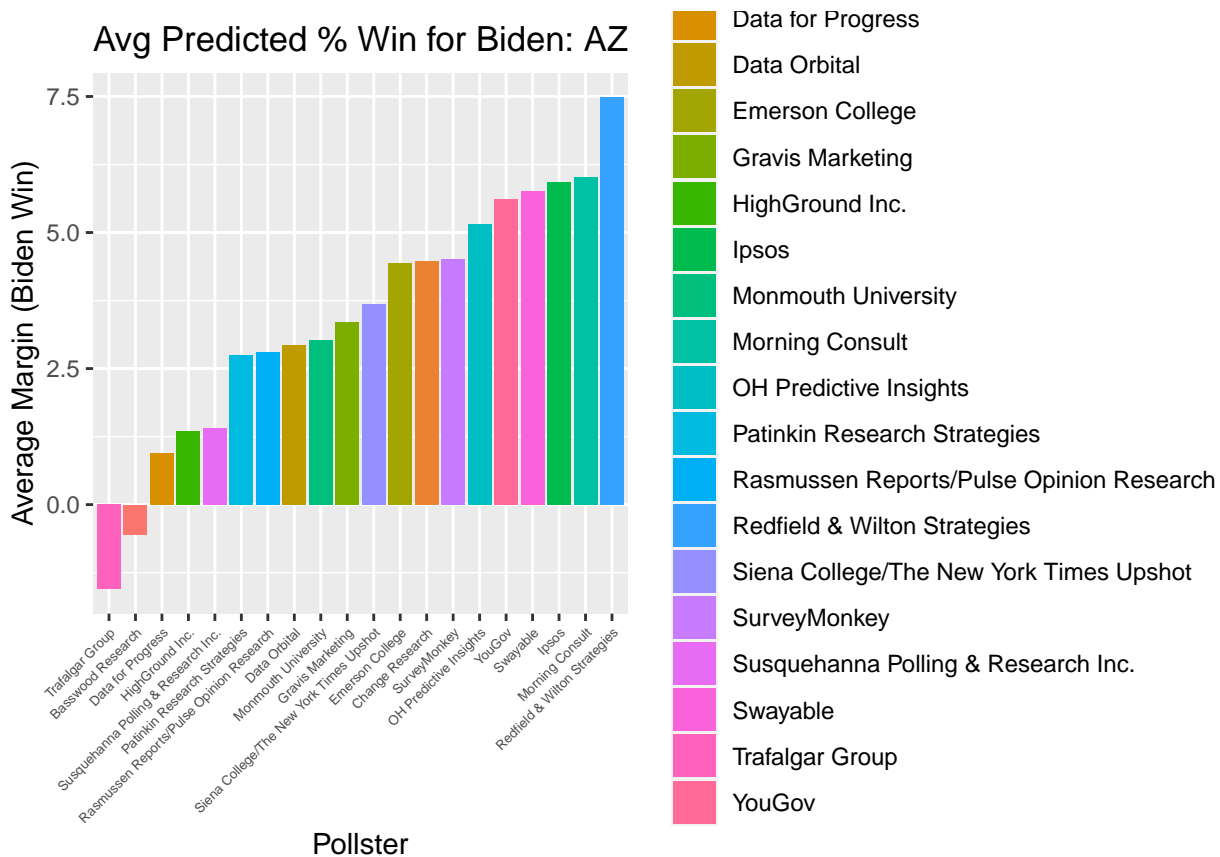


There is significant variation among pollsters in terms of their predicted margins for Biden in Arizona. Some variation is expected, due to the random nature of polling. But it seems interesting that while Biden only barely ended up winning Arizona, polls in Arizona had so much variation, with some predicting a narrow Trump victory to others predicting a large Biden lead. This may be worth investigating.

If we look at the actual election data, we can examine the states that were most heavily won for Biden and won for Trump. These again are evidence of non-swing states, as they lean so heavily one way or the other that the other campaign does not bother spending time in them.

```
filtered_results %>%
  arrange(desc(actual_margin)) %>%
  head(10) %>%
  ggplot(aes(x = reorder(state, actual_margin), y = actual_margin, fill = state)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1)) +
  labs(title = "Highest Actual Win for Biden", x = "State", y = "Average Margin (Biden Win)")
```

# Highest Actual Win for Biden
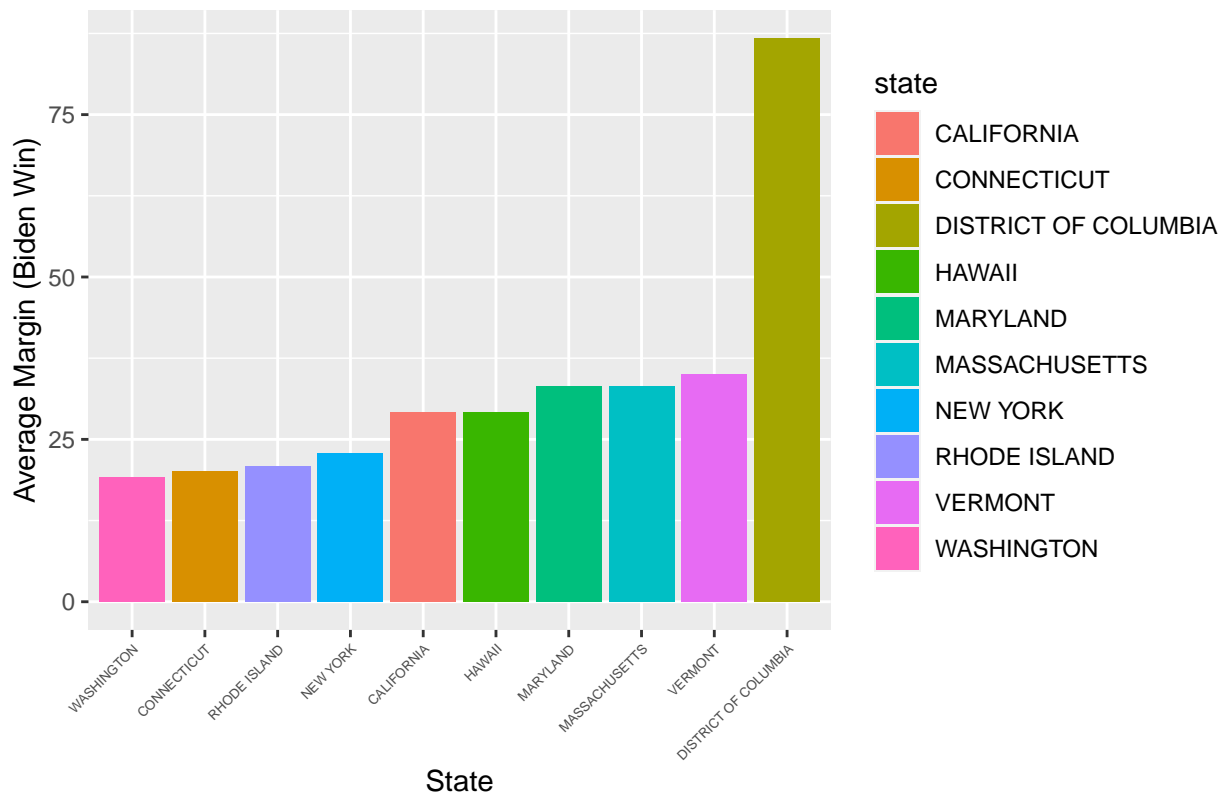


```
filtered_results %>%
  mutate(actual_margin = -actual_margin) %>%
  arrange(desc(actual_margin)) %>%
  head(10) %>%
  ggplot(aes(x = reorder(state, actual_margin), y = actual_margin, fill = state)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1)) +
  labs(title = "Highest Actual Win for Trump", x = "State", y = "Average Margin (Biden Win)")
```

## Highest Actual Win for Trump
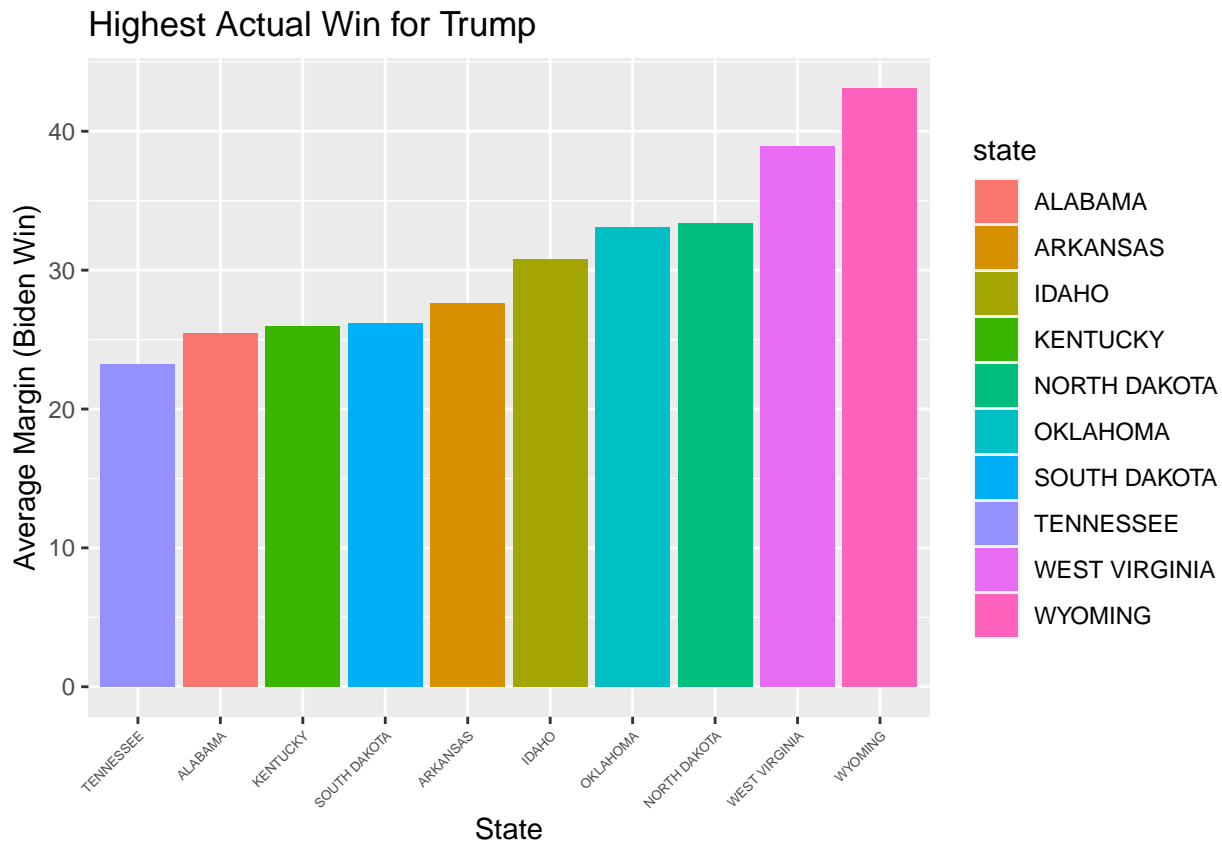


For Biden, his strongholds are in the Northeast and West particularly, and Trump's are in the Midwest and South. The one interesting outlier is the District of Columbia, because it is extremely Democratic-leaning and has a much higher percentage win for Biden than the other states. However, overall, most of these "safe states" fall within a margin of victory of around 18% and more for either candidate.

But now let's take a look at the polling data for these "safe states:":

```r
state_results <-
  compare_both %>%
  group_by(state) %>%
  mutate(avg_margin = mean(poll_margin)) %>%
  select(state, avg_margin) %>%
  unique() %>%
  arrange(avg_margin)

#top 10 for Biden Victory
Biden10 <- tail(state_results, 10)
Biden10 %>%
  ggplot(aes(x = reorder(state, avg_margin), y = avg_margin, fill = state)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1)) +
  labs(title = "Highest Predicted Win for Biden", x = "State", y = "Average Margin (Biden Win)")
```

# Highest Predicted Win for Biden
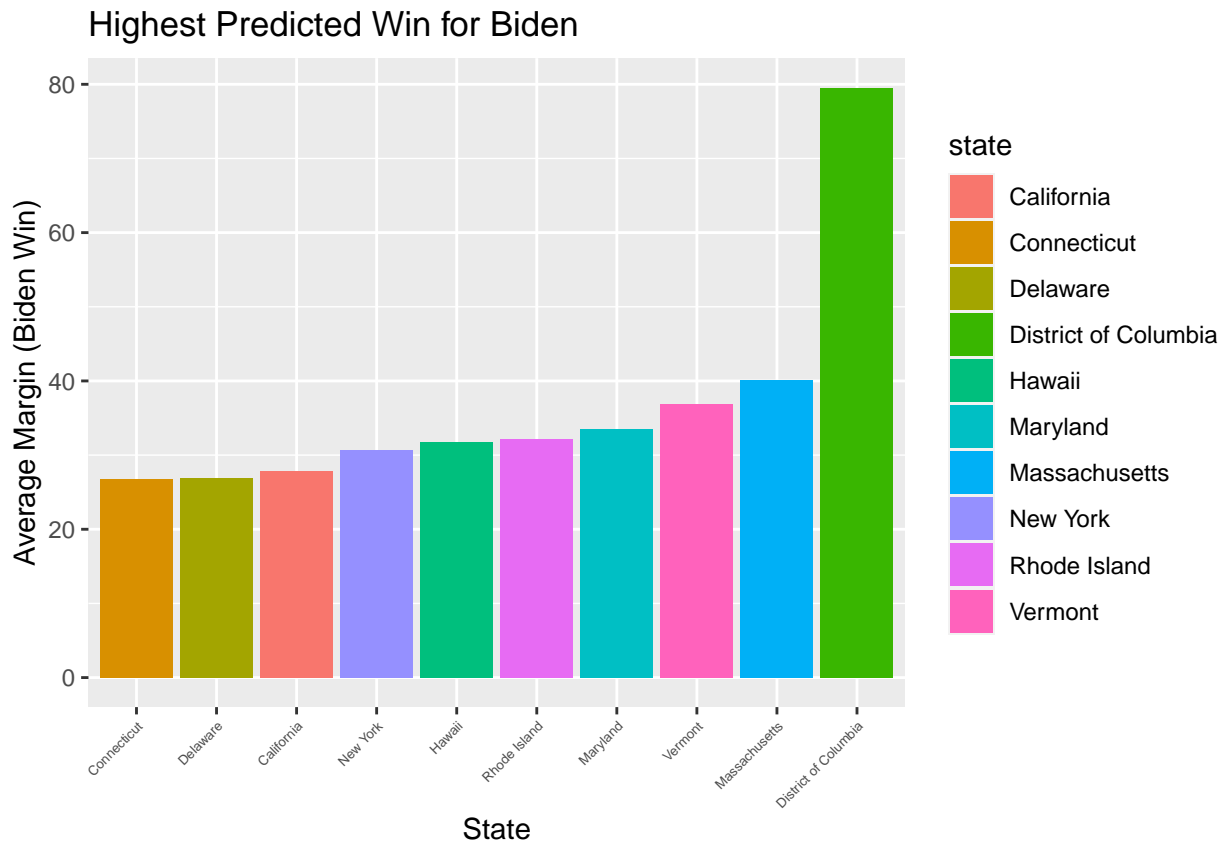


```r
#top5 for Trump Victory
Trump10 <- head(state_results, 10)
Trump10 %>%
  mutate(avg_margin = -avg_margin) %>%
  ggplot(aes(x = reorder(state, avg_margin), y = avg_margin, fill = state)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1)) +
  labs(title = "Highest Predicted Win for Trump", x = "State", y = "Average Margin (Biden Win)")
```

## Highest Predicted Win for Trump
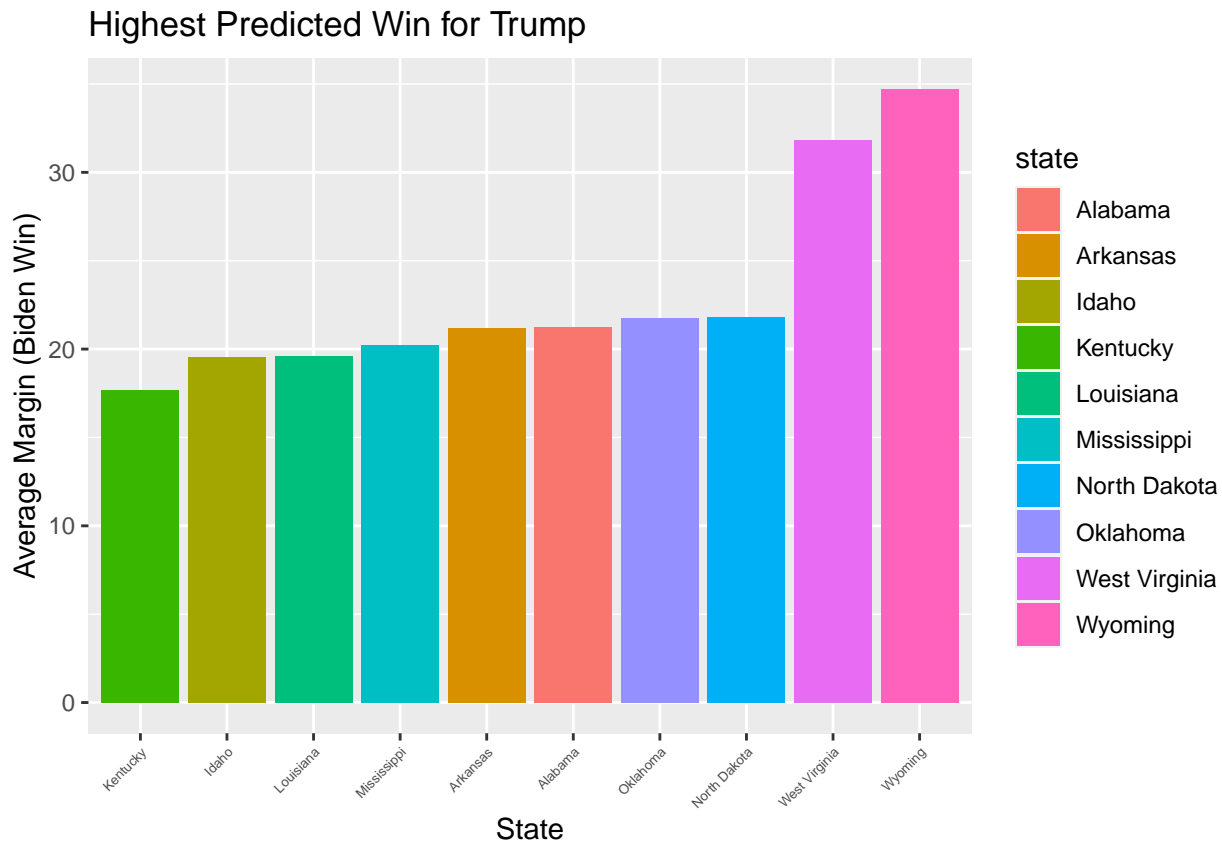


Overall, it looks pretty close, meaning that at such high margins, even if there is polling error, it wouldn't be enough to sway the election. Some of the predicted percentages for Biden also tend to be higher than the actual margins of victory. For example we can see that Biden was predicted to win Massachusetts by 40%, however the actual percentage they won by was around 33%. But, when it comes to these safe states, there are often less polls so our insights will be less reliable.

Another relationship we wanted to explore was how the poll results changed as we got closer to the election.

If we take a look at the time the poll was conducted, we can see that polls only really start populating heavily near May 2020 and beyond, likely because that is when the primary season for both parties is wrapping up and the general election campaigning period is about to start. We decided to focus our data on analyzing polls after August 1st, 2020, because we believe that's when most pollsters have started conducting polls heavily (increased concentration of dots) and the political environment has at least somewhat settled.

```
compare_both %>%
  filter(state == "Wisconsin") %>%
 ggplot(aes(x = date, y = poll_margin)) +
  geom_point(color = "blue") +
  geom_smooth(color = "black", method = "lm") +
  scale_x_date(date_breaks = "1 month", limit = c(as.Date("2020-01-01"), as.Date("2020-11-2"))) +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Poll Margin over Time in WI", y = "Average Margin (Biden Win)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

## Poll Margin over Time in WI
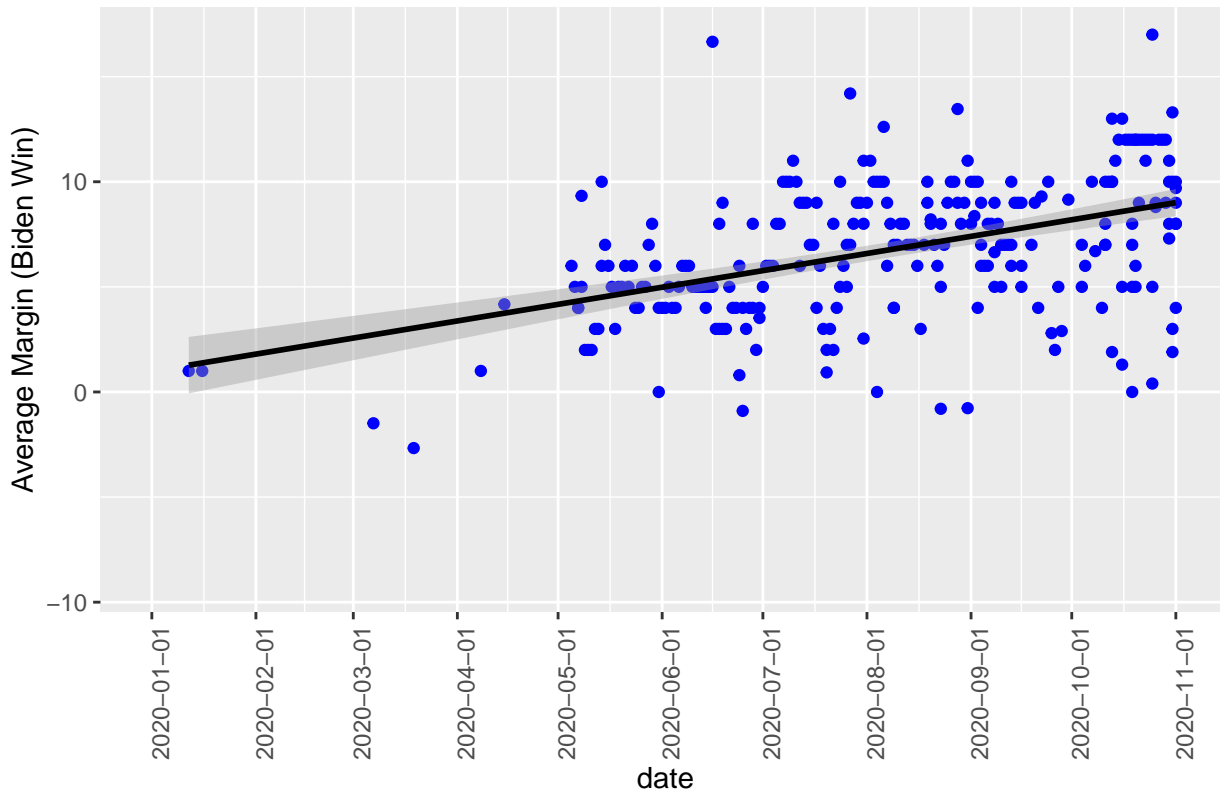


While Biden did gain some support from August 1st to November 1st, we would not be able to limit our analysis further because we need more polls to be able to make informed analyses. For instance, let's say we limited our analysis to polls only conducted in the last month.

```
compare_both %>%
  filter(state == "wisconsin", date > '2020-10-01') %>%
  group_by(pollster) %>%
  summarize(count = n()) %>%
  arrange(desc(count))
```

```
## # A tibble: 0 x 2
## # ... with 2 variables: pollster <chr>, count <int>
```

One of the issues if we limit our polling to polls conducted extremely close to the election is that we simply do not have enough polls to make an informed analysis. While some pollsters conducted many polls — notably SurveyMonkey — others did not in the last month, and so widening the date period to August would allow more polls from other pollsters to be included in our analysis. We might be able to identify short patterns with 2 or 3 polls, but having more would definitely add more weight to our analysis.

To further explore this relationship we wanted to check if samplesize had any effect on this. (some polls didn't have a sample size so drop_na() is used here, for reference there are 39) We decided that a small sample would be 500 or less responses, medium samples would be 1000 or less responses, and large samples would be anything over that.

```
compare_both %>%
  mutate(sample_cat = case_when(samplesize < 500 ~ "Small", samplesize < 1000 ~ "Medium", samplesize > 100
  drop_na() %>%
  ggplot(aes(x = date, y = poll_margin, color = sample_cat)) +
  geom_point() +
  geom_smooth(method = "lm") +
  scale_x_date(date_breaks = "1 month", limit = c(as.Date("2020-08-01"), as.Date("2020-11-2"))) +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Poll Margin over Time (Zoomed out)", y = "Average Margin (Biden Win)")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

```
## Warning: Removed 660 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 660 rows containing missing values (geom_point).
```
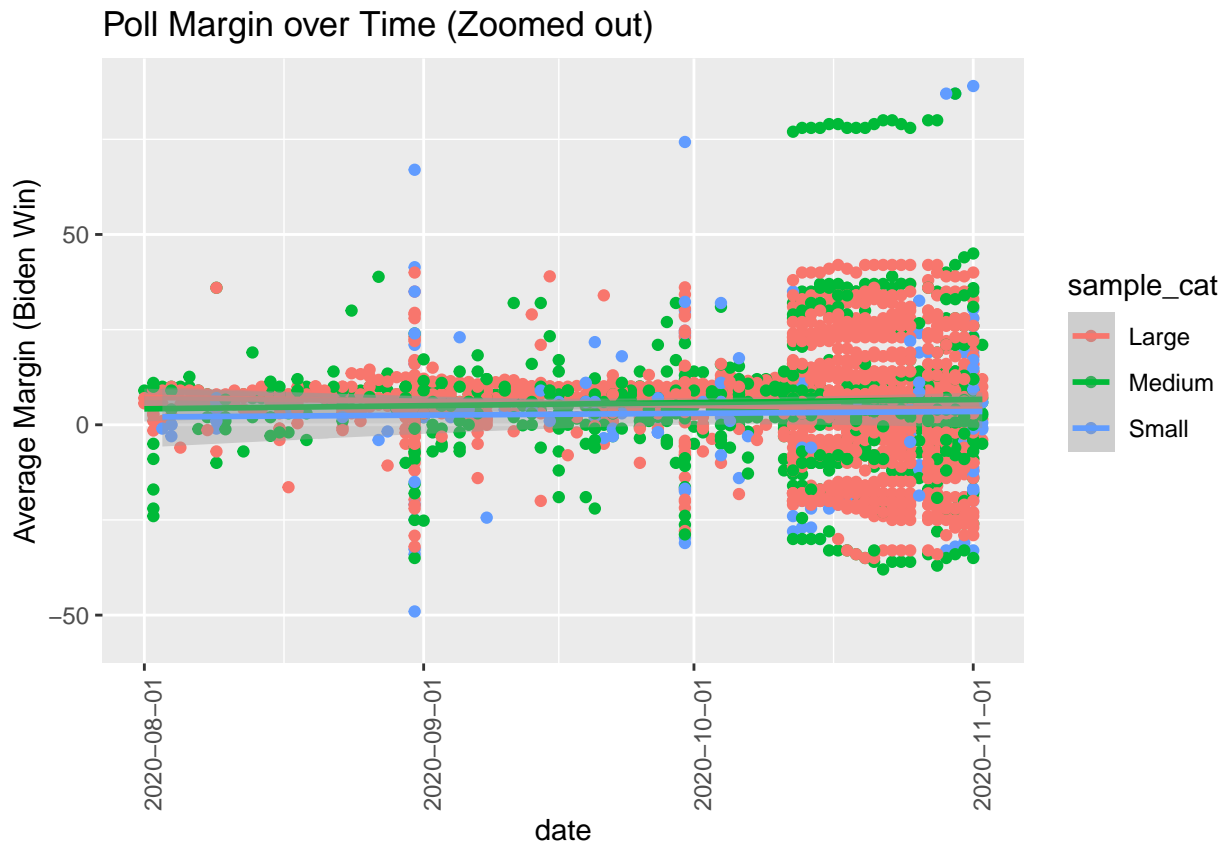


```
compare_both %>%
  mutate(sample_cat = case_when(samplesize < 500 ~ "Small", samplesize < 1000 ~ "Medium", samplesize > 1000
  drop_na() %>%
  ggplot(aes(x = date, y = poll_margin, color = sample_cat)) +
  geom_point() +
  geom_smooth(method = "lm") +
  scale_x_date(date_breaks = "1 month", limit = c(as.Date("2020-08-01"), as.Date("2020-11-2"))) +
  theme(axis.text.x = element_text(angle = 90)) +
  coord_cartesian(ylim=c(-1,10)) +
  labs(title = "Poll Margin over Time (Zoomed in)", y = "Average Margin (Biden Win)")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

```
## Warning: Removed 660 rows containing non-finite values (stat_smooth).
## Removed 660 rows containing missing values (geom_point).
```

## Poll Margin over Time (Zoomed in)



We can see that the best trends between sample size greatly vary. Overall, sample size doesn't appear to have a large effect on the margins delivered by polls, explaining why we will likely not be using sample size in our final pollster error analysis. But in our zoomed out graph, we noticed that there was a huge influx of polls between mid October and November. Why?

It turns out they are SurveyMonkey polls, which started to get HEAVILY conducted across the country during that time. In fact, they started conducting polls, including in safe states nearly EVERY SINGLE DAY before the election, which explains why some of their polls had such high average margins of victory for Biden — because a lot of other pollsters didn't bother to poll safe states.

```
compare_both %>%
  filter(pollster == "SurveyMonkey") %>%
  arrange(desc(date)) %>%
  filter(poll_id > 70400) %>%
  filter(state != "national")
```

```
## # A tibble: 1,261 x 8
##    state         pollster      Biden~1 Trump~2 date       poll_id sampl~3 poll_~4
##    <chr>         <chr>           <dbl>   <dbl> <date>       <dbl>   <dbl>   <dbl>
##  1 Wyoming       SurveyMonkey       33      66 2020-11-01   72663     330     -33
##  2 Wisconsin     SurveyMonkey       54      44 2020-11-01   72665    2475      10
##  3 West Virginia SurveyMonkey       32      67 2020-11-01   72664     738     -35
##  4 Washington    SurveyMonkey       62      36 2020-11-01   72666    3666      26
##  5 Virginia      SurveyMonkey       57      41 2020-11-01   72668    4145      16
##  6 Vermont       SurveyMonkey       71      26 2020-11-01   72667     719      45
##  7 Utah          SurveyMonkey       44      54 2020-11-01   72669    1435     -10
##  8 Texas         SurveyMonkey       47      51 2020-11-01   72670    8250      -4
##  9 Tennessee     SurveyMonkey       45      54 2020-11-01   72671    2943      -9
## 10 South Dakota  SurveyMonkey       36      62 2020-11-01   72672     556     -26
## # ... with 1,251 more rows, and abbreviated variable names 1: Biden_pct,
## #   2: Trump_pct, 3: samplesize, 4: poll_margin
```

If we look at the sample size of states in an example swing state, such as Michigan:

```
compare_both %>%
  mutate(sample_cat = case_when(samplesize < 500 ~ "Small", samplesize < 1000 ~ "Medium", samplesize > 100
  filter(state == "Michigan") %>%
  drop_na() %>%
  ggplot(aes(x = sample_cat, y = poll_margin, color = sample_cat)) +
  geom_boxplot(outlier.size = 0.75) +
  labs(title = "Sample Size vs Poll Margins", x = "Sample Size", y = "Average Margin (Biden Win)")
```
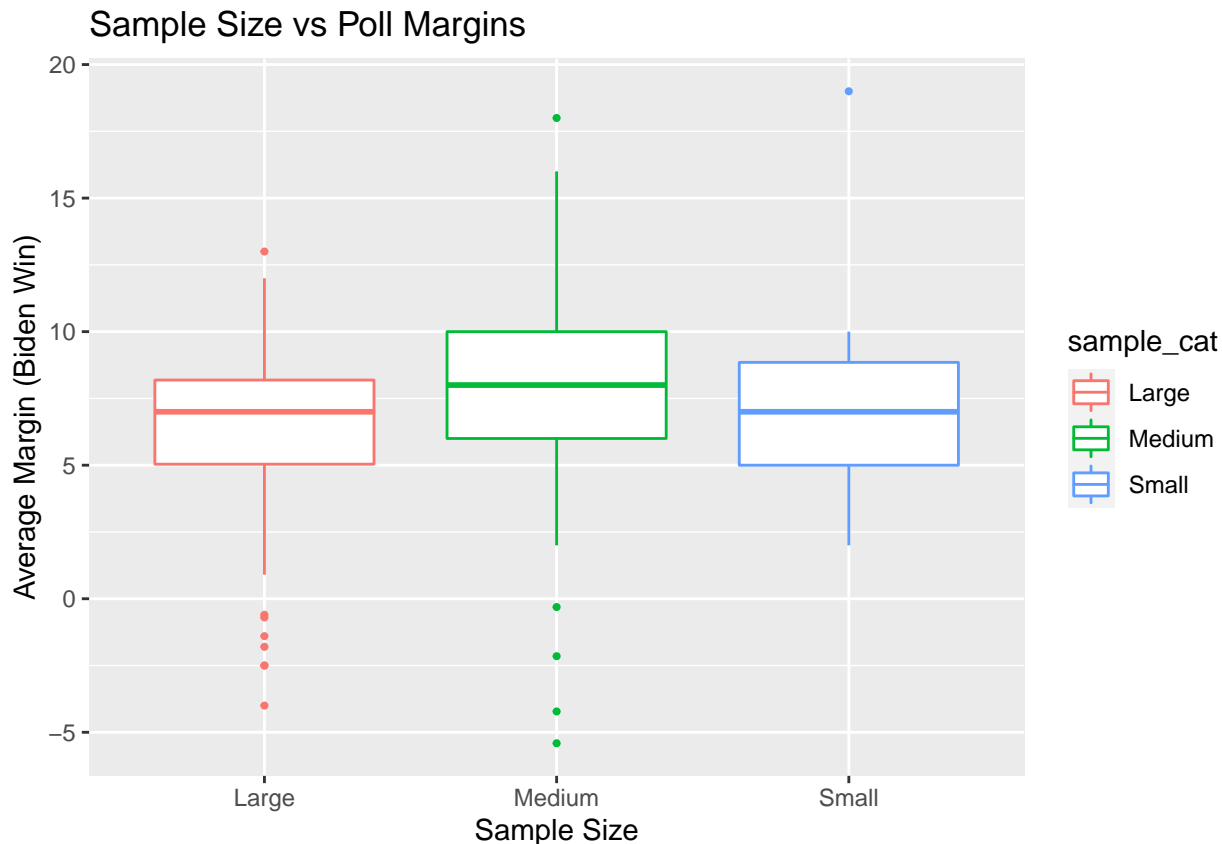


As we can we see from the boxplot, the mean poll margins for samples that are large, medium, and small are extremely close to one another, and so again we don't believe that there are much differences between polls of varying sample sizes. Interestingly, the polls with small samples have significantly more variation; we would expect that newer or more experimental pollsters might have smaller samples they can collect, which may explain this extra variance. Regardless, since our main goal is to compare pollster to pollster, and the differences in predicted margins of victory between large, medium, and small sample polls doesn't seem great, we will not be including it as a factor in our analysis.

Next, we looked at polling error. We are calculating polling error by subtracting the average predicted margin of victory by the actual margin of victory of the state.

```
states_map <- map_data("state")
  biden_aggregate<-raw_results%>%
    filter(year==2020, candidate %in% c("BIDEN, JOSEPH R. JR"))%>%
    mutate(pct=candidatevotes/totalvotes*100)%>%
    select(state, pct)
  biden_aggregate$state<-tolower(biden_aggregate$state)


compare_both$state <- tolower(compare_both$state)
 filtered_results$state <- tolower(filtered_results$state)
```

```
 map_natError <- compare_both %>%
   group_by(state) %>%
   summarize(avg_margin = mean(poll_margin)) %>%
   inner_join(filtered_results, by = "state") %>%
   mutate(error = avg_margin - actual_margin)


ggplot(map_natError, aes(map_id = state))+
  geom_map(aes(fill=error), map = states_map, color = "black") +
  expand_limits(x = states_map$long, y = states_map$lat) +
  coord_equal() +
  labs(x= "", y="", fill="Margin Error", title="Error of all Pollsters by State", caption="*Positive value
  scale_x_discrete() +
  scale_y_discrete()
```

## Error of all Pollsters by State
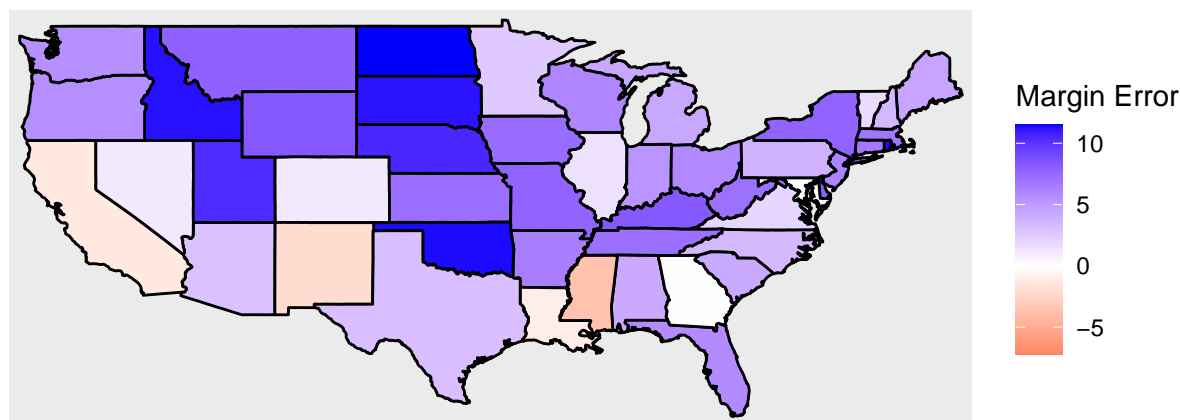


*Positive values: overestimated Biden, negative values: overestimated Trump.

Darker blue states represent a higher margin of error for Biden winning, white states represent results closest to the actual results, and states that are more orange represent higher margin of error for Trump winning. States like Oklahoma, Montana, Dakota, and Idaho had very high margins of error (in favor of Biden).

Now, let's break down this polling error map by pollster, starting with SurveyMonkey:

```
compare_both$state <- tolower(compare_both$state)
filtered_results$state <- tolower(filtered_results$state)

map_SM <- compare_both %>%
  group_by(state) %>%
  filter(pollster == "SurveyMonkey") %>%
  filter(date > '2020-08-01') %>%
  summarize(avg_margin = mean(poll_margin)) %>%
  inner_join(filtered_results, by = "state") %>%
  mutate(error = avg_margin - actual_margin)
```

```
ggplot(map_SM, aes(map_id = state))+
geom_map(aes(fill=error), map = states_map, color = "black") +
expand_limits(x = states_map$long, y = states_map$lat) +
coord_equal() +labs(x= "", y="", fill="Margin Error", title=" Error of SurveyMonkey Polls after August 1s
scale_x_discrete() +
scale_y_discrete()
```

## Error of SurveyMonkey Polls after August 1st by State
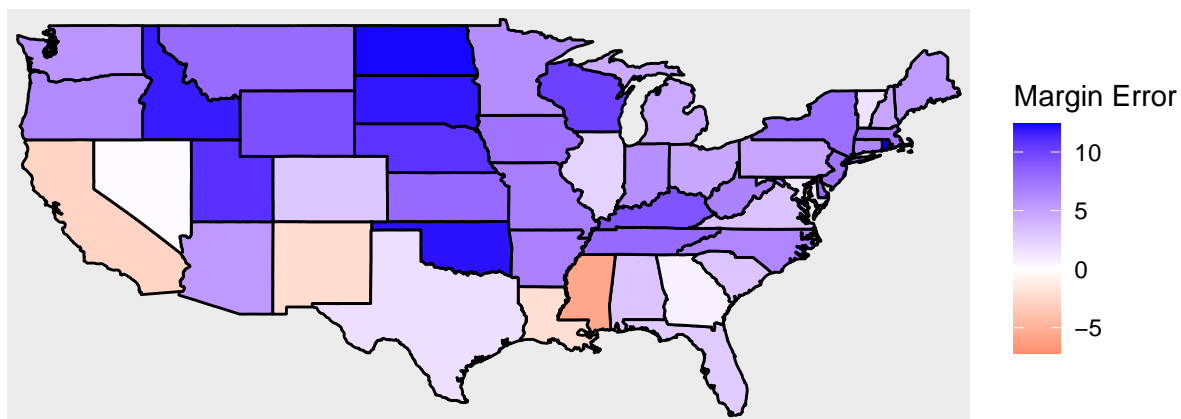


*Positive values: overestimated Biden, negative values: overestimated Trump.

As we can see, SurveyMonkey's polls seem to have error towards Biden in most states, with exception for some like Georgia and Nevada. What we noted earlier is true – the entire map is filled in meaning SurveyMonkey surveyed every single state in the country, including swing states.

For another pollster such as Mornign Consult:

```
compare_both$state <- tolower(compare_both$state)
filtered_results$state <- tolower(filtered_results$state)

map_MC <- compare_both %>%
  group_by(state) %>%
  filter(pollster == "Morning Consult") %>%
  filter(date > '2020-08-01') %>%
  summarize(avg_margin = mean(poll_margin)) %>%
  inner_join(filtered_results, by = "state") %>%
  mutate(error = avg_margin - actual_margin)


ggplot(map_MC, aes(map_id = state))+
geom_map(aes(fill=error), map = states_map, color = "black") +
expand_limits(x = states_map$long, y = states_map$lat) +
coord_equal() +labs(x= "", y="", fill="Margin Error", title=" Error of Morning Consult Polls after August
scale_x_discrete() +
```

```
scale_y_discrete()
```

## Error of Morning Consult Polls after August 1st by State
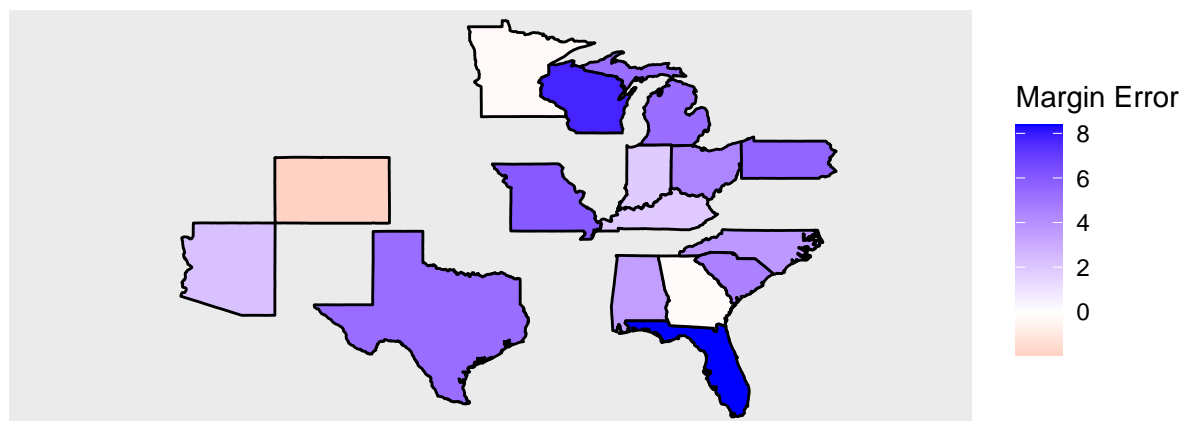


*Positive values: overestimated Biden, negative values: overestimated Trump.

One noticeable difference between Morning Consult is the predicted margin of error between Florida and the actual margin of error. This is very important because Florida is worth a lot of electoral college points, and Morning Consult predicted Florida to win by a large margin. This would make it very hard for people to actually use their polls for an accurate prediction for the presidential election results. In addition, they clearly did not survey every state — instead focusing on close ones — which helps explain why the map is not fully filled in.

Overall, some of their margins of error seem higher than SurveyMonkey in some states and lower in others.

Now if we look at Redfield & Wilton Strategies:

```
compare_both$state <- tolower(compare_both$state)
filtered_results$state <- tolower(filtered_results$state)

map_MC <- compare_both %>%
  group_by(state) %>%
  filter(pollster == "Redfield & Wilton Strategies") %>%
  filter(date > '2020-08-01') %>%
  summarize(avg_margin = mean(poll_margin)) %>%
  inner_join(filtered_results, by = "state") %>%
  mutate(error = avg_margin - actual_margin)


  ggplot(map_MC, aes(map_id = state))+
  geom_map(aes(fill=error), map = states_map, color = "black") +
  expand_limits(x = states_map$long, y = states_map$lat) +
  coord_equal() +labs(x= "", y="", fill="Margin Error", title=" Error of Redfield & Wilton Strategies afte
  scale_x_discrete() +
  scale_y_discrete()
```

# Error of Redfield & Wilton Strategies after August 1st by State
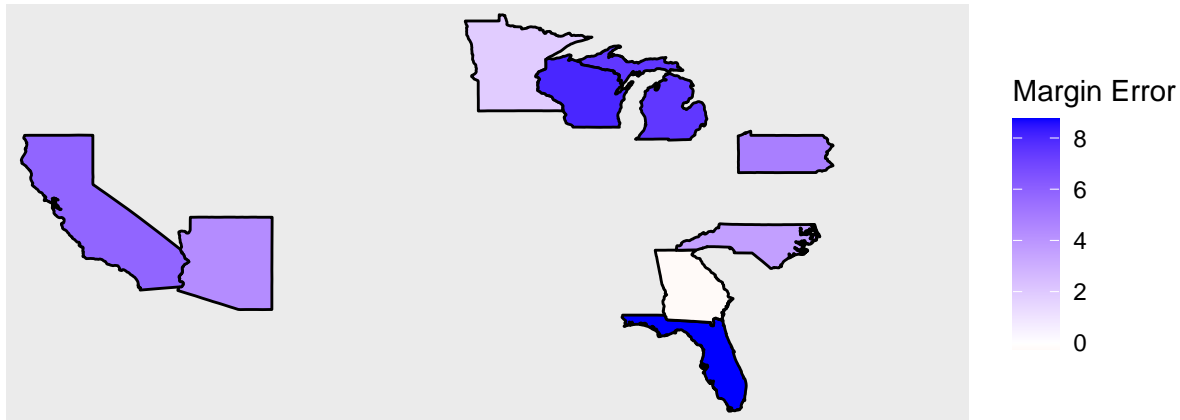


*Positive values: overestimated Biden, negative values: overestimated Trump.

They surveyed even less states, and again they seemed to have varyign errors across states. Now while overall these pollsters all appeared to have some error for Biden, we don't know numerically from these graphs how much that error is. This is what we'll be analyzing in the next few parts.

Last, looking at Change Research polls:

```
compare_both$state <- tolower(compare_both$state)
filtered_results$state <- tolower(filtered_results$state)

map_CR <- compare_both %>%
  group_by(state) %>%
  filter(pollster == "Change Research") %>%
  filter(date > '2020-08-01') %>%
  summarize(avg_margin = mean(poll_margin)) %>%
  inner_join(filtered_results, by = "state") %>%
  mutate(error = avg_margin - actual_margin)


ggplot(map_CR, aes(map_id = state))+
geom_map(aes(fill=error), map = states_map, color = "black") +
expand_limits(x = states_map$long, y = states_map$lat) +
coord_equal() +labs(x= "", y="", fill="Margin Error", title=" Error of Change Research Polls after August
scale_x_discrete() +
scale_y_discrete()
```

Error of Change Research Polls after August 1st by State
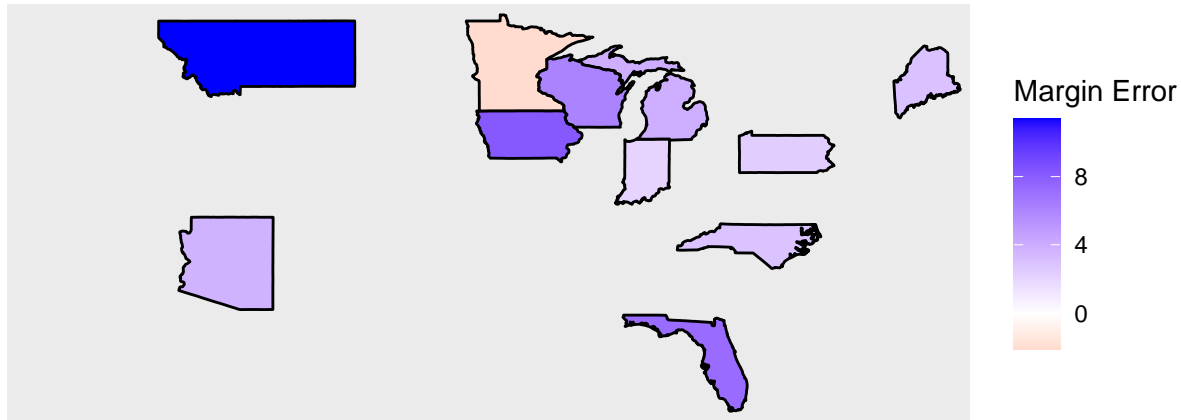


*Positive values: overestimated Biden, negative values: overestimated Trump.

The highest margin of error Change Research has is for the state of Montana. However, this is not as concerning becuse Montana is only worth 3 electoral votes, so even if it is predicted wrong it's not significant. In addition, it appears to have less error compared to the other polls, but still some error towards Biden (lighter blue).

For Ipsos:

```
compare_both$state <- tolower(compare_both$state)
filtered_results$state <- tolower(filtered_results$state)

map_CR <- compare_both %>%
  group_by(state) %>%
  filter(pollster == "Ipsos") %>%
  filter(date > '2020-08-01') %>%
  summarize(avg_margin = mean(poll_margin)) %>%
  inner_join(filtered_results, by = "state") %>%
  mutate(error = avg_margin - actual_margin)


  ggplot(map_CR, aes(map_id = state))+
  geom_map(aes(fill=error), map = states_map, color = "black") +
  expand_limits(x = states_map$long, y = states_map$lat) +
  coord_equal() +labs(x= "", y="", fill="Margin Error", title=" Error of Ipsos Polls after August 1st by S
  scale_x_discrete() +
  scale_y_discrete()
```

## Error of Ipsos Polls after August 1st by State
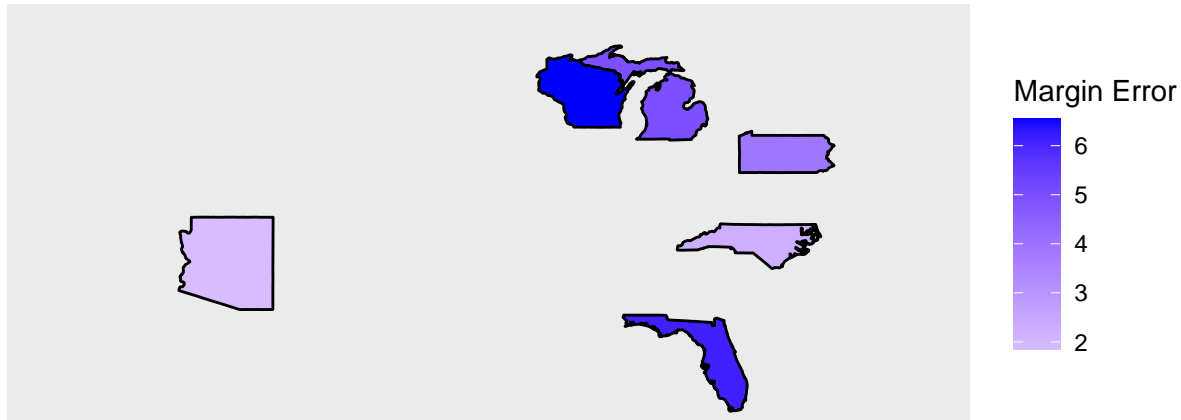


*Positive values: overestimated Biden, negative values: overestimated Trump.

What do we see? We see that first, there are certain swing states where the margins are extremely close and campaigns AND pollsters focus most of their attention.

Second, there appears to be clear differences in terms of how the top pollsters forecast the election. While many had an error towards Biden, there were still differences in terms of how large that error was. We did not notice many trends in terms of other variables, such as polling sample size or date (after August 1st, 2020).

The vast differences we saw between pollsters in their predictions is what informed our research question and future analysis: how large EXACTLY were these differences in polling error among pollsters?

## Part 4: Exploring and Modeling the Data

For our t-test analysis, as noted earlier, we will focus on the top 5 states with the closest margins of victory for either Biden or Trump: Georgia, Arizona, Wisconsin, Pennsylvania, and North Carolina — stored in `final_states`. Our primary goals are to 1) identify differences in polling error between pollsters in the most competitive states NUMERICALLY, 2) potentially discover whether we can say those errors are statistically significant or if we do not have enough evidence to support that claim, and 3) identify errors among pollsters nationwide.

First, in each state, we will look at all of the polls conducted from August 1st, 2020 to Election Day (November 3rd, 2020). Next, we will filter for pollsters that had at least 3 polls recorded in the state; we believe this is a good balance between 1) allowing as many viable pollsters as possible to be part of our analysis and 2) ensuring that any potential accuracy / inaccuracy from pollsters can be attributed to greater patterns beyond random chance.

Our data is currently composed of the predicted margins of victory for EACH poll and the actual margins of victory in each state. Each state will be composed of multiple t-tests — one for each pollster. The inputted data will be the error calculated for EACH poll (predicted margin of victory - actual margin of victory). Our null hypothesis for each t-test will be that the error is 0, while our alternate hypothesis is that the error is not equal to 0. We will be testing with 95% confidence.

We will be looking for 1) a 95% confidence interval of the actual error by the pollster in the state, 2) the actual mean error calculated based on each polls' error, and 3) the t-statistic predicted based on the data, which can either reject the null hypothesis (saying that we are sufficiently confident the pollsters' polls had some degree of error)

or fail to reject the null hypothesis (saying that we do not have enough evidence to say the pollsters' polls had some degree of error). **As a reminder, we'll be calculating the margin of victory through Biden's % - Trump's %, so a positive value would mean that Biden won the state, and a negative value means that Trump won.**

It's important to note that, if we do end up failing to reject the null hypothesis, it does **NOT** mean that the pollster had no error. The absence (or lack) of sufficient evidence does not indicate the accuracy of each pollster. Instead, the result of rejecting / not rejecting the null hypothesis can hone down if there are pollsters whose results are consistently off of the actual results.

On a broader level, we must note that our results will not be definitive or conclusive because of the low number of polls by some pollsters. However, we hope this analysis can provide more of a numerical backing in terms of identifiying potential differences in polling error between pollsters.

This is the code we will be using for all of our t-tests. It allows us to input a designated pollster (des_pollster) and designated state (des_state) to perform a t-test and calculate a 95% confidence interval. The code is explained in the comments.

```r
func_ttest <- function(des_pollster, des_state) {
  compare_both %>% #dataframe of all of our polls
    # filter for only our designated pollster and state
    filter(pollster == des_pollster, state == des_state) %>%
    # filter for polls conducted after August 1st, 2020
    filter(date > '2020-08-01') %>%
    # join the each poll margin results with the actual margin of victory of either
    # Biden / Trump in the state
    inner_join(filtered_results, by = "state") %>%
    # calculate EACH poll's error by subtracting the predicted poll margin (poll_margin) by
    # the actual winning margin (actual_margin)
    mutate(each_error = poll_margin - actual_margin) %>%

    # we are only interested in performing a t-test on the errors,
    # so we select it, pull it, and run a t test on it.
    select(each_error) %>%
    pull() %>%
    t.test(conf.level = 0.95)
}
```

The first state we will be analyizng is Georgia.

## 4.1: Georgia (+0.24%)

First, we will find the pollsters that have had published at least 3 polls in Georgia from August 2020 to Election Day, and store that in `top_pollsters_GA`.

```r
top_pollsters_GA  <- compare_both %>%
  filter(state %in% c("georgia")) %>%
  filter(date > '2020-08-01') %>%
  group_by(pollster) %>%
  summarize(count = n()) %>%
  filter(count > 2) %>%
  arrange(desc(count)) %>%
  select(pollster) %>%
  pull()

top_pollsters_GA
```

```
## [1] "SurveyMonkey"          "Landmark Communications"
## [3] "Morning Consult"       "Data for Progress"
## [5] "University of Georgia" "YouGov"
```

Next, we will run a t-test of the pollsters listed to determine the predicted polling error given each pollsters' submitted polls. Each t-test for each pollster is listed below.

```r
for (x in top_pollsters_GA) {
  print(x)
  print(func_ttest(x, "georgia")) # print
}
```

```
## [1] "SurveyMonkey"
##
##  One Sample t-test
##
## data:  .
## t = 4.6219, df = 21, p-value = 0.000147
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   0.4117199 1.0852981
## sample estimates:
## mean of x
##   0.748509
##
## [1] "Landmark Communications"
##
##  One Sample t-test
##
## data:  .
## t = -2.8032, df = 6, p-value = 0.03103
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -5.6585548 -0.3840375
## sample estimates:
## mean of x
## -3.021296
##
## [1] "Morning Consult"
##
##  One Sample t-test
##
## data:  .
## t = -0.18991, df = 5, p-value = 0.8568
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -2.455305  2.117474
## sample estimates:
##  mean of x
## -0.1689152
##
## [1] "Data for Progress"
##
##  One Sample t-test
##
## data:  .
## t = 0.64663, df = 2, p-value = 0.5842
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -2.43735  3.29952
## sample estimates:
## mean of x
## 0.4310848
```

```
## 
## [1] "University of Georgia"
## 
##  One Sample t-test
## 
## data:  .
## t = -1.4851, df = 2, p-value = 0.2758
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.347050  1.142552
## sample estimates:
##  mean of x
## -0.6022486
## 
## [1] "YouGov"
## 
##  One Sample t-test
## 
## data:  .
## t = -0.40804, df = 2, p-value = 0.7228
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.719720  2.248556
## sample estimates:
##  mean of x
## -0.2355819
```

**SurveyMonkey**

SurveyMonkey conducted 22 polls in Georgia from August 2020 to November 2020. The average error of those polls was **0.7485%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 0.412% to 1.085%. The calculated t-statistic was 4.62, and the calculated p-value was 0.000147, less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that SurveyMonkey polls in Georgia differed systematically from the actual results.

**Landmark Communications**

Landmark Communications conducted 7 polls in Georgia from August 2020 to November 2020. The average error of those polls was **3.02196%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -5.659% to -0.38404%. The calculated t-statistic was -2.80, and the calculated p-value was 0.03103, less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Landmark Communications polls in Georgia differed systematically from the actual results.

**Morning Consult**

Morning Consult conducted 6 polls in Georgia from August 2020 to November 2020. The average error of those polls was **0.1689%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -2.4553% to 2.117%. The calculated t-statistic was -0.190, and the calculated p-value was 0.8568, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Morning Consult polls in Georgia differed systematically from the actual results.

**Data for Progress**

Data for Progress conducted 3 polls in Georgia from August 2020 to November 2020. The average error of those polls was **0.431%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -2.43735% to 3.29952%. The calculated t-statistic was 0.6466, and the calculated p-value was 0.5842, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Data for Progress polls in Georgia differed systematically from the actual results.

**University of Georgia**

University of Georgia conducted 3 polls in Georgia from August 2020 to November 2020. The average error of those polls was **0.60225%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -2.34705% to 1.143%. The calculated t-statistic was -1.4851, and the calculated p-value was 0.2758, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that University of Georgia polls in Georgia differed systematically from the actual results.

**YouGov**

YouGov conducted 3 polls in Georgia from August 2020 to November 2020. The average error of those polls was **0.02356%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -0.2356% to 2.2486%. The calculated t-statistic was -0.40804, and the calculated p-value was 0.7228, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that YouGov polls in Georgia differed systematically from the actual results.

We compiled our results below into a table, and then plotted the estimated error on the y axis, pollster on the x axis, and whether we were able to reject the null with 95% confidence or not with either a light red (false) or light blue (true), arranged in ascending order by error.

```
ga_results <- compare_both %>%
   filter(state == "georgia") %>%
   filter(date > '2020-08-01') %>%
  filter(pollster %in% top_pollsters_GA) %>%
   inner_join(filtered_results, by = "state") %>%
   mutate(each_error = poll_margin - actual_margin) %>%
   group_by(pollster) %>%
  do(tidy(t.test(.$each_error)))


ga_results %>%
  arrange(abs(estimate)) %>%
  select(pollster, estimate, p.value, conf.low, conf.high) %>%
  mutate(reject95 = p.value < 0.05) %>%
  print() %>%
  arrange(abs(estimate)) %>%
  ggplot(aes(x = reorder(pollster, abs(estimate)), y = estimate, fill = reject95,
             ymin = conf.low, ymax = conf.high)) + geom_col() + theme_linedraw() +
   labs(x = "Pollster", y = "Estimated Error", fill = "Reject null?", title = "Polling Error in Georgia du
   scale_x_discrete(guide = guide_axis(n.dodge = 5)) +geom_errorbar()
```

```
## # A tibble: 6 x 6
## # Groups:   pollster [6]
##   pollster                estimate  p.value conf.low conf.high reject95
##   <chr>                      <dbl>    <dbl>    <dbl>     <dbl> <lgl>
## 1 Morning Consult           -0.169 0.857      -2.46      2.12 FALSE
## 2 YouGov                    -0.236 0.723      -2.72      2.25 FALSE
## 3 Data for Progress          0.431 0.584      -2.44      3.30 FALSE
## 4 University of Georgia     -0.602 0.276      -2.35      1.14 FALSE
## 5 SurveyMonkey               0.749 0.000147    0.412     1.09 TRUE
## 6 Landmark Communications   -3.02  0.0310     -5.66     -0.384 TRUE
```

Polling Error in Georgia during the 2020 presidential election
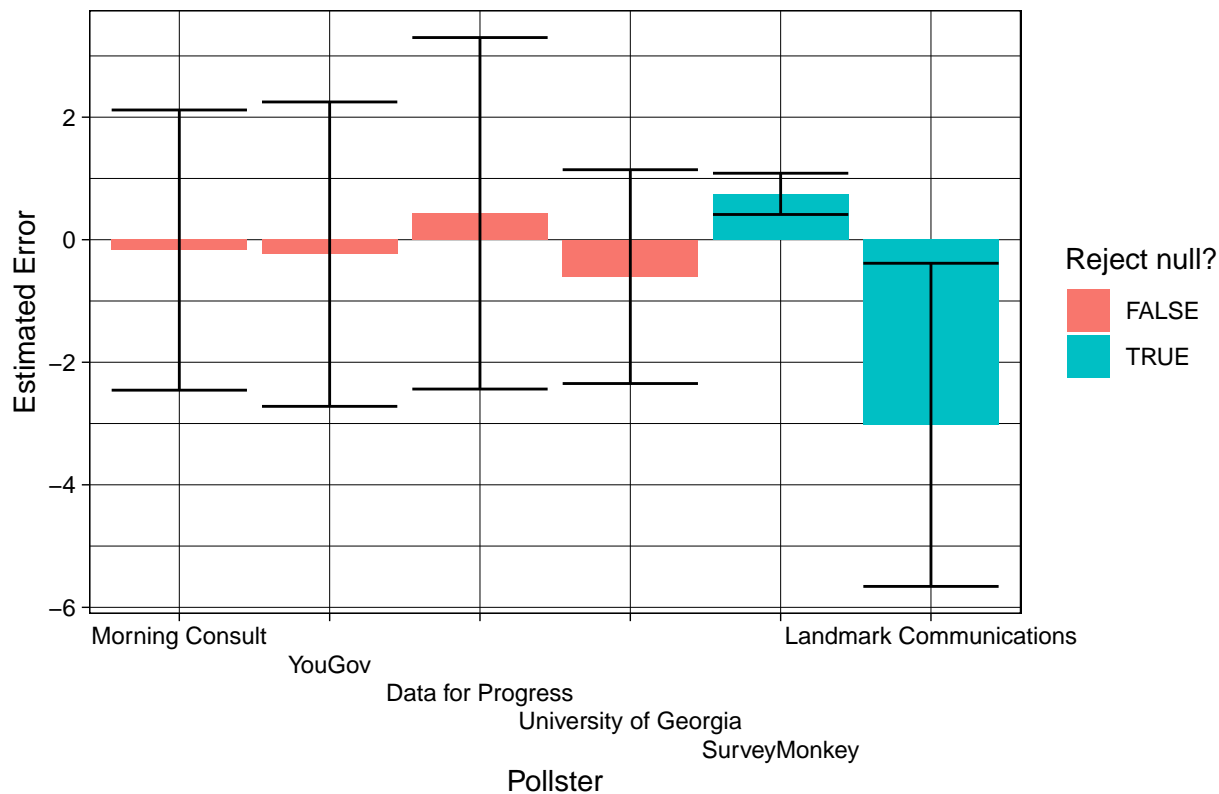
(Combine analyses to provide see which pollsters were most accurate nationwide)

We were able to reject the null hypothesis (that their error was not equal 0) with 95% confidence for 2 pollsters: SurveyMonkey and Landmark Communications. However, this does not mean that the pollsters for which we did not reject the null hypothesis are accurate; absence of evidence of systematic error is not the same as evidence for accuracy. Like for the other states, one of the largest reasons for variation / lack of evidence is simply because of the lack of polls. While some pollsters had significant numbers of polls conducted (such as SurveyMonkey), allowing us to narrow our confidence interval, others had far less, so the confidence interval is much wider.

Overall, polls seem relatively close overall to the actual results, and they are mixed in terms of whether they favored Biden (positive values) or Trump (negative values).

The two most accurate pollsters from our analysis are: Morning Consult (-0.17%) and YouGov (-0.24%). The two most inaccurate pollsters from our analysis are: SurveyMonkey (+0.749%) and Landmark Communications (-3.02%).

## 4.2: Arizona (+0.31%)

First, we will find the pollsters that have had published at least 3 polls in Arizona from August 2020 to Election Day, and store that in `top_pollsters_AZ`.

```
top_pollsters_AZ  <- compare_both %>%
  filter(state %in% c("arizona")) %>%
  filter(date > '2020-08-01') %>%
  group_by(pollster) %>%
  summarize(count = n()) %>%
  filter(count > 2) %>%
  arrange(desc(count)) %>%
  select(pollster) %>%
  pull()

top_pollsters_AZ
```

```
## [1] "SurveyMonkey"
## [2] "Redfield & Wilton Strategies"
## [3] "Change Research"
## [4] "Ipsos"
## [5] "Morning Consult"
## [6] "Data for Progress"
## [7] "Data Orbital"
## [8] "OH Predictive Insights"
## [9] "Patinkin Research Strategies"
## [10] "Siena College/The New York Times Upshot"
## [11] "Trafalgar Group"
```

Next, we will run a t-test of the pollsters listed to determine the predicted polling error given each pollsters' submitted polls. Each t-test for each pollster is listed below.

```
#func_ttest will run a t-test given a designated pollster (des_pollster) and des_state (des_state)

for (x in top_pollsters_AZ) {
  print(x)
  print(func_ttest(x, "arizona")) # print
}
```

```
## [1] "SurveyMonkey"
##
##  One Sample t-test
##
## data:  .
## t = 8.9249, df = 21, p-value = 1.365e-08
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  4.174096 6.710303
## sample estimates:
## mean of x
##    5.4422
##
## [1] "Redfield & Wilton Strategies"
##
##  One Sample t-test
##
## data:  .
## t = 5.5496, df = 7, p-value = 0.0008602
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.477180 6.155401
## sample estimates:
## mean of x
##    4.31629
##
## [1] "Change Research"
##
##  One Sample t-test
##
## data:  .
## t = 4.6915, df = 6, p-value = 0.003355
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.76607 5.61651
## sample estimates:
```

```
## mean of x
##   3.69129
##
## [1] "Ipsos"
##
##  One Sample t-test
##
## data:  .
## t = 4.4065, df = 5, p-value = 0.006979
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.7741046 2.9418096
## sample estimates:
## mean of x
##  1.857957
##
## [1] "Morning Consult"
##
##  One Sample t-test
##
## data:  .
## t = 1.2993, df = 5, p-value = 0.2505
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.193118  6.675699
## sample estimates:
## mean of x
##   2.24129
##
## [1] "Data for Progress"
##
##  One Sample t-test
##
## data:  .
## t = 1.6704, df = 3, p-value = 0.1934
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -1.644086  5.276667
## sample estimates:
## mean of x
##   1.81629
##
## [1] "Data Orbital"
##
##  One Sample t-test
##
## data:  .
## t = 2.8602, df = 3, p-value = 0.06457
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.3313725  6.2139533
## sample estimates:
## mean of x
##   2.94129
##
## [1] "OH Predictive Insights"
##
##  One Sample t-test
##
```

```
## data:  .
## t = 2.9408, df = 3, p-value = 0.06047
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.3958257 10.0284066
## sample estimates:
## mean of x
##   4.81629
##
## [1] "Patinkin Research Strategies"
##
##  One Sample t-test
##
## data:  .
## t = 3.626, df = 2, p-value = 0.06835
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.8131879  9.5291021
## sample estimates:
## mean of x
##  4.357957
##
## [1] "Siena College/The New York Times Upshot"
##
##  One Sample t-test
##
## data:  .
## t = 8.3431, df = 2, p-value = 0.01406
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   3.563374 11.152540
## sample estimates:
## mean of x
##  7.357957
##
## [1] "Trafalgar Group"
##
##  One Sample t-test
##
## data:  .
## t = -3.9044, df = 2, p-value = 0.05978
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -6.1841361  0.3000502
## sample estimates:
## mean of x
## -2.942043
```

**SurveyMonkey**

SurveyMonkey conducted 22 polls in Arizona from August 2020 to November 2020. The average error of those polls was **5.4422%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 4.174% to 6.710%. The calculated t-statistic was 8.9249, and the calculated p-value was significantly less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that SurveyMonkey polls in Arizona differed systematically from the actual results.

**Redfield & Wilton Strategies**

Redfield & Wilton Strategies conducted 8 polls in Arizona from August 2020 to November 2020. The average error

of those polls was **4.3163%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 2.477% to 6.155%. The calculated t-statistic was 5.5496, and the calculated p-value was 0.0009, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Redfield & Wilton Strategies polls in Arizona differed systematically from the actual results.

### Change Research

Change Research conducted 7 polls in Arizona from August 2020 to November 2020. The average error of those polls was **3.69129%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 1.7661% to 5.6165%. The calculated t-statistic was 4.6915, and the calculated p-value was 0.003, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Change Research polls in Arizona differed systematically from the actual results.

### Ipsos

Ipsos conducted 6 polls in Arizona from August 2020 to November 2020. The average error of those polls was **1.8580%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 0.7741% to 2.9418%. The calculated t-statistic was 4.4065, and the calculated p-value was 0.007, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Ipsos polls in Arizona differed systematically from the actual results.

### Morning Consult

Morning Consult conducted 5 polls in Arizona from August 2020 to November 2020. The average error of those polls was **2.24129%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -2.1931% to 6.6757%. The calculated t-statistic was 1.2993, and the calculated p-value was 0.2505, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Morning Consult polls in Arizona differed systematically from the actual results.

### Data for Progress

Data for Progress conducted 4 polls in Arizona from August 2020 to November 2020. The average error of those polls was **1.81629%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -1.6441% to 5.2767%. The calculated t-statistic was 1.6704, and the calculated p-value was 0.1934, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Data for Progress polls in Arizona differed systematically from the actual results.

### Data Orbital

Data Orbital conducted 4 polls in Arizona from August 2020 to November 2020. The average error of those polls was **2.94129%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -0.33137% to 6.2140%. The calculated t-statistic was 2.8602, and the calculated p-value was 0.0646, which is just barely greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Data Orbital polls in Arizona differed systematically from the actual results.

### OH Predictive Insights

OH Predictive Insights conducted 4 polls in Arizona from August 2020 to November 2020. The average error of those polls was **4.81629%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -0.3958% to 10.0284%. The calculated t-statistic was 2.9408, and the calculated p-value was 0.06047, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that OH Predictive Insights polls in Arizona differed systematically from the actual results.

### Patinkin Research Strategies

Patinkin Research Strategies conducted 3 polls in Arizona from August 2020 to November 2020. The average error of those polls was **4.357957%** more positive for Biden than the actual final result. The 95% confidence interval for

the error is -0.813% to 9.529%. The calculated t-statistic was 3.626, and the calculated p-value was 0.06835, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Patinkin Research Strategies polls in Arizona differed systematically from the actual results.

**Siena College / The New York Times Upshot**

Siena College / The New York Times Upshot conducted 3 polls in Arizona from August 2020 to November 2020. The average error of those polls was **7.3580%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 3.5634% to 11.1525%. The calculated t-statistic was 8.3431, and the calculated p-value was 0.01406, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Siena College / The New York Times Upshot polls in Arizona differed systematically from the actual results.

**Trafalgar Group**

Trafalgar Group conducted 3 polls in Arizona from August 2020 to November 2020. The average error of those polls was **2.942043%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -6.1841% to 0.300%. The calculated t-statistic was -3.9044, and the calculated p-value was 0.05978, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Trafalgar Group polls in Arizona differed systematically from the actual results.

We compiled our results below into a table, and then plotted the estimated error on the y axis, pollster on the x axis, and whether we were able to reject the null with 95% confidence or not with either a light red (false) or light blue (true), arranged in ascending order by error.

```
 az_results <- compare_both %>%
    filter(state == "arizona") %>%
    filter(date > '2020-08-01') %>%
  filter(pollster %in% top_pollsters_AZ) %>%
    inner_join(filtered_results, by = "state") %>%
    mutate(each_error = poll_margin - actual_margin) %>%
    group_by(pollster) %>%
  do(tidy(t.test(.$each_error)))


az_results %>%
  arrange(abs(estimate)) %>%
  select(pollster, estimate, p.value, conf.low, conf.high) %>%
  mutate(reject95 = p.value < 0.05) %>%
  print() %>%
  arrange(abs(estimate)) %>%
  ggplot(aes(x = reorder(pollster, abs(estimate)), y = estimate, fill = reject95,
            ymin = conf.low, ymax = conf.high)) + geom_col() + theme_linedraw() +
   labs(x = "Pollster", y = "Estimated Error", fill = "Reject null?", title = "Polling Error in Arizona du
   scale_x_discrete(guide = guide_axis(n.dodge = 5)) +geom_errorbar()
```

```
## # A tibble: 11 x 6
## # Groups:   pollster [11]
##    pollster                        estim~1 p.value conf.~2 conf.~3 rejec~4
##    <chr>                             <dbl>   <dbl>   <dbl>   <dbl> <lgl>
##  1 Data for Progress                  1.82 1.93e-1   -1.64    5.28 FALSE
##  2 Ipsos                              1.86 6.98e-3   0.774    2.94 TRUE
##  3 Morning Consult                    2.24 2.51e-1   -2.19    6.68 FALSE
##  4 Data Orbital                       2.94 6.46e-2  -0.331    6.21 FALSE
##  5 Trafalgar Group                   -2.94 5.98e-2   -6.18   0.300 FALSE
##  6 Change Research                    3.69 3.36e-3    1.77    5.62 TRUE
##  7 Redfield & Wilton Strategies       4.32 8.60e-4    2.48    6.16 TRUE
```

```
##  8 Patinkin Research Strategies              4.36 6.84e-2  -0.813   9.53  FALSE
##  9 OH Predictive Insights                    4.82 6.05e-2  -0.396  10.0   FALSE
## 10 SurveyMonkey                              5.44 1.37e-8   4.17    6.71  TRUE
## 11 Siena College/The New York Times Ups~     7.36 1.41e-2   3.56   11.2   TRUE
## # ... with abbreviated variable names 1: estimate, 2: conf.low, 3: conf.high,
## #   4: reject95
```
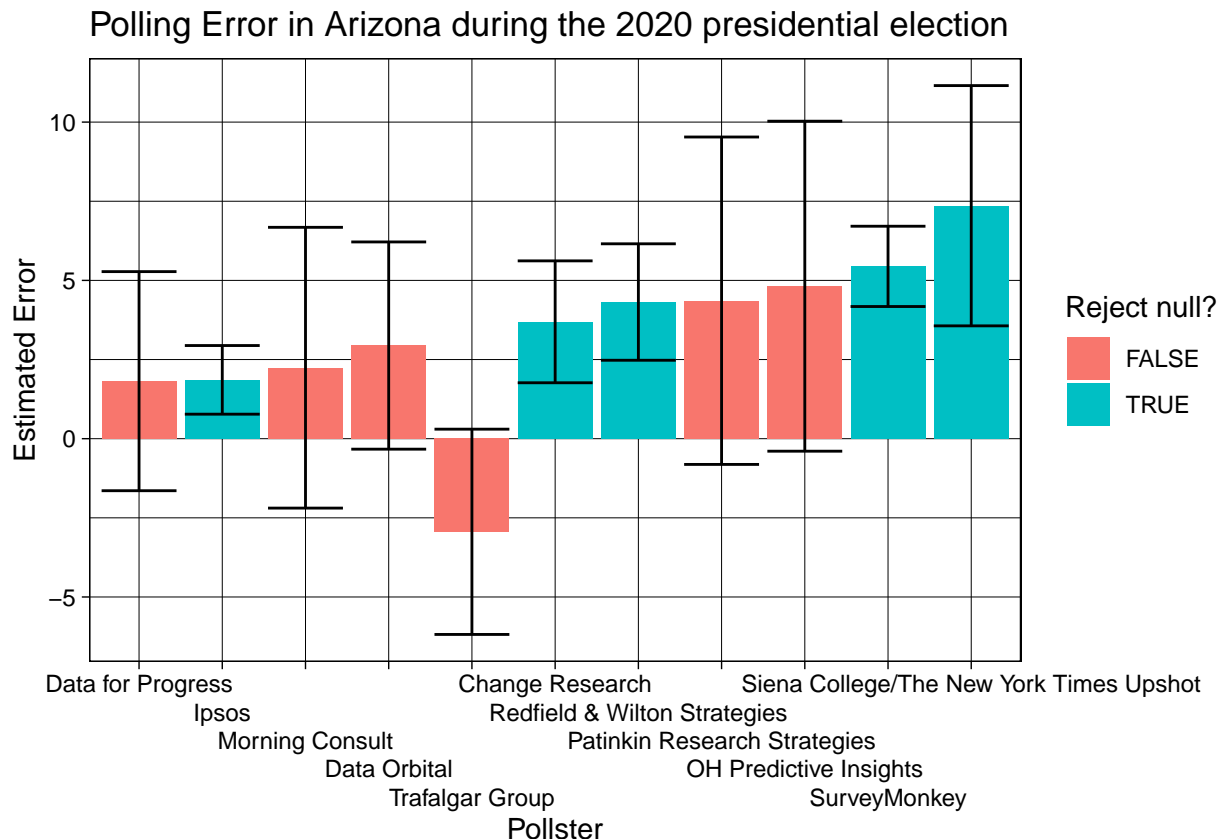
## Polling Error in Arizona during the 2020 presidential election



(Combine analyses to provide see which pollsters were most accurate nationwide)

We were able to reject the null hypothesis (that their error was not equal 0) with 95% confidence for 5 pollsters: Ipsos, Change Research, Redfield & Wilton Strategies, Survey Monkey, and the Siena College / The New York Times Upshot. However, this does not mean that the pollsters for which we did not reject the null hypothesis are accurate; absence of evidence of systematic error is not the same as evidence for accuracy. In fact, some of the pollsters for which we saw systematic error (such as Ipsos) still had lower average error than some of the pollsters where we failed to reject the null hypothesis. Like for the other states, one of the largest reasons for variation / lack of evidence is simply because of the lack of polls.

Compared to Georgia, polls here appeared to sway farther from the actual results, and most error was in Biden's direction, rather than a mix of both.

The three most accurate pollsters from our analysis are: Data for Progress (+1.82%), Ipsos (+1.86%), and Morning Consult (+2.24%). The three most inaccurate pollsters from our analysis are: OH Predictive Insights (+4.82%), SurveyMonkey (+5.44%), and Siena College / The New York Times Upshot (+7.36%).

## 4.3 Wisconsin (+0.63%)

First, we will find the pollsters that have had published at least 3 polls in Wisconsin from August 2020 to Election Day, and store that in `top_pollsters_WI`.

```
top_pollsters_WI <- compare_both %>%
  filter(state %in% c("wisconsin")) %>%
```

```
  filter(date > '2020-08-01') %>%
  group_by(pollster) %>%
  summarize(count = n()) %>%
  filter(count > 2) %>%
  arrange(desc(count)) %>%
  select(pollster) %>%
  pull()

top_pollsters_WI
```

```
##  [1] "Morning Consult"
##  [2] "SurveyMonkey"
##  [3] "Redfield & Wilton Strategies"
##  [4] "Change Research"
##  [5] "Ipsos"
##  [6] "Trafalgar Group"
##  [7] "YouGov"
##  [8] "Marquette University Law School"
##  [9] "Siena College/The New York Times Upshot"
## [10] "Susquehanna Polling & Research Inc."
```

Next, we will run a t-test of the pollsters listed to determine the predicted polling error given each pollsters'
submitted polls. Each t-test for each pollster is listed below.

```
for (x in top_pollsters_WI) {
  print(x)
  print(func_ttest(x, "wisconsin")) # print
}
```

```
## [1] "Morning Consult"
##
##  One Sample t-test
##
## data:  .
## t = 30.393, df = 48, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  7.214856 8.237067
## sample estimates:
## mean of x
##  7.725962
##
## [1] "SurveyMonkey"
##
##  One Sample t-test
##
## data:  .
## t = 16.856, df = 21, p-value = 1.112e-13
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   8.829322 11.314660
## sample estimates:
## mean of x
##  10.07199
##
## [1] "Redfield & Wilton Strategies"
##
```

```
##  One Sample t-test
##
## data:  .
## t = 6.871, df = 7, p-value = 0.0002374
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   5.245463 10.750338
## sample estimates:
## mean of x
##    7.9979
##
## [1] "Change Research"
##
##  One Sample t-test
##
## data:  .
## t = 8.95, df = 6, p-value = 0.0001086
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  4.422953 7.751419
## sample estimates:
## mean of x
##  6.087186
##
## [1] "Ipsos"
##
##  One Sample t-test
##
## data:  .
## t = 9.9986, df = 5, p-value = 0.0001711
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  4.858285 8.220849
## sample estimates:
## mean of x
##  6.539567
##
## [1] "Trafalgar Group"
##
##  One Sample t-test
##
## data:  .
## t = 1.347, df = 5, p-value = 0.2358
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -0.7172656  2.2963998
## sample estimates:
## mean of x
## 0.7895671
##
## [1] "YouGov"
##
##  One Sample t-test
##
## data:  .
## t = 6.4218, df = 4, p-value = 0.003023
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.049960 7.695841
```

```
## sample estimates:
## mean of x
##    5.3729
##
## [1] "Marquette University Law School"
##
##  One Sample t-test
##
## data:  .
## t = 13.416, df = 3, p-value = 0.0008953
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.954207 4.791594
## sample estimates:
## mean of x
##    3.8729
##
## [1] "Siena College/The New York Times Upshot"
##
##  One Sample t-test
##
## data:  .
## t = 4.3318, df = 2, p-value = 0.04938
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   0.05418161 16.02495264
## sample estimates:
## mean of x
##  8.039567
##
## [1] "Susquehanna Polling & Research Inc."
##
##  One Sample t-test
##
## data:  .
## t = 1.1788, df = 2, p-value = 0.3597
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.755016  4.834150
## sample estimates:
## mean of x
##  1.039567
```

**Morning Consult**

Morning Consult conducted 49 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **7.726%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 7.215% to 8.237%. The calculated t-statistic was 30.393, and the calculated p-value was $2.2 * 10^{-16}$, which is much less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Morning Consult polls in Wisconsin differed systematically from the actual results.

**SurveyMonkey**

SurveyMonkey conducted 22 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **10.072%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 8.829% to 11.315%. The calculated t-statistic was 16.856, and the calculated p-value was $1.1 * 10^{-13}$, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that SurveyMonkey polls in Wisconsin differed systematically from the actual results.

### Redfield & Wilton Strategies

Redfield & Wilton Strategies conducted 8 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **7.9979%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 5.245% to 10.750%. The calculated t-statistic was 6.871, and the calculated p-value was 0.00024, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Redfield & Wilton Strategies polls in Wisconsin differed systematically from the actual results.

### Change Research

Change Research conducted 7 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **6.087186%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 4.423% to 7.751%. The calculated t-statistic was 8.95, and the calculated p-value was 0.00011, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Change Research polls in Wisconsin differed systematically from the actual results.

### Ipsos

Ipsos conducted 6 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **6.5396%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 4.858% to 8.220849%. The calculated t-statistic was 9.999, and the calculated p-value was 0.00017, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Ipsos polls in Wisconsin differed systematically from the actual results.

### Trafalgar Group

Trafalgar Group conducted 6 polls in Georgia from August 2020 to November 2020. The average error of those polls was **0.7896%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -0.71727% to 2.2964%. The calculated t-statistic was 1.347, and the calculated p-value was 0.2358, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Trafalgar polls in Wisconsin differed systematically from the actual results.

### YouGov

Ipsos conducted 5 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **5.3729%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 3.050% to 7.696%. The calculated t-statistic was 6.4218, and the calculated p-value was 0.003023, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that YouGov polls in Wisconsin differed systematically from the actual results.

### Marquette University Law School

Marquette University Law School conducted 4 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **3.8729%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 2.954% to 4.792%. The calculated t-statistic was 13.416, and the calculated p-value was 0.0008953, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Marquette University Law School polls in Wisconsin differed systematically from the actual results.

### Siena College / The New York Times Upshot

Siena College / The New York Times Upshot conducted 3 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **8.039567%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 0.05418% to 16.0250% (wide because of variation between the 3 polls). The calculated t-statistic was 4.3318, and the calculated p-value was 0.04938, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Siena College / The New York Times Upshot polls in Wisconsin differed systematically from the actual results.

### Susquehanna Polling & Research Inc.

Susquehanna Polling and Research conducted 3 polls in Wisconsin from August 2020 to November 2020. The average error of those polls was **1.0396%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -2.755% to 4.834%. The calculated t-statistic was 1.1788, and the calculated p-value was 0.3597, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Susquehanna Polling & Research polls in Wisconsin differed systematically from the actual results.

We compiled our results below into a table, and then plotted the estimated error on the y axis, pollster on the x axis, and whether we were able to reject the null with 95% confidence or not with either a light red (false) or light blue (true), arranged in ascending order by error.

```r
 wi_results <- compare_both %>%
    filter(state == "wisconsin") %>%
    filter(date > '2020-08-01') %>%
  filter(pollster %in% top_pollsters_WI) %>%
    inner_join(filtered_results, by = "state") %>%
    mutate(each_error = poll_margin - actual_margin) %>%
    group_by(pollster) %>%
  do(tidy(t.test(.$each_error)))


wi_results %>%
  arrange(abs(estimate)) %>%
  select(pollster, estimate, p.value, conf.low, conf.high) %>%
  mutate(reject95 = p.value < 0.05) %>%
  print() %>%
  arrange(abs(estimate)) %>%
  ggplot(aes(x = reorder(pollster, abs(estimate)), y = estimate, fill = reject95,
            ymin = conf.low, ymax = conf.high)) + geom_col() + theme_linedraw() +
   labs(x = "Pollster", y = "Estimated Error", fill = "Reject null?", title = "Polling Error in Wisconsin
   scale_x_discrete(guide = guide_axis(n.dodge = 5)) +geom_errorbar()
```

```
## # A tibble: 10 x 6
## # Groups:   pollster [10]
##    pollster                           estim~1  p.value conf.~2 conf.~3 rejec~4
##    <chr>                                <dbl>    <dbl>   <dbl>   <dbl> <lgl>
##  1 Trafalgar Group                      0.790 2.36e- 1 -0.717    2.30 FALSE
##  2 Susquehanna Polling & Research Inc.  1.04  3.60e- 1 -2.76     4.83 FALSE
##  3 Marquette University Law School      3.87  8.95e- 4  2.95     4.79 TRUE
##  4 YouGov                               5.37  3.02e- 3  3.05     7.70 TRUE
##  5 Change Research                      6.09  1.09e- 4  4.42     7.75 TRUE
##  6 Ipsos                                6.54  1.71e- 4  4.86     8.22 TRUE
##  7 Morning Consult                      7.73  5.23e-33  7.21     8.24 TRUE
##  8 Redfield & Wilton Strategies         8.00  2.37e- 4  5.25    10.8  TRUE
##  9 Siena College/The New York Times Up~ 8.04  4.94e- 2  0.0542  16.0  TRUE
## 10 SurveyMonkey                        10.1   1.11e-13  8.83    11.3  TRUE
## # ... with abbreviated variable names 1: estimate, 2: conf.low, 3: conf.high,
## #   4: reject95
```

## Polling Error in Wisconsin during the 2020 presidential election
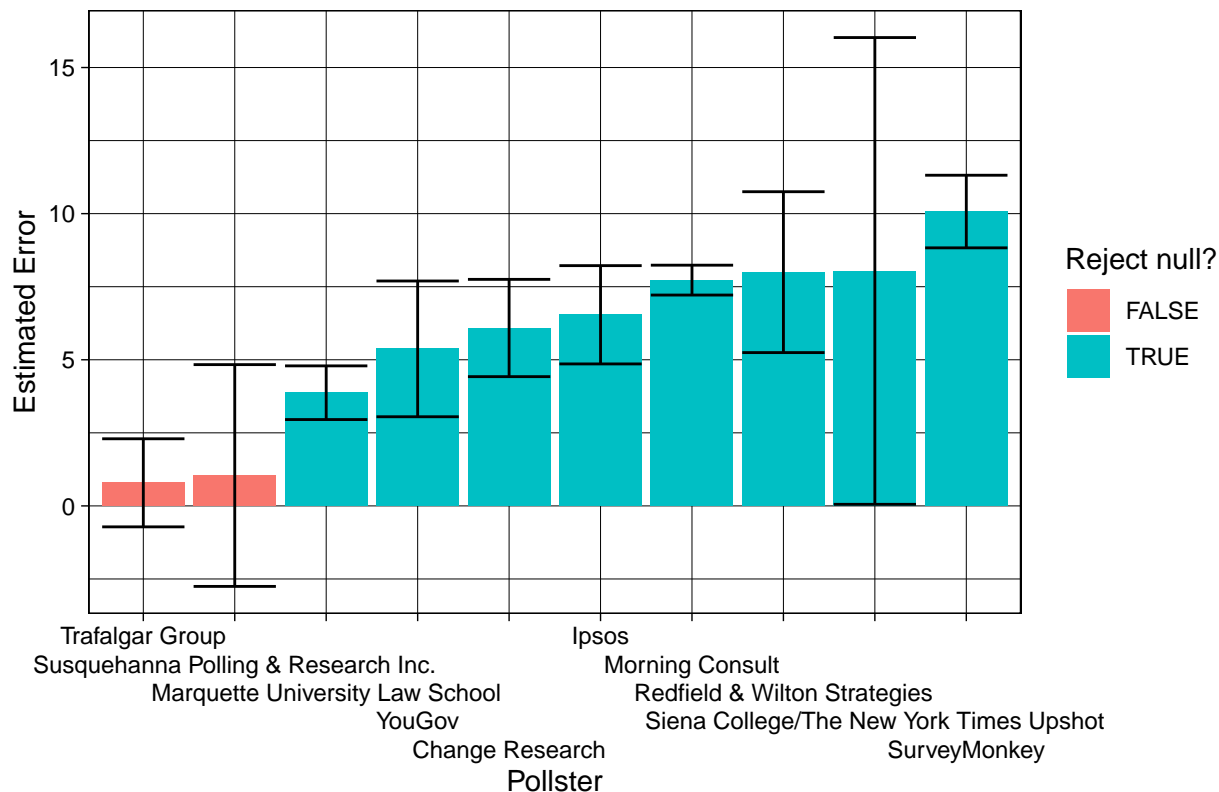


(Combine analyses to provide see which pollsters were most accurate nationwide)

We were able to reject the null hypothesis (that their error was not equal 0) with 95% confidence for 8 pollsters, and only failed to reject for two. Again, this does not mean that the pollsters for which we did not reject the null hypothesis are accurate; absence of evidence of systematic error is not the same as evidence for accuracy.

However, compared to Arizona and Georgia, overall polling error seemed to be significantly higher in favor of Biden compared to Arizona & Georgia. This may be due to some of the demographic reasons mentioned in Part 1 — where certain likely voters in the Midwest are harder to reach than other states. This may explain why Trafalgar's polls succeeded here; they advertised heavily that they are able to reach rural and conservative voters — while not saying specifically how — but that promise may have been necessary in a state like Wisconsin.

The two most accurate pollsters were the Trafalgar Group (+0.79%) and Susquehanna Polling & Research Inc. (+1.04%), while the two least accurate were Siena College / The New York Times Upshot (+8.04%) and SurveyMonkey (+10.1%).

## 4.4 Pennslyvania (+1.16%)

First, we will find the pollsters that have had published at least 3 polls in Pennsylvania from August 2020 to Election Day, and store that in `top_pollsters_PA`.

```
top_pollsters_PA <- compare_both %>%
  filter(state %in% c("pennsylvania")) %>%
  filter(date > '2020-08-01') %>%
  group_by(pollster) %>%
  summarize(count = n()) %>%
  filter(count > 2) %>%
  arrange(desc(count)) %>%
  select(pollster) %>%
  pull()

top_pollsters_PA
```

```
##  [1] "SurveyMonkey"
##  [2] "Redfield & Wilton Strategies"
##  [3] "Change Research"
##  [4] "Ipsos"
##  [5] "Morning Consult"
##  [6] "Trafalgar Group"
##  [7] "YouGov"
##  [8] "Quinnipiac University"
##  [9] "Emerson College"
## [10] "Monmouth University"
## [11] "Muhlenberg College"
## [12] "Opinion Savvy/InsiderAdvantage"
## [13] "Rasmussen Reports/Pulse Opinion Research"
## [14] "Siena College/The New York Times Upshot"
```

Next, we will run a t-test of the pollsters listed to determine the predicted polling error given each pollsters' submitted polls. Each t-test for each pollster is listed below.

```r
for (x in top_pollsters_PA) {
  print(x)
  print(func_ttest(x, "pennsylvania")) # print
}
```

```
## [1] "SurveyMonkey"
##
##  One Sample t-test
##
## data:  .
## t = 25.484, df = 21, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  4.445135 5.235098
## sample estimates:
## mean of x
##  4.840116
##
## [1] "Redfield & Wilton Strategies"
##
##  One Sample t-test
##
## data:  .
## t = 11.442, df = 7, p-value = 8.747e-06
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.835880 5.834352
## sample estimates:
## mean of x
##  4.835116
##
## [1] "Change Research"
##
##  One Sample t-test
##
## data:  .
## t = 8.0925, df = 6, p-value = 0.0001909
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.67888 3.13421
```

```
## sample estimates:
## mean of x
##   2.406545
##
## [1] "Ipsos"
##
##   One Sample t-test
##
## data:  .
## t = 7.0727, df = 5, p-value = 0.0008742
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.494281 5.342618
## sample estimates:
## mean of x
##    3.91845
##
## [1] "Morning Consult"
##
##   One Sample t-test
##
## data:  .
## t = 6.6201, df = 5, p-value = 0.001184
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.436818 7.800081
## sample estimates:
## mean of x
##    5.61845
##
## [1] "Trafalgar Group"
##
##   One Sample t-test
##
## data:  .
## t = -0.37772, df = 5, p-value = 0.7211
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.197645  1.634544
## sample estimates:
##   mean of x
## -0.2815503
##
## [1] "YouGov"
##
##   One Sample t-test
##
## data:  .
## t = 7.263, df = 4, p-value = 0.001909
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.357415 7.512818
## sample estimates:
## mean of x
##  5.435116
##
## [1] "Quinnipiac University"
##
##   One Sample t-test
```

```
##
## data:   .
## t = 5.7866, df = 3, p-value = 0.01027
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##    3.526064 12.144169
## sample estimates:
## mean of x
##   7.835116
##
## [1] "Emerson College"
##
##   One Sample t-test
##
## data:   .
## t = 4.094, df = 2, p-value = 0.0548
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -0.1941224  7.8110218
## sample estimates:
## mean of x
##    3.80845
##
## [1] "Monmouth University"
##
##   One Sample t-test
##
## data:   .
## t = 2.1547, df = 2, p-value = 0.164
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -4.653965 13.990864
## sample estimates:
## mean of x
##    4.66845
##
## [1] "Muhlenberg College"
##
##   One Sample t-test
##
## data:   .
## t = 4.7266, df = 2, p-value = 0.04196
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   0.3738667 7.9630327
## sample estimates:
## mean of x
##    4.16845
##
## [1] "Opinion Savvy/InsiderAdvantage"
##
##   One Sample t-test
##
## data:   .
## t = -0.82047, df = 2, p-value = 0.4982
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -9.355059  6.358625
## sample estimates:
```

```
## mean of x
## -1.498217
##
## [1] "Rasmussen Reports/Pulse Opinion Research"
##
##  One Sample t-test
##
## data:  .
## t = 0.65326, df = 2, p-value = 0.5807
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.050828  5.501061
## sample estimates:
## mean of x
## 0.7251164
##
## [1] "Siena College/The New York Times Upshot"
##
##  One Sample t-test
##
## data:  .
## t = 6.9944, df = 2, p-value = 0.01983
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.373867 9.963033
## sample estimates:
## mean of x
##  6.16845
```

**SurveyMonkey**

SurveyMonkey conducted 49 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **4.840%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 4.445% to 5.235%. The calculated t-statistic was 25.484, and the calculated p-value was close to 0, which is definitely less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that SurveyMonkey polls in Pennsylvania differed systematically from the actual results.

**Redfield & Wilton Strategies**

Redfield & Wilton Strategies conducted 8 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **4.835%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 3.836% to 5.834%. The calculated t-statistic was 11.442, and the calculated p-value was 8.747 * 10^-6, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Redfield & Wilton Strategies polls in Pennsylvania differed systematically from the actual results.

**Change Research**

Change Research conducted 7 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **2.4065%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 1.679% to 3.134%. The calculated t-statistic was 8.0925, and the calculated p-value was 0.0001909, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Change Research polls in Pennsylvania differed systematically from the actual results.

**Ipsos**

Ipsos conducted 6 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **3.91845%** more positive for Biden than the actual final result. The 95% confidence interval for the error is

2.4943% to 5.343%. The calculated t-statistic was 7.0727, and the calculated p-value was 0.0008742, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Ipsos polls in Pennsylvania differed systematically from the actual results.

### Morning Consult

Morning Consult conducted 6 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **5.61845%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 3.4368% to 7.8%. The calculated t-statistic was 6.6201, and the calculated p-value was 0.001184, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Morning Consult polls in Pennsylvania differed systematically from the actual results.

### Trafalgar Group

Trafalgar Group conducted 6 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **0.28155%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -2.197645% to 1.634544%. The calculated t-statistic was -0.37772, and the calculated p-value was 0.7211, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Trafalgar polls in Pennsylvania differed systematically from the actual results.

### YouGov

YouGov conducted 5 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **5.435116%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 3.357415% to 7.512818%. The calculated t-statistic was 7.263, and the calculated p-value was 0.001909, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that YouGov polls in Pennsylvania differed systematically from the actual results.

### Quinnipiac University

Quinnipiac University conducted 4 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **7.835116%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 3.526064% to 12.144169%. The calculated t-statistic was 5.7866, and the calculated p-value was 0.01027, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Quinnipiac University polls in Pennsylvania differed systematically from the actual results.

### Emerson College

Emerson College conducted 3 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **3.80845%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -0.1941224% to 7.8110218%. The calculated t-statistic was 4.094, and the calculated p-value was 0.0548, which is slightly greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Emerson College polls in Pennsylvania differed systematically from the actual results.

### Monmouth University

Monmouth University conducted 3 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **4.66845%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -4.653965% to 13.990864%. The calculated t-statistic was 2.1547, and the calculated p-value was 0.164, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Monmouth University polls in Pennsylvania differed systematically from the actual results.

### Muhlenberg University

Muhlenberg University conducted 3 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **4.16845%** more positive for Biden than the actual final result. The 95% confidence interval

for the error is 0.3738667% to 7.9630327%. The calculated t-statistic was 4.7266, and the calculated p-value was 0.04196, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Muhlenberg University polls in Pennsylvania differed systematically from the actual results.

**Opinion Savvy/InsiderAdvantage**

Opinion Savvy/InsiderAdvantage conducted 3 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **1.498217%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -9.355059% to 6.358625%. The calculated t-statistic was -0.82047, and the calculated p-value was 0.4982, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Opinion Savvy/InsiderAdvantage polls in Pennsylvania differed systematically from the actual results.

**Rasmussen Reports / Pulse Opinion Research**

Rasmussen Reports / Pulse Opinion Research conducted 3 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **0.7251164%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -4.05% to 5.50%. The calculated t-statistic was 0.65326, and the calculated p-value was 0.5807, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Rasmussen Reports / Pulse Opinion Research polls in Pennsylvania differed systematically from the actual results.

**Siena College / The New York Times Upshot**

Siena College / The New York Times Upshot conducted 3 polls in Pennsylvania from August 2020 to November 2020. The average error of those polls was **6.16845%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 2.373867% to 9.963033%. The calculated t-statistic was 6.9944, and the calculated p-value was 0.01983, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, meaning we are 95% confident that Siena College / The New York Times Upshot polls in Pennsylvania differed systematically from the actual results.

We compiled our results below into a table, and then plotted the estimated error on the y axis, pollster on the x axis, and whether we were able to reject the null with 95% confidence or not with either a light red (false) or light blue (true), arranged in ascending order by error.

```
pa_results <- compare_both %>%
    filter(state == "pennsylvania") %>%
    filter(date > '2020-08-01') %>%
  filter(pollster %in% top_pollsters_PA) %>%
    inner_join(filtered_results, by = "state") %>%
    mutate(each_error = poll_margin - actual_margin) %>%
    group_by(pollster) %>%
  do(tidy(t.test(.$each_error)))


pa_results %>%
  arrange(abs(estimate)) %>%
  select(pollster, estimate, p.value, conf.low, conf.high) %>%
  mutate(reject95 = p.value < 0.05) %>%
  print() %>%
  arrange(abs(estimate)) %>%
  ggplot(aes(x = reorder(pollster, abs(estimate)), y = estimate, fill = reject95,
             ymin = conf.low, ymax = conf.high)) + geom_col() + theme_linedraw() +
   labs(x = "Pollster", y = "Estimated Error", fill = "Reject null?", title = "Polling Error in Pennsylvan
    scale_x_discrete(guide = guide_axis(n.dodge = 5)) +geom_errorbar()


## # A tibble: 14 x 6
```

```
## # Groups:   pollster [14]
##    pollster                         estim~1  p.value conf.~2 conf.~3 rejec~4
##    <chr>                              <dbl>    <dbl>   <dbl>   <dbl> <lgl>
##  1 Trafalgar Group                   -0.282 7.21e- 1   -2.20    1.63 FALSE
##  2 Rasmussen Reports/Pulse Opinion Res~ 0.725 5.81e- 1  -4.05    5.50 FALSE
##  3 Opinion Savvy/InsiderAdvantage    -1.50  4.98e- 1   -9.36    6.36 FALSE
##  4 Change Research                    2.41  1.91e- 4    1.68    3.13 TRUE
##  5 Emerson College                    3.81  5.48e- 2   -0.194   7.81 FALSE
##  6 Ipsos                              3.92  8.74e- 4    2.49    5.34 TRUE
##  7 Muhlenberg College                 4.17  4.20e- 2    0.374   7.96 TRUE
##  8 Monmouth University                4.67  1.64e- 1   -4.65   14.0  FALSE
##  9 Redfield & Wilton Strategies       4.84  8.75e- 6    3.84    5.83 TRUE
## 10 SurveyMonkey                       4.84  2.81e-17    4.45    5.24 TRUE
## 11 YouGov                             5.44  1.91e- 3    3.36    7.51 TRUE
## 12 Morning Consult                    5.62  1.18e- 3    3.44    7.80 TRUE
## 13 Siena College/The New York Times Up~ 6.17 1.98e- 2   2.37    9.96 TRUE
## 14 Quinnipiac University              7.84  1.03e- 2    3.53   12.1  TRUE
## # ... with abbreviated variable names 1: estimate, 2: conf.low, 3: conf.high,
## #   4: reject95
```
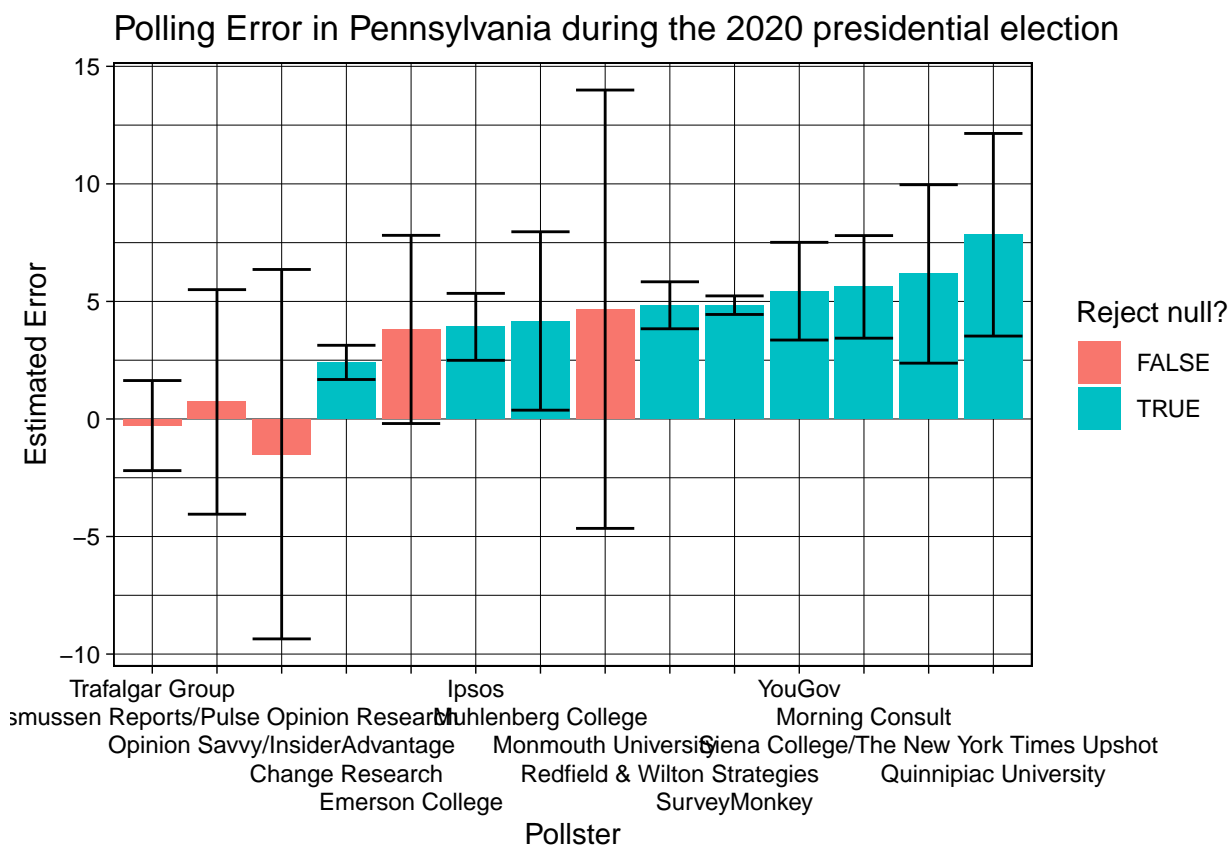


(Combine analyses to provide see which pollsters were most accurate nationwide)

We were able to reject the null hypothesis (that their error was not equal 0) with 95% confidence for 9 pollsters, and failed to reject for 5. Again, this does not mean that the pollsters for which we did not reject the null hypothesis are accurate; absence of evidence of systematic error is not the same as evidence for accuracy.

Polling error appeared to be relatively high in the state similar to Wisconsin, especially compared to Arizona and Georgia. Pennsylvania is also considered part of the Midwest like Wisconsin, which may explain some of the error seen here — and why some Republican-leaning pollsters like Trafalgar were closer to the actual result. Most pollsters overestimated Biden's chances, but there were some that overestimated Trump. While Biden did end up winning the state, the margin of victory by which he did so was significantly lower compared to what many of these pollsters predicted.

The two most accurate pollsters were the Trafalgar Group (-0.282%) and Rasmussen Reports / Pulse Opinion Research (+0.725%), while the two least accurate were Siena College / The New York Times Upshot (+6.17%) and Quinnipiac University (+7.84%).

## 4.5 North Carolina (-1.35%)

First, we will find the pollsters that have had published at least 3 polls in North Carolina from August 2020 to Election Day, and store that in `top_pollsters_NC`.

```
top_pollsters_NC  <- compare_both %>%
  filter(state %in% c("north carolina")) %>%
  filter(date > '2020-08-01') %>%
  group_by(pollster) %>%
  summarize(count = n()) %>%
  filter(count > 2) %>%
  arrange(desc(count)) %>%
  select(pollster) %>%
  pull()

top_pollsters_NC
```

```
##  [1] "SurveyMonkey"
##  [2] "Redfield & Wilton Strategies"
##  [3] "Change Research"
##  [4] "Morning Consult"
##  [5] "Ipsos"
##  [6] "Data for Progress"
##  [7] "East Carolina University"
##  [8] "Emerson College"
##  [9] "Rasmussen Reports/Pulse Opinion Research"
## [10] "Harper Polling"
## [11] "Siena College/The New York Times Upshot"
## [12] "SurveyUSA"
## [13] "Trafalgar Group"
## [14] "YouGov"
```

Next, we will run a t-test of the pollsters listed to determine the predicted polling error given each pollsters' submitted polls. Each t-test for each pollster is listed below.

```
#func_ttest will run a t-test given a designated pollster (des_pollster) and des_state (des_state)

for (x in top_pollsters_NC) {
  print(x)
  print(func_ttest(x, "north carolina")) # print
}
```

```
## [1] "SurveyMonkey"
##
##  One Sample t-test
##
## data:  .
## t = 15.424, df = 21, p-value = 6.276e-13
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  5.593283 7.336595
## sample estimates:
## mean of x
```

```
## 6.464939
##
## [1] "Redfield & Wilton Strategies"
##
##  One Sample t-test
##
## data:  .
## t = 3.4907, df = 7, p-value = 0.01012
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.160722 6.035520
## sample estimates:
## mean of x
##  3.598121
##
## [1] "Change Research"
##
##  One Sample t-test
##
## data:  .
## t = 6.0706, df = 6, p-value = 0.000907
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.742759 4.096340
## sample estimates:
## mean of x
##  2.919549
##
## [1] "Morning Consult"
##
##  One Sample t-test
##
## data:  .
## t = 8.2283, df = 6, p-value = 0.000174
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.62347 4.84420
## sample estimates:
## mean of x
##  3.733835
##
## [1] "Ipsos"
##
##  One Sample t-test
##
## data:  .
## t = 5.7168, df = 5, p-value = 0.00229
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.246420 3.283155
## sample estimates:
## mean of x
##  2.264787
##
## [1] "Data for Progress"
##
##  One Sample t-test
##
## data:  .
```

```
## t = 4.9514, df = 3, p-value = 0.01581
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   1.910686 8.785556
## sample estimates:
## mean of x
##   5.348121
##
## [1] "East Carolina University"
##
##  One Sample t-test
##
## data:  .
## t = 2.4056, df = 3, p-value = 0.0954
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -1.002143  7.208384
## sample estimates:
## mean of x
##   3.103121
##
## [1] "Emerson College"
##
##  One Sample t-test
##
## data:  .
## t = 2.6655, df = 3, p-value = 0.07598
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -0.2973393  3.3635807
## sample estimates:
## mean of x
##   1.533121
##
## [1] "Rasmussen Reports/Pulse Opinion Research"
##
##  One Sample t-test
##
## data:  .
## t = 5.5954, df = 3, p-value = 0.01128
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   0.1630602 0.5931813
## sample estimates:
## mean of x
## 0.3781207
##
## [1] "Harper Polling"
##
##  One Sample t-test
##
## data:  .
## t = 2.5222, df = 2, p-value = 0.1278
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -1.186981  4.549889
## sample estimates:
## mean of x
##   1.681454
```

```
##
## [1] "Siena College/The New York Times Upshot"
##
##  One Sample t-test
##
## data:  .
## t = 4.5523, df = 2, p-value = 0.04502
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.2202043 7.8093704
## sample estimates:
## mean of x
##  4.014787
##
## [1] "SurveyUSA"
##
##  One Sample t-test
##
## data:  .
## t = 1.8089, df = 2, p-value = 0.2122
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.156301 10.185875
## sample estimates:
## mean of x
##  3.014787
##
## [1] "Trafalgar Group"
##
##  One Sample t-test
##
## data:  .
## t = -2.4933, df = 2, p-value = 0.1302
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.2310865  0.5939945
## sample estimates:
## mean of x
## -0.818546
##
## [1] "YouGov"
##
##  One Sample t-test
##
## data:  .
## t = 7.0222, df = 2, p-value = 0.01968
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.813019 7.549889
## sample estimates:
## mean of x
##  4.681454
```

**SurveyMonkey**

SurveyMonkey conducted 21 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **6.464939%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 5.593% to 7.337%. The calculated t-statistic was 15.424, and the calculated p-value was 6.276 * 10^-13, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that SurveyMonkey polls in North Carolina differed systematically from the actual results.

### Redfield & Wilton Strategies

Redfield & Wilton Strategies conducted 8 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **3.598121%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 1.160722% to 6.035520%. The calculated t-statistic was 3.4907, and the calculated p-value was 0.01012, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Redfield & Wilton Strategies polls in North Carolina differed systematically from the actual results.

### Change Research

Change Research conducted 7 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **2.919549%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 1.742759% to 4.096340%. The calculated t-statistic was 6.0706, and the calculated p-value was 0.000907, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Change Research polls in North Carolina differed systematically from the actual results.

### Morning Consult

Morning Consult conducted 7 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **3.733835%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 2.62347% to 4.84420%. The calculated t-statistic was 8.2283, and the calculated p-value was 0.000174, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Morning Consult polls in North Carolina differed systematically from the actual results.

### Ipsos

Ipsos conducted 6 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **2.264787%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 1.25% to 3.28%. The calculated t-statistic was , and 5.7168the calculated p-value was 0.00229, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Ipsos polls in North Carolina differed systematically from the actual results.

### Data for Progress

Data for Progress conducted 4 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **5.348121%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 1.910686% to 8.785556%. The calculated t-statistic was 4.9514, and the calculated p-value was 0.01581, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Data for Progress polls in North Carolina differed systematically from the actual results.

### East Carolina University

East Carolina University conducted 4 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **3.103121%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -1.00% to 7.21%. The calculated t-statistic was 2.41, and the calculated p-value was 0.0954, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that East Carolina University polls in North Carolina differed systematically from the actual results.

### Emerson College

Emerson College conducted 4 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **1.533121%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -0.2973393% to 3.3635807%. The calculated t-statistic was 2.6655, and the calculated p-value was 0.07598, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Emerson College polls in North Carolina differed systematically from the actual results.

### Rasmussen Reports/Pulse Opinion Research

Rasmussen Reports/Pulse Opinion Research conducted 4 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **0.3781207%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 0.1630602% to 0.5931813%. The calculated t-statistic was 5.5954, and the calculated p-value was 0.01128, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Rasmussen Reports/Pulse Opinion Research polls in North Carolina differed systematically from the actual results — albeit by a very small margin.

### Harper Polling

Harper Polling conducted 3 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **1.681454%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -1.187% to 4.550%. The calculated t-statistic was 2.5222, and the calculated p-value was 0.1278, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Harper Polling polls in North Carolina differed systematically from the actual results.

### Siena College / The New York Times Upshot

Siena College / The New York Times Upshot conducted 3 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **4.014787%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 0.2202% to 7.8094%. The calculated t-statistic was 4.5523, and the calculated p-value was 0.04502, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that Siena College / The New York Times Upshot polls in North Carolina differed systematically from the actual results.

### SurveyUSA

SurveyUSA conducted 3 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **3.014787%** more positive for Biden than the actual final result. The 95% confidence interval for the error is -4.156% to 10.186%. The calculated t-statistic was 1.8089, and the calculated p-value was 0.2122, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that SurveyUSA polls in North Carolina differed systematically from the actual results.

### Trafalgar Group

Trafalgar Group conducted 3 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **0.818546%** more negative for Biden than the actual final result. The 95% confidence interval for the error is -2.2310865% to 0.5939945%. The calculated t-statistic was -2.4933, and the calculated p-value was 0.1302, which is greater than 0.05.

Because of this, we do not have sufficient evidence to reject the null hypothesis, meaning we do not have sufficient evidence to say that Trafalgar Group polls in North Carolina differed systematically from the actual results.

### YouGov

YouGov conducted 3 polls in North Carolina from August 2020 to November 2020. The average error of those polls was **4.681454%** more positive for Biden than the actual final result. The 95% confidence interval for the error is 1.813% to 7.550%. The calculated t-statistic was 7.0222, and the calculated p-value was 0.01968, which is less than 0.05.

Because of this, we have sufficient evidence to reject the null hypothesis, and to say that we are 95% confident that YouGov polls in North Carolina differed systematically from the actual results.

We compiled our results below into a table, and then plotted the estimated error on the y axis, pollster on the x axis, and whether we were able to reject the null with 95% confidence or not with either a light red (false) or light blue (true), arranged in ascending order by error.

```
 nc_results <- compare_both %>%
    filter(state == "north carolina") %>%
    filter(date > '2020-08-01') %>%
   filter(pollster %in% top_pollsters_NC) %>%
    inner_join(filtered_results, by = "state") %>%
    mutate(each_error = poll_margin - actual_margin) %>%
    group_by(pollster) %>%
  do(tidy(t.test(.$each_error)))

nc_results %>%
  arrange(abs(estimate)) %>%
  select(pollster, estimate, p.value, conf.low, conf.high) %>%
  mutate(reject95 = p.value < 0.05) %>%
  print() %>%
  arrange(abs(estimate)) %>%
  ggplot(aes(x = reorder(pollster, abs(estimate)), y = estimate, fill = reject95,
             ymin = conf.low, ymax = conf.high)) + geom_col() + theme_linedraw() +
   labs(x = "Pollster", y = "Estimated Error", fill = "Reject null?", title = "Polling Error in North Carol
    scale_x_discrete(guide = guide_axis(n.dodge = 5)) +geom_errorbar()
```

```
## # A tibble: 14 x 6
## # Groups:   pollster [14]
##    pollster                       estim~1  p.value conf.~2 conf.~3 rejec~4
##    <chr>                            <dbl>    <dbl>   <dbl>   <dbl> <lgl>
##  1 Rasmussen Reports/Pulse Opinion Res~  0.378 1.13e- 2   0.163   0.593 TRUE
##  2 Trafalgar Group                -0.819 1.30e- 1  -2.23    0.594 FALSE
##  3 Emerson College                 1.53  7.60e- 2  -0.297   3.36  FALSE
##  4 Harper Polling                  1.68  1.28e- 1  -1.19    4.55  FALSE
##  5 Ipsos                           2.26  2.29e- 3   1.25    3.28  TRUE
##  6 Change Research                 2.92  9.07e- 4   1.74    4.10  TRUE
##  7 SurveyUSA                       3.01  2.12e- 1  -4.16   10.2   FALSE
##  8 East Carolina University        3.10  9.54e- 2  -1.00    7.21  FALSE
##  9 Redfield & Wilton Strategies    3.60  1.01e- 2   1.16    6.04  TRUE
## 10 Morning Consult                 3.73  1.74e- 4   2.62    4.84  TRUE
## 11 Siena College/The New York Times Up~  4.01  4.50e- 2   0.220   7.81  TRUE
## 12 YouGov                          4.68  1.97e- 2   1.81    7.55  TRUE
## 13 Data for Progress               5.35  1.58e- 2   1.91    8.79  TRUE
## 14 SurveyMonkey                    6.46  6.28e-13   5.59    7.34  TRUE
## # ... with abbreviated variable names 1: estimate, 2: conf.low, 3: conf.high,
## #   4: reject95
```
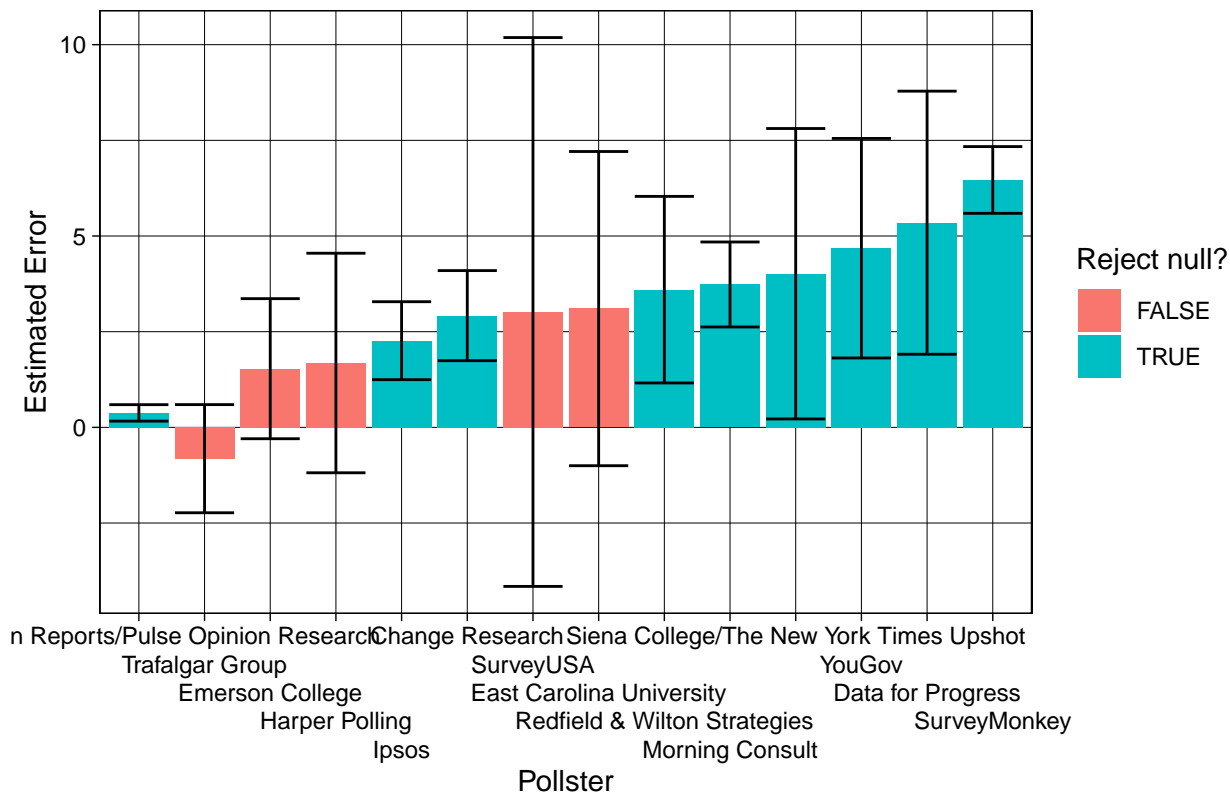
## Polling Error in North Carolina during the 2020 presidential election



(Combine analyses to provide see which pollsters were most accurate nationwide)

We were able to reject the null hypothesis (that their error was not equal 0) with 95% confidence for 9 pollsters, and failed to reject for 5: Trafalgar Group, Emerson College, Harper Polling, SurveyUSA, and East Carolina University However, this does not mean that the pollsters for which we did not reject the null hypothesis are accurate; absence of evidence of systematic error is not the same as evidence for accuracy.

Compared to Georgia, polls here appeared to sway farther from the actual results, and most error was in Biden's direction, rather than a mix of both. While error was not extreme on average compared to Wisconsin or Pennsylvania, there was still more error than a state like Georgia. Most polling errors tended in the direction of overestimating Biden's chance of victory; while some pollsters predicted up to a 6.5% Biden victory, Biden ended up losing the state by around 1.35%. Among the five states we've analyzed, North Carolina was the only state out of the five we analyzed where the error was large enough that the predicted victor by many pollsters (Biden) did not end up winning.

The three most accurate pollsters from our analysis are: Rasmussen Reports / Pulse Opinion Research (+0.378%), Trafalgar Group (-0.819%), and Emerson College (+1.53%). The three most inaccurate pollsters from our analysis are: YouGov (+4.68%), Data for Progress (+5.35%), and SurveyMonkey (+6.46%).

## 4.6 Compiling the top 5 most competitive states

For our final analysis, we will be combining poll errors from all five of the closest states: Arizona, Georgia, Wisconsin, Pennsylvania, and North Carolina. We will be looking at pollsters who have conducted at least 1 poll in EACH of the states from August 2020 to Election Day. The process is similar to the previous sections, but we have commented our code below:

```
# Our step by step process will:
# 1. Collect all pollsters who ran at least 1 poll in Arizona. Store that.
# 2. Run through EACH pollster to check that they ran at least 1 poll in Georgia,
# Wisconsin, Pennslyvania, & North Carolina - states listed in xy.
# Remove those that do not exist in 1 or more states.
# 3. Then, with our compressed list of top pollsters, we will run a t-test on errors
```

```r
# among all 5 states and report our results.

#xy contains the top 5 states we will analyze, except for Arizona.
xy <- c("georgia", "wisconsin", "pennsylvania", "north carolina")

func_ttest <- function(des_pollster, des_state) {
  compare_both %>% #dataframe of all of our polls
    # filter for only our designated pollster and state
    filter(pollster == des_pollster, state == des_state) %>%
    # filter for polls conducted after August 1st, 2020
    filter(date > '2020-08-01') %>%
    # join the each poll margin results with the actual margin of victory of either
    # Biden / Trump in the state
    inner_join(filtered_results, by = "state") %>%
    # calculate EACH poll's error by subtracting the predicted poll margin (poll_margin) by
    # the actual winning margin (actual_margin)
    mutate(each_error = poll_margin - actual_margin) %>%
    # we are only interested in performing a t-test on the errors,
    # so we select it, pull it, and run a t test on it.
    select(each_error) %>%
    pull() %>%
    t.test(conf.level = 0.95)
}

# identify top pollsters in Arizona
top_pollsters <- compare_both %>%
  filter(state %in% c("arizona")) %>%
  filter(date > '2020-08-01') %>%
  group_by(pollster) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  select(pollster) %>%
  pull()

# create a vector to specify which pollsters to remove
to_remove <- c("")

# loop through each pollster
for(p in top_pollsters) {
  # loop through each state outside of AZ
  for(s in xy) {
    temp3 <- compare_both %>%
      filter(state == s) %>%
      filter(date > '2020-08-01') %>%
      filter(pollster == p) %>%
      summarize(count = n()) %>%
      pull()
    # do they have at least 1 poll in this state? if so, go into the if statemenet
    if(temp3 == 0) {
      # add that pollster to be removed
      to_remove <- append(to_remove, p)
    }
  }
}

#remove all pollsters that are in to_remove. there can be duplicates in to_remove, but since
#once a pollster is removed, it is gone forever, we are not too worried about the extra
# runtime.
```

```
top_pollsters <- top_pollsters[! top_pollsters %in% to_remove]


#then run a t-test among all top 5 states among the remaining pollsters!
nat_results <- compare_both %>%
    filter(state %in% xy) %>%
    filter(date > '2020-08-01') %>%
   filter(pollster %in% top_pollsters) %>%
    inner_join(filtered_results, by = "state") %>%
    mutate(each_error = poll_margin - actual_margin) %>%
    group_by(pollster) %>%
 do(tidy(t.test(.$each_error))) %>%
  arrange(abs(estimate)) %>%
  select(pollster, estimate, statistic, p.value, conf.low, conf.high) %>%
  mutate(reject95 = p.value < 0.05)

nat_results
```

```
## # A tibble: 9 x 7
## # Groups:   pollster [9]
##   pollster                    estim~1 stati~2  p.value conf.~3 conf.~4 rejec~5
##   <chr>                         <dbl>   <dbl>    <dbl>   <dbl>   <dbl> <lgl>
## 1 AtlasIntel                    -1.02   -1.18 3.23e- 1   -3.77    1.73 FALSE
## 2 Emerson College                2.53    3.04 1.25e- 2   0.673    4.38 TRUE
## 3 YouGov                         4.21    6.39 1.21e- 5    2.81    5.61 TRUE
## 4 Siena College/The New York T~  4.93    4.73 8.07e- 4    2.60    7.25 TRUE
## 5 Redfield & Wilton Strategies   5.04    7.54 6.77e- 8    3.66    6.41 TRUE
## 6 SurveyMonkey                   5.53   13.6  2.69e-23    4.72    6.34 TRUE
## 7 Swayable                       5.59    4.79 1.98e- 3    2.83    8.35 TRUE
## 8 Garin-Hart-Yang Research Gro~  5.82    4.54 1.05e- 2    2.26    9.37 TRUE
## 9 Morning Consult                6.43   17.7  6.62e-27    5.71    7.16 TRUE
## # ... with abbreviated variable names 1: estimate, 2: statistic, 3: conf.low,
## #   4: conf.high, 5: reject95
```

As shown above, there are 9 pollsters who have run at least 1 poll in the top 5 most competitive states (Georgia, Arizona, Wisconsin, Pennsylvania, and North Carolina) between August 2020 and November 2020. Our null hypothesis was that there was no error in a pollsters' results in the 5 competitive states, and our alternate was that there was some error.

For all of the pollsters EXCEPT ONE, we were able to reject the null hypothesis (their p.value was less than 0.05) and say that we are confident that those pollsters' results differed systematically from the actual elections results. For the pollster we did not reject — a failure to reject is not a confirmation of its accuracy. A failure to reject could due to be a low number of polls (AtlasIntel only had 1 poll in each top state — albeit those polls were relatively close to the actual results), or variation between polls. We will take caution when we are presenting these findings.

From lowest to greatest absolute value of error (+ values overestimated Biden, - values overestimated Trump): AtlasIntel (-1.02%); Emerson College (+2.53%), YouGov (+4.21%), Siena College / The New York Times Upshot (+4.93%), Redfield & Wilton Strategies (+5.04%), SurveyMonkey (+5.53%), Swayable (+5.593%), Garin-Hart-Yang Research Group (+5.8170%), and Morning Consult (+6.43%). The full data and confidence intervals can be found in the table above.

We compiled our results below into a table, and then plotted the estimated error on the y axis, pollster on the x axis, and whether we were able to reject the null with 95% confidence or not with either a light red (false) or light blue (true), arranged in ascending order by error. As the graph clearly shows, most error went towards in the direction of Biden, but as we noted earlier, in only one state was the margin of victory small enough that that error was enough to flip the predicted winner of the state — North Carolina.

```
nat_results %>%
  arrange(abs(estimate)) %>%
```

```
  select(pollster, estimate, p.value, conf.low, conf.high) %>%
  mutate(reject95 = p.value < 0.05) %>%
  print() %>%
  arrange(abs(estimate)) %>%
  ggplot(aes(x = reorder(pollster, abs(estimate)), y = estimate, fill = reject95,
          ymin = conf.low, ymax = conf.high)) + geom_col() + theme_linedraw() +
  labs(x = "Pollster", y = "Estimated Error", fill = "Reject null?", title = "Polling Error in top states
  scale_x_discrete(guide = guide_axis(n.dodge = 5)) +geom_errorbar()
```

```
## # A tibble: 9 x 6
## # Groups:   pollster [9]
##   pollster                        estim~1  p.value conf.~2 conf.~3 rejec~4
##   <chr>                             <dbl>    <dbl>   <dbl>   <dbl> <lgl>
## 1 AtlasIntel                        -1.02 3.23e- 1   -3.77    1.73 FALSE
## 2 Emerson College                    2.53 1.25e- 2   0.673    4.38 TRUE
## 3 YouGov                             4.21 1.21e- 5    2.81    5.61 TRUE
## 4 Siena College/The New York Times Ups~  4.93 8.07e- 4    2.60    7.25 TRUE
## 5 Redfield & Wilton Strategies       5.04 6.77e- 8    3.66    6.41 TRUE
## 6 SurveyMonkey                       5.53 2.69e-23    4.72    6.34 TRUE
## 7 Swayable                           5.59 1.98e- 3    2.83    8.35 TRUE
## 8 Garin-Hart-Yang Research Group     5.82 1.05e- 2    2.26    9.37 TRUE
## 9 Morning Consult                    6.43 6.62e-27    5.71    7.16 TRUE
## # ... with abbreviated variable names 1: estimate, 2: conf.low, 3: conf.high,
## #   4: reject95
```



Polling Error in top states during the 2020 presidential election
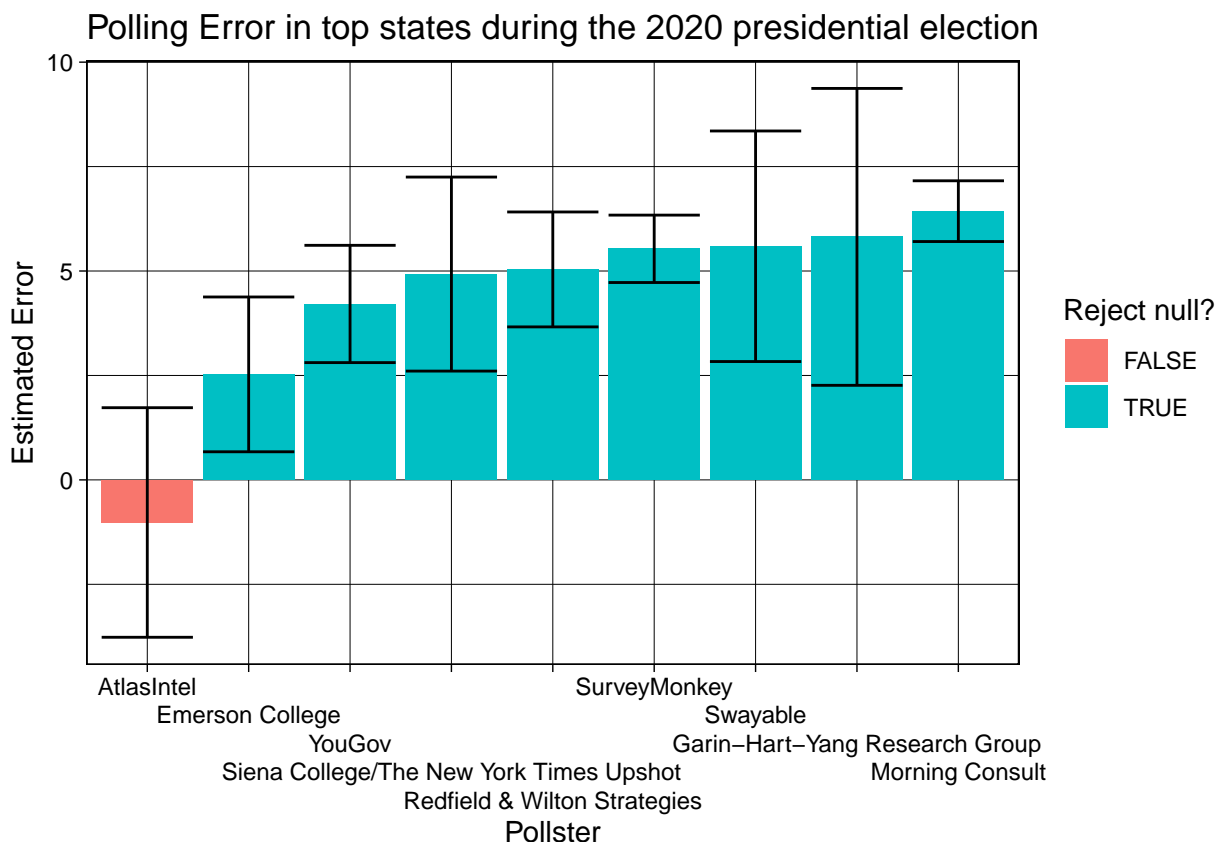
Overall, among the top 5 most competitive states, the amount of error varied between pollsters — to an extent. We
are unfortunately unable to directly compare some pollsters to one another, often because their confidence intervals
overlap, but there are other pollsters — such as AtlasIntel and Morning Consult — whose confidence intervals
clearly do not overlap, so, for instance, we can say with 95% confidence Morning Consult's polls had systematically
more error than AtlasIntel's.

But what might cause these differences in error among these different pollsters? Unfortunately, the data set we were provided is limited in terms of each pollster's actual methodology and sample demographics; however, we can make inferences about data found online.

## POLLSTERS FROM LOWEST TO HIGHEST ABSOLUTE ERRORS

*Please note that we included FiveThirtyEight pollster grades below, which can be found from their website. Those grades are not just based off of polling accuracy; they also are determined based on empirical analyses of each pollster's methodology, bias, and reputation. While this was not available through the spreadsheets we analyzed, we are including it as another potential data point from which we might potentially identify patterns.*

Starting with **AtlasIntel**, which appears to have the lowest absolute error, they advertise that they use **online polling** using " proprietary methodology", which is not particularly informative [11]. However, they claim that their random selection methods are well calibrated to minimize error, which appears to have worked to an extent. FiveThirtyEight's internal analysis confirms this, giving AtlasIntel an **A** in terms of polling accuracy, with 75% race calling accuracy [12].

**Emerson College** has the next lowest absolute error. Notably, Emerson College is the only pollster on our list that is operated out of a university. They use a completely automated system; first, they collect a panel of participants **ONLINE** to gather information, and then they randomly **CALL** those selected participants [13]. They note how this may reach more voters who have inaccurate or nonexistent contact information, especially those who may not be targeted by traditional pollsters. Being the second lowest pollster in terms of error — albeit, like the others, having a significant confidence interval — it appears those methods may have worked. FiveThirtyEight's analysis confirms this, giving Emerson College an **A-** in terms of polling accuracy, with 76% race calling accuracy [14].

**YouGov** is the next pollster. YouGov commissions their polls completely **ONLINE**, but they use a relatively unique method. YouGov's polls are not available to the public; instead, they randomly sample a small subset of their own members / respondents, which they believe is representative of the voting population of each state - known as Active Sampling. Some of their polls are tracking polls - meaning they track the opinions of the same sample over time — while others are not. YouGov also has software installed on some participants' computers and phones which tracks their viewing history in addition to asking questions about presidential polls at random times to maximize participation and answering rates [15]. FiveThirtyEight gives YouGov a **B+** in terms of polling accuracy, with 89% of races called correctly [16].

**Siena College/The New York Times Upshot** is next. Nearly all of their polls are commissioned through the **LIVE PHONE CALLS**, meaning that the voters they contact all reach a real person asking them questions — not a bot. They collect numbers from a voter file purchased online, which contains cell phones and landline numbers that were provided by voters when they registered to vote. They note that, while they call random samples from those lists, they adjust based on response rate and likeliness to vote (as voting history is public information), and subset by age, race, region, party, gender, and education [17]. FiveThirtyEight gives Siena College / The New York Times Upshot a **A+** in terms of polling accuracy, with 73% of races called correctly [18].

**Redfield & Wilton Strategies** is next. Their U.S. polls are all commissioned **ONLINE**, but unfortunately, they are vague in terms of their exact methodology & sampling decisions. They are primarily European-focused, but we are unable to find exact details publicly available, and they do not publish full polling data for free [19]. FiveThirtyEight gives Redfield & Wilton Strategies a provisional grade (due to a low number of polls) of **B/C** in temrs of polling accuracy, with 78% of races called correctly [20]. For some reason, some of the polls listed in the original FiveThirtyEight data set — such as all of the Wisconsin polls — are not listed on FiveThirtyEight's actual website, so we will have to take their grade with some caution.

**SurveyMonkey** is next, and is probably the most well known as they are a form creation platform available to the general public. What is less well known is their political side, which is commissioned completely **ONLINE**. They target people who have just completed any SurveyMonkey form; after they complete it, some are randomly

---

[11]https://ropercenter.cornell.edu/atlasintel

[12]https://projects.fivethirtyeight.com/pollster-ratings/atlasintel/

[13]https://emerson.edu/academics/academic-departments/communication-studies/emerson-college-polling

[14]https://projects.fivethirtyeight.com/pollster-ratings/emerson-college/

[15]https://polsci.umass.edu/research/umass-poll/methodology

[16]https://projects.fivethirtyeight.com/pollster-ratings/yougov/

[17]https://www.nytimes.com/2018/09/06/upshot/live-poll-method.html

[18]https://projects.fivethirtyeight.com/pollster-ratings/siena-college-the-new-york-times-upshot/

[19]https://redfieldandwiltonstrategies.com/about-us/

[20]https://projects.fivethirtyeight.com/pollster-ratings/redfield-wilton-strategies/

invited to participate in a political poll. They adjust it to ensure the sample is representative, based on gender, race, education, etc [21]. FiveThirtyEight gives SurveyMonkey a grade of **C**, with 88% of races called correctly [22].

**Swayable** is next. Their polls are commissioned through **MOBILE AND INTERNET ADS**, where they run ads advertising people to take their survey on popular apps and websites in exchange for a non-monetary reward. They stratify their sample based on demographics - such as age, ethnicity, gender, education, and geography [23]. FiveThirtyEight gives Swayable a grade of **C**, with 89% of races called correctly [24].

**Garin-Hart-Yang Research Group** is next. Their polls are all commissioned through **LIVE PHONE CALLS**, similar to Siena College / The New York Times Upshot. One difference is that, while some of their polls are done through samples from voter files, others are done through random-digit dialing, where they just call random phone numbers, check to see if they are a voter, and ask them questions [25]. They aim to make their samples representative by stratifiying for demographic variables — much like the other pollsters. FiveThirtyEight gives Garin-Hart-Yang Research Group a grade of **B-**, with 81% of races called correctly [26].

Finally, the pollster with the highest averaged error among the top 5 states we've selected is **Morning Consult**. They commission all of their polls **ONLINE**, by sending surveys to "vendors", such as Lucid and PureSpectrum, who place polls on websites to be taken by random Americans. They then weight their results by demographic variables. Unfortunately, no more information is provided on how exactly those vendors distribute surveys, but we can infer that they attempt to gather a random sample from online website visitors, while trying to weight for the fact that not everyone visits those particular websites or has access to the internet [27]. FiveThirtyEight gives Morning Consult a grade of **B**, with 83% of races called correctly [28].

Overall, there are clear differences between pollsters' accuracy and their methodology for sampling voters. Some use traditional phone calls, some use online methods, and others use a combination of both. It seems like pollsters who implement more targeted and creatively-chosen samples — such as by using a combination of phone calls and online polling or specific software to collect data — may have done better in terms of overcoming some of the demographic issues mentioned earlier with traditional polling. While we can rank each pollsters' average errors with one another, we cannot definitively say that one is "more accurate" than another because of the high amount of variance between them.

In addition, there do appear to be differences in error among states. While that was not the target of this analysis, we saw that certain states (like Georgia) had significantly less error compared others like (Pennsylvania and Wisconsin). This adds weight to the idea that possible error could originate from certain demographic traits of the voting population that are not fully captured in traditional polling — leading more nontraditional pollsters to perform better.

Again, we must emphasize that our analysis is limited to an extent due to variation and the limited information we are provided with, but we hope the above information could potentially shed some light on how / why different pollsters differ in terms of error.

# Part 5: Conclusion

Overall, yes, there were differences in polling accuracy in pollsters in terms of polling error among the most competitive swing states, although many of their effects are unclear due to high variance. Most of those errors went in Biden's direction, and one state's (North Carolina) error was large enough to flip the predicted winner of that election (Biden –> Trump).

We also saw differences between states in terms of polling error. Interestingly, states that were traditionally Republican (Arizona, Georgia) were the ones that had the most accurate polling for Biden, compared to traditional Democratic / union strongholds (Wisconsin, Pennsylvania), which had much larger errors among many pollsters.

Overall, polls had a high degree of variation, which is to be expected given that this was such a contentious election with close margins. Some pollsters were more accurate in some states, and less accurate than others. We identified

[21] https://www.surveymonkey.com/mp/survey-methodology/
[22] https://projects.fivethirtyeight.com/pollster-ratings/surveymonkey/
[23] https://www.swayable.com/polls/2020-10-28/
[24] https://projects.fivethirtyeight.com/pollster-ratings/swayable/
[25] https://hartresearch.com/our-story/
[26] https://projects.fivethirtyeight.com/pollster-ratings/garin-hart-yang-research-group/
[27] https://edchoice.morningconsultintelligence.com/methodology/
[28] https://projects.fivethirtyeight.com/pollster-ratings/morning-consult/

potential differences between pollsters in terms of how they collected results, but we cannot definitively say that those reasons are why certain pollsters are more accurate than others. However, it was interesting to see how different pollsters tried to tackle many of the demographic underrepresentation issues mentioned earlier — both to gather representative samples and minimize the rate of nonresponse to their surveys.

However, our analysis was limited by the fact that most pollsters don't publish their full methodology and poll / sample results online. In a future analysis with more resources, we would like to 1) analyze more states, possibly across multiple election years, 2) acquire full methodology and polling data from pollsters to understand their process in detail, understand shortfalls, and see potential differences in sample representation from one poll to another, and 3) perform deeper analyses on error beyond just margin of victory — were there other differences, such as in predicted turnout, the demographics of those who would turn out, etc?

On a whole, in the 2020 election, Joe Biden won by large enough margins that errors in polling didn't change the outcome for the election. And to an extent, polling error is natural, and pollsters are working to combat it constantly; for example, while we did not get a chance to analyze new data, many of the polls coming out of the recent 2022 midterms got accurate results — or estimated Republicans. Variation between polls is natural, but there is still more that needs to be analyzed to truly understand how to tackle polling error and ensure that polls accurately reflect the will of the people during our country's most pivotal times.