MAKE FORECASTING GREAT AGAIN: THREE LESSONS FROM THE 2016 U.S. PRESIDENTIAL ELECTION

Tangbufan Wei, Halilou Junior Tmite, Gaurav Daven

**Executive Summary**

Pre-election polls of the 2016 U.S. Presidential Election showed the Democratic candidate, Hillary Clinton, led the Republican candidate, Donald Trump, in the swing states that are crucial for the presidency. While all data pointed to her victory, Trump won the election by a landslide (306 electors vs 232 electors). Therefore, this report examines why the forecasts were inaccurate.

Using a representative sample of polls for the 2016 U.S. Presidential Election from FiveThirtyEight, we identify the following characteristics of the polls pertinent to this error.

Raw polls incorrectly favoured Clinton.

Too few quality pollsters in battleground states. The ill-identification of likely voters.

These three problems exacerbate for swing state (Florida, Michigan, Ohio, Pennsylvania and Wisconsin) forecasts.

These features reflect three main problems of polls for the 2016 U.S. Presidential Election. For each problem, we summarize the lessons and provide recommendations on how not to repeat the same mistake.

**Problems Recommendations**

**Unprecedented Election** More polls should be conducted near the election to capture the

preferences of undecided voters. Polling methodology should be designed to weigh polls conducted after major announcements affecting either candidate more heavily.

**Traditional Polling Method Unsuitable** Well-resourced survey institutions can collectively finance high-quality

state-level polls and encourage more quality pollsters to carry out state-level polls.

**The Johnson Effect** Our exercise shows that it is possible for Johnson, as a third candidate,

to affect the election outcome. Therefore, we recommend further studies on the John Effect.

This summary is accompanied by a data dictionary on the next page to aid reader’s understanding of the rationale and each variable of the data set from FiveThirtyEight. An interactive report is also available at <https://uottawa-statistics.shinyapps.io/report/> to facilitate readers’ understanding of this data set.

1Department of Mathematics and Statistics, University of Ottawa, Ottawa

**Email**: [htimi079@uottawa.ca [](mailto:htimi079@uottawa.ca)](mailto:htimi079@uottawa.ca)



# Data Dictionary

This data dictionary presents data items in alphabetical order by item names. Each data item contains a definition, item-specific codes, variable name and data type.

The variable **"Grade of the Pollster"** is included as it determines the quality and reliability of the pollster by Five Thirty Eight. Some pollsters have a grade of NA because FiveThirtyEight did not have enough information to assess their quality. **"Pollster"** variable is also included to analyze which organizations conducted the most number of polls and what proportion of them were in battleground states.

The **"Population"** variable helps determine the target population for each poll and if any conclusions can be reached from it.

The variable **"Raw Poll"** for the three candidates is in- cluded as it represents the raw, unadjusted poll results and can be directly compared with the adjusted poll results to see if poll adjustment affected reliability.

The variable **"Adjusted Poll"** for the three candidates represents the five adjustments applied to the **"Raw Poll"** variable by Five Thirty Eight. These are the likely voter adjustment, convention bounce adjustment, omitted third- party candidate adjustment, trend line adjustment and house effects adjustment. These adjustments are important in that they help address the issues associated with the polls and improve their credibility.

**Variables**

**Adjusted Poll: Clinton / Johnson / McMullin / Trump**

 **Definition**: the percentage for Hillary Clinton / Gary Johnson / Evan McMullin / Donald Trump from the poll, adjusted by FiveThirtyEight.

 **Codes**: between -3.66789 and 100, including NA

## **Name**: adjpoll\_clinton, adjpoll\_johnson, adjpoll\_mcmullin, adjpoll\_trump

 **Data type**: numeric

**Grade of the Pollster**

 **Definition**: the grade of the pollster assessed by FiveThirtyEight.

 **Codes**: in descending order, A+ (the best), A, A-, B+, B, B-, C+, C, C-, D (the worst), including NA

 **Name**: grade

 **Data type**: factor

**Pollster**

 **Definition**: the pollster’s name.

 **Codes**: 196 distinct organizations

 **Name**: pollster

 **Data type**: string

**Population**

 **Definition**: the type of voters being polled.

 **Codes**:

Notably, the sum of the raw and/or adjusted percent- ages received by all candidates in each poll may not equal

100. We treat the remaining percentage points to represent "undecided" interviewees.

* a: all adults
* lv: likely voters **Name**: population **Data type**: factor
* rv: registered voters
* v: voters

The **"Sample Size"** variable is one of the most important variables to include since a quality poll often has a large enough sample size that is representative of the underlying population the poll is trying to cover.

The variable **"Start Date*/*End Date"** is included to help in the analysis by only looking at polls close to major events surrounding the general election. These include party con- ventions, candidate debates, etc.

The **"State"** variable is useful in analyzing poll results in only battleground States which are often critical for both candidates in the general election. It can also be used to look at national polls which are generally well funded and comprehensive.

**Raw Poll: Clinton / Johnson / McMullin / Trump**

 **Definition**: the percentage for Hillary Clinton / Gary Johnson / Evan McMullin / Donald Trump as reported from the poll.

 **Codes**: between 0 and 100, including NA

## **Name**: rawpoll\_clinton, rawpoll\_johnson, rawpoll\_mcmullin, rawpoll\_trump

 **Data type**: numeric

**Sample Size**

 **Definition**: sample size of the poll.

 **Codes**: from 35 to 84292 (with 1 instance of NA)

 **Name**: samplesize

 **Data type**: Non-negative integer

**Start Date / End Date**

 **Definition**: the start date (or end date ) of the poll

 **Codes**:

* start date: dates from November 6th, 2015 to November 6th, 2016
* end date: dates from November 8th, 2015 to November 7th, 2016

 **Name**: startdate / enddate

 **Data type**: date (M/D/YYYY)

**State**

**Definition:** the U.S. state where the poll took place.

**Codes:** "U.S." indicates national polls.

* + Alabama
  + Alaska
  + Arizona
  + Arkansas
  + California
  + Colorado
  + Connecticut
  + Delaware
  + District of Columbia
  + Florida
  + Georgia
  + Hawaii
  + Idaho
  + Illinois
  + Indiana
  + Iowa
  + Kansas
  + Kentucky
  + Louisiana
  + Maine
  + Maine CD-1
  + Maine CD-2
  + Maryland
  + Massachusetts
  + Michigan
  + Minnesota
  + Mississippi
  + Missouri
  + Montana

**Name**: state

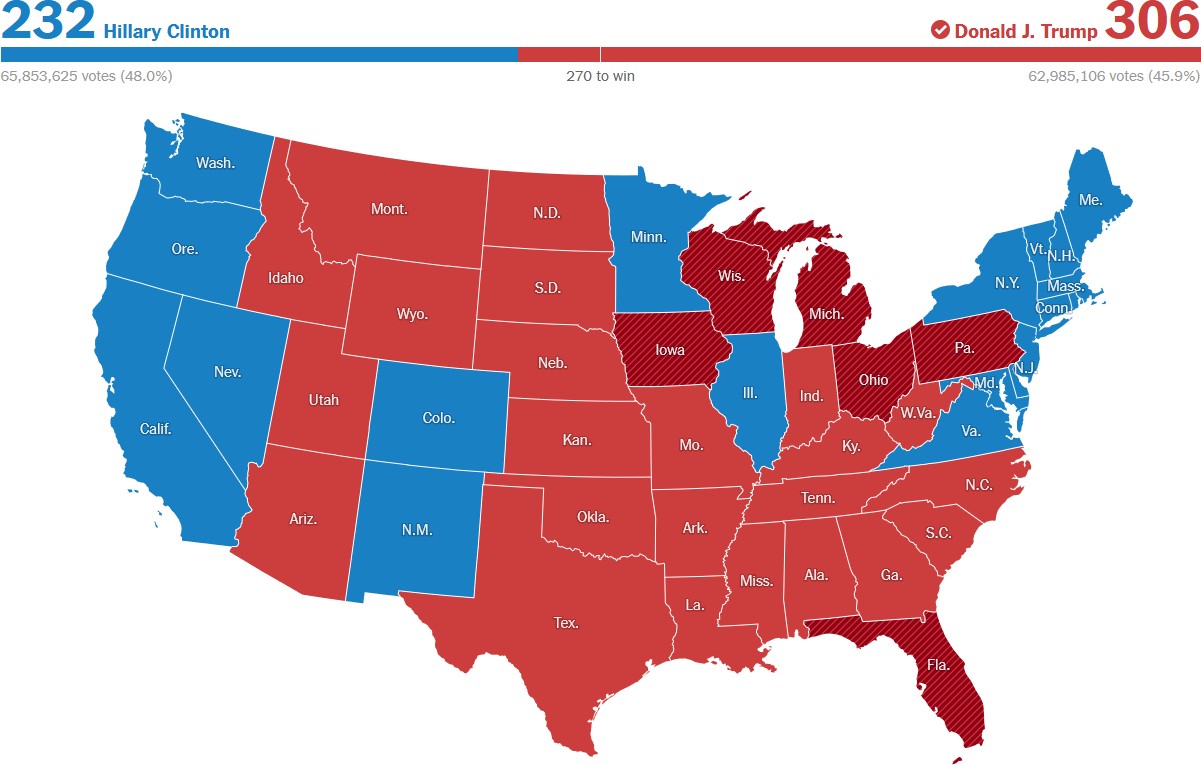
**Data type**: factor

* Nebraska
* Nebraska CD-1
* Nebraska CD-2
* Nebraska CD-3
* Nevada
* New Hampshire
* New Jersey
* New Mexico
* New York
* North Carolina
* North Dakota
* Ohio
* Oklahoma
* Oregon
* Pennsylvania
* Rhode Island
* South Carolina
* South Dakota
* Tennessee
* Texas
* U.S.
* Utah
* Vermont
* Virginia
* Washington
* West Virginia
* Wisconsin
* Wyoming

# Contents

1. [Introduction 5](#_TOC_250005)
2. [Background 5](#_TOC_250004)
3. [Exploratory Analysis 6](#_TOC_250003)
   1. Trump Had More Frequent Gain in Percentage of Polls after Adjustment. 6
   2. Post-adjustment Support for Trump is Lower than Clinton’s in Most Polls 6
   3. Ipsos was the Dominant Grade-A Pollster 6
   4. Only a Few Small-scale Pollsters Conducted Polls in the Swing States 6
   5. Swing States Have More Polls Conducted by Grade-A Pollsters than Non-Swing States 6
   6. Pollsters Attempted to Use Larger Sample Size to Correct for Methodological Weakness 7
   7. Most Polls are Targeted to Likely Voters 7
4. [Problems, Lessons and Recommendations 8](#_TOC_250002)
   1. Problem 1: An Unprecedented Election 8
   2. Problem 2: Traditional Polling Method Unsuitable 9
   3. Problem 3: The Johnson Effect 11
5. [Conclusion 12](#_TOC_250001)
6. [Bibliography 12](#_TOC_250000)

# 1. Introduction

This report uses a representative sample of polls for the 2016 U.S. Presidential Election from FiveThirtyEight to re- flect the stark contrast between the forecast and the turnout in reality. FiveThirtyEight used a weighted average of polls in this data set to forecast Clinton’s chance of winning to be over 70% (71.4%). However, in reality, the opponent, Trump, won by a landslide (306 electors vs. 232 electors, as in Figure 1). This drastic mistake was prevalent among all other political forecasts, while the actual popular vote (48.2%) was close to FiveThirtyEight’s predicted popular vote (46.1%).

**Figure 1.** Actual U.S. Presidential Election Results, 2016 For simplicity, this report considers groups of poll grades

instead of every individual. Specifically, we group grades A+, A and A- as grade A, grades B+, B and B- as grade B and grade NA as NA. We group the remaining grades as "lower than C+". Our analysis mainly focuses on the entire data set and polls associated with grade-A pollsters.

**Section 2: Background** provides an overview of the circumstances that lead to the result of the 2016 Presiden- tial Election. It highlights both candidates’ strengths and weaknesses as well as other events that impacted the result. **Section 3: Exploratory Analysis** helps analyze the polling results by looking at adjustment, characteristics of pollsters, number of polls, population type and also using these vari- ables to understand the results in swing states. **Section 4: Problems, Lessons and Recommendation** aims to de- fine the key issues affecting the reliability of polls, lessons learned and recommendations for future elections. Visual- izations are used throughout the report for the reader to better understand the analysis.

This report is accompanied by a data dictionary (af- ter the Executive Summary) to aid readers’ understanding of each variable in the data set. An interactive report is also available at [https://uottawa-statistics.](https://uottawa-statistics.shinyapps.io/report/) [shinyapps.io/report/](https://uottawa-statistics.shinyapps.io/report/) to facilitate readers’ under- standing of this data set.

# 2. Background

The 2016 U.S. Presidential Election was one of the most transformative political events in U.S. history that sought to change how polls are conducted. The Democratic party nominee was Hillary Clinton, a well-known establishment candidate was pitted against Donald Trump, the Republi- can nominee, a businessman and reality TV star with no previous experience in the field of politics or military. The Libertarian party candidate, Gary Johnson, was able to win 3.3% of the popular vote, the most in his party’s history, partly due to there being an unusually large number of undecided voters who viewed both Clinton and Trump as generally unlikeable and polarizing.

Trump portrayed himself as an outsider who could help fix what is broken with the Washington establishment, re- ferring to it as the "swamp". This message resonated with his voters due to bleak economic conditions, growing dis- content with the government and increasing political po- larization. Trump often referred to his primary voter base as the "forgotten people" who, in his opinion, were largely ignored by the establishment class. Trump used this narra- tive along with fear-mongering to energize his core base. In his presidential bid announcement speech, he labelled people coming from Mexico to the U.S. as “in many cases, criminals, drug dealers, rapists, etc.” and said that if elected, he would build a border wall with Mexico and make Mexico pay for it. Later, Trump also proposed a ban on Muslims from entering the U.S. Misinformation was frequently used by Trump to support his claims and it proved to be a bene- ficial strategy for him, with social media companies being ineffective in stopping the spread of misinformation.

Clinton, on the other hand, was a widely recognized political figure in the U.S. Her experience in politics in- cluded being the former First Lady, New York senator, and most recently, the Secretary of State. In addition to her political experience, Clinton was also the first woman to be a presumptive major-party nominee in U.S. history. The Clinton campaign, disillusioned by polling results and an in- experienced outsider rival candidate, failed to convey their message and outreach effectively in the battleground states. Her campaign failed to consider that she was campaigning to win a third consecutive term for the Democratic party, replacing a vastly more popular and charismatic predeces- sor in Barack Obama. Clinton’s candidacy, like Trump, was also shrouded in scandals and investigations. Her use of a private email server for official business was the subject of an ongoing FBI investigation and devastated her polling in the last days of the election when an announcement to re-open the investigation was made. No charges were filed but the damage was already done. The investigation domi- nated news headlines in the last crucial days and may have played a role in shifting undecided voters.

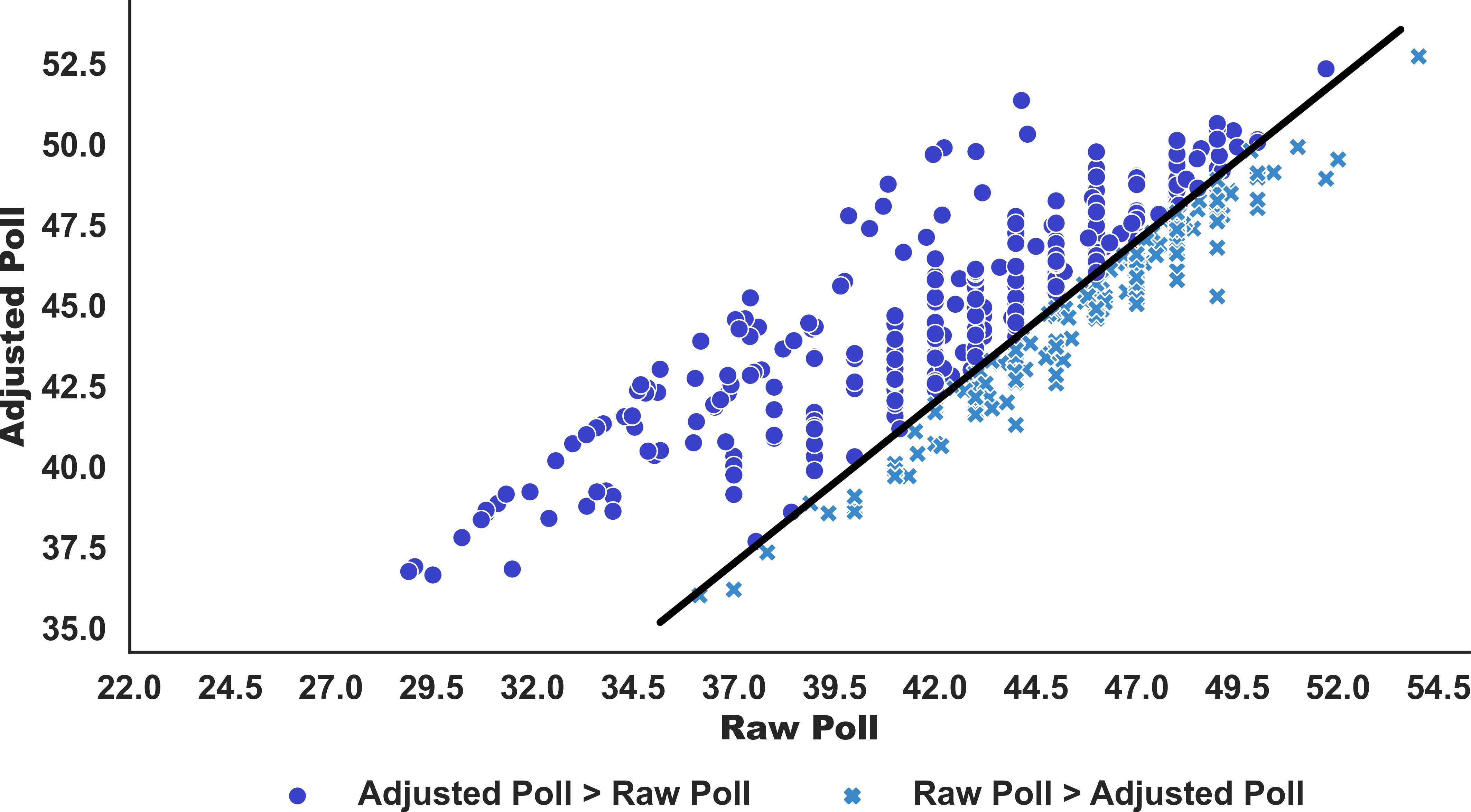
# 3. Exploratory Analysis

Through a series of data visualization, this section examines the failure of pollsters and the FiveThirtyEight adjustment to capture Trump’s lead in the swing states. We identify the lack of resourceful, quality pollsters for state-level polls and the ill-definition of most polls’ target population as causes of this huge deviation between the forecasted and the actual election outcome.

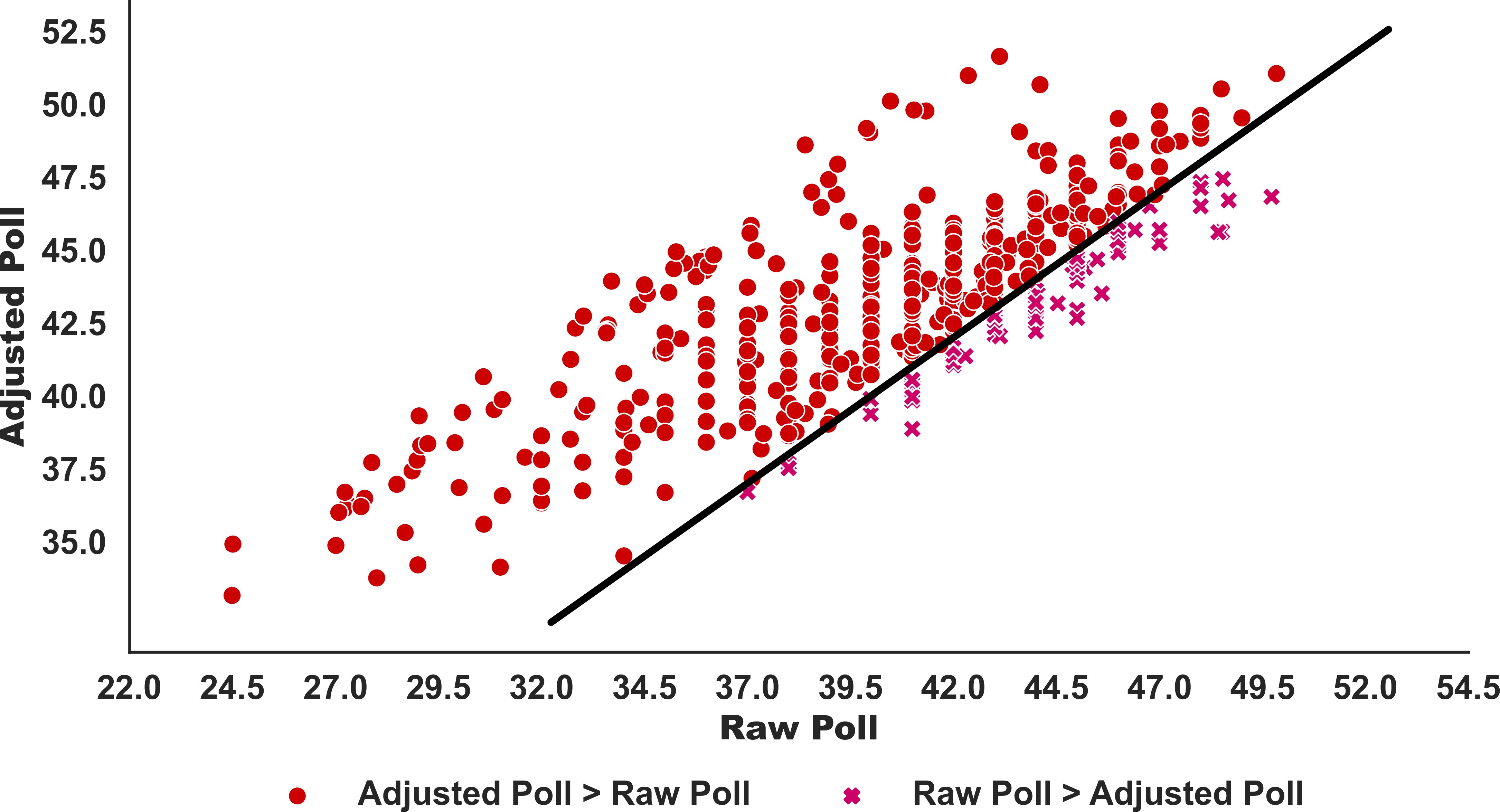
* 1. **Trump Had More Frequent Gain in Percentage of Polls after Adjustment.**

All data points on Figure 2 above each scatter plot’s 45- degree line correspond to polls with higher post-adjustment percentages than the reported. Many more points are above the line on panel (b) than panel (a). Specifically, 83% of the polls had Trump’s support raised after adjustment, while 62% of the polls had Clinton’s support increased post- adjustment. Hence, it is more frequent for FiveThirtyEight to raise Trump’s support in swing states.

1. **Clinton**



1. **Trump**



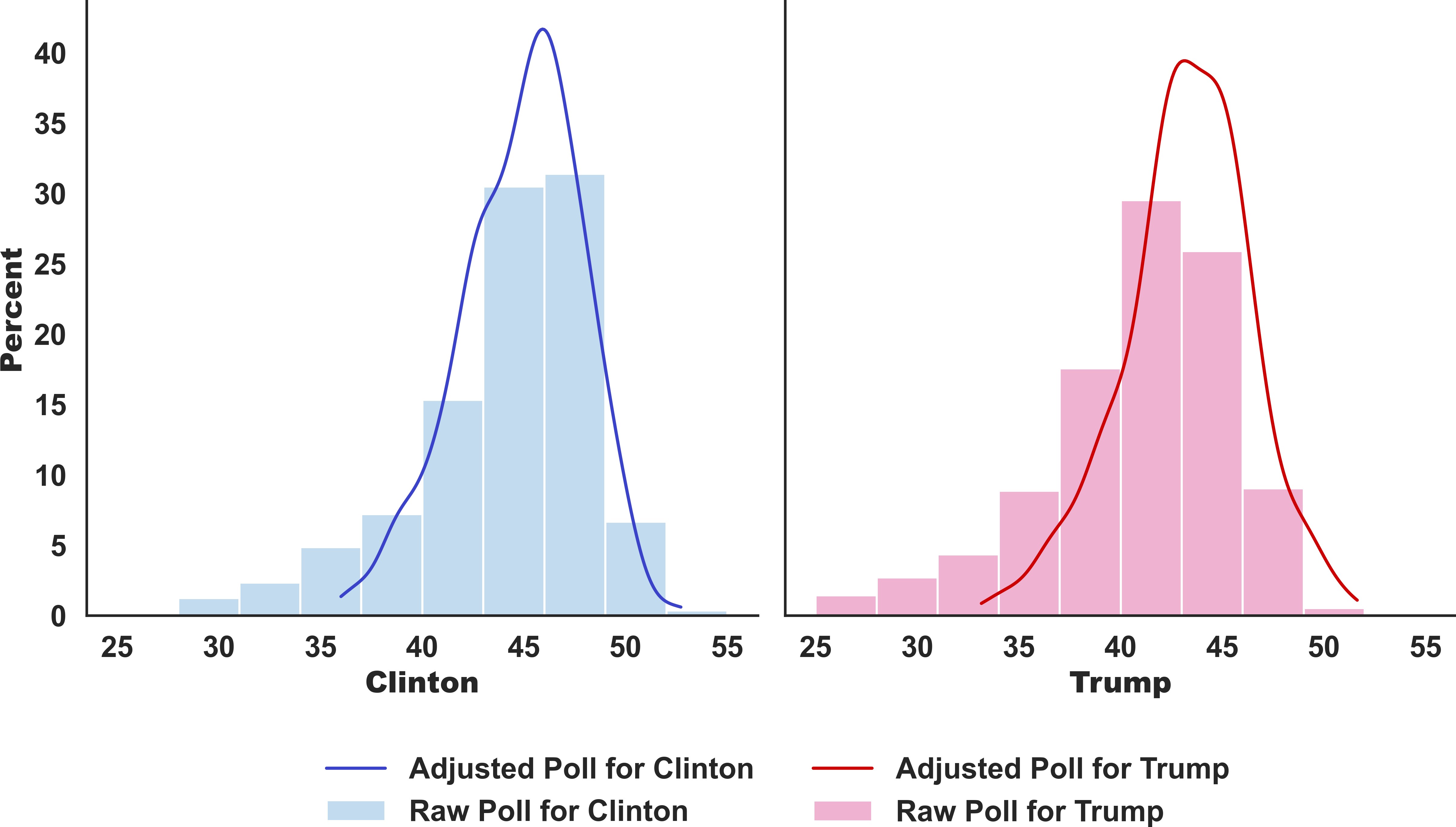
**Figure 2.** Raw Vs. Adjusted Polls,Clinton and Trump, All Grades, Swing States, November 12, 2015 to November 6,

2016

* 1. **Post-adjustment Support for Trump is Lower than Clin- ton’s in Most Polls.**

Despite Trump’s more frequent gain, the magnitude of Trump’s support after adjustment is lower than Clinton’s on average (Figure 3). Indeed, Clinton’s median shifted from

* + 1. **Clinton (b) Trump**



**Figure 3.** The Percentage of Raw and Adjusted Polls, Clinton and Trump, Swing States, All Grades, November 12, 2015 to November 6, 2016

### Remark 1

The lower magnitude of Trump’s support than Clinton explains why most election forecasts, including FiveThir- tyEight’s, predicted a landslide victory for Clinton in the 2016 U.S. Presidential Election. However, it is notice- able that FiveThirtyEight’s predicted chance of winning for Trump is higher than other forecasts. This distinction comes from the more frequent positive adjustment from FiveThirtyEight.

* 1. **Ipsos was the Dominant Grade-A Pollster.**

Among all grade-A pollsters, Ipsos provided a majority of the polls for the election (69%) (panel (a) of Figure 4). Other grade-A institutions carried out only a significantly smaller proportion of polls, no matter how large. In particular, the pollster with the second most polls (Quinnipiac University) only accounts for 5%. Notably, among all 196 pollsters, Ipsos also carried out the most polls (22%), followed by grade-B pollster Google Consumer Surveys (18%) (panel

(b) of Figure 4).1

* 1. **Only a Few Small-scale Pollsters Conducted Polls in the Swing States.**

Swing states have much fewer pollsters than the non-swing-

state average (Figure 5). Most of them are small institutions or universities. They have a tight budget and hence cannot easily adapt to elections. For example, the pollster who conducted the second most polls in Michigan, EPIC-MRA, is a small Michigan-based survey research firm.

* 1. **Swing States Have More Polls Conducted by Grade-A Pollsters than Non-Swing States.**

All swing states have more polls conducted by grade-A

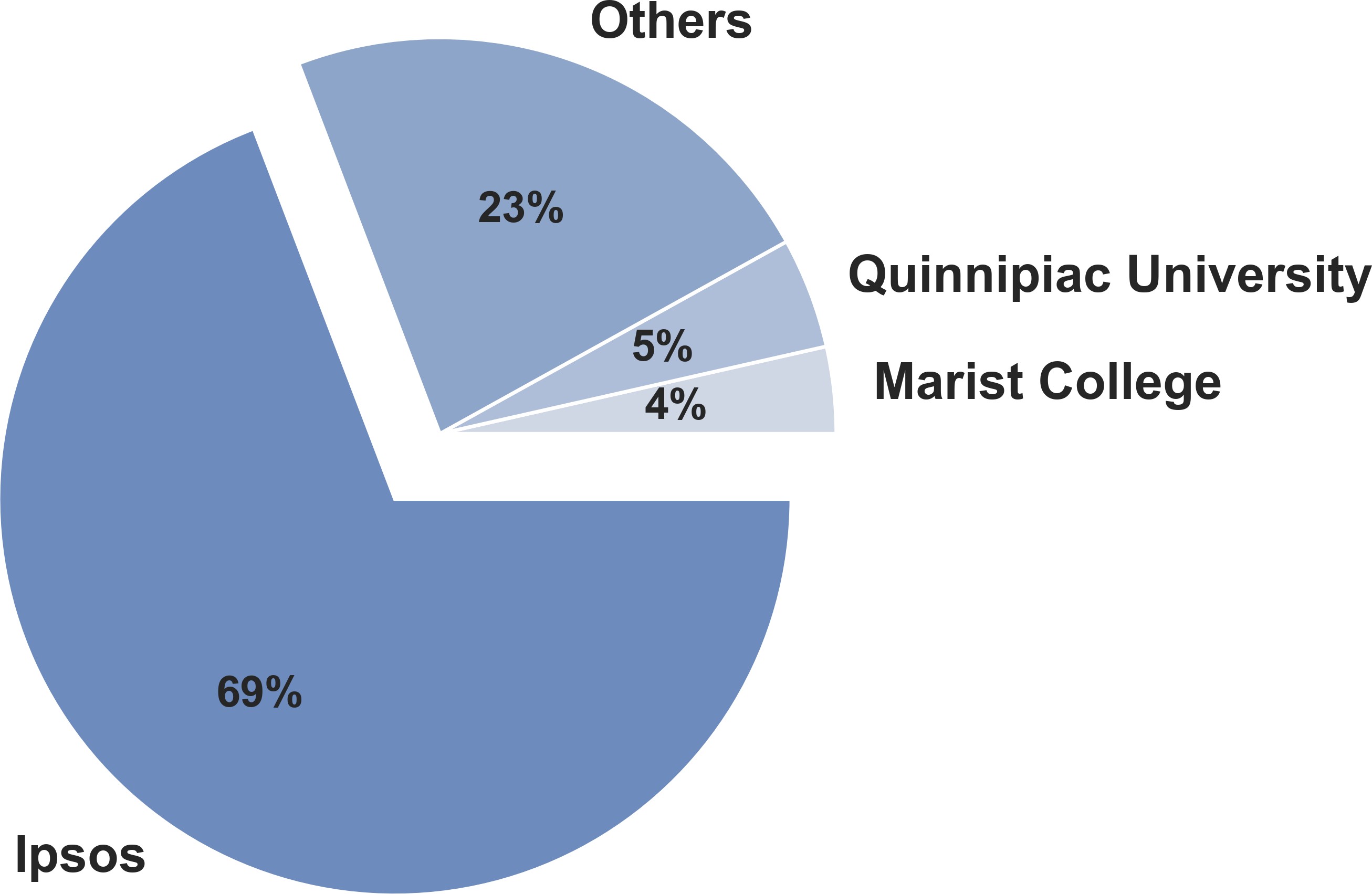
pollsters than the non-swing-state average (Figure 6). In particular, the numbers of polls in Florida and Pennsylvania

44.8% to 45.2% after adjustment, but that for Trump went

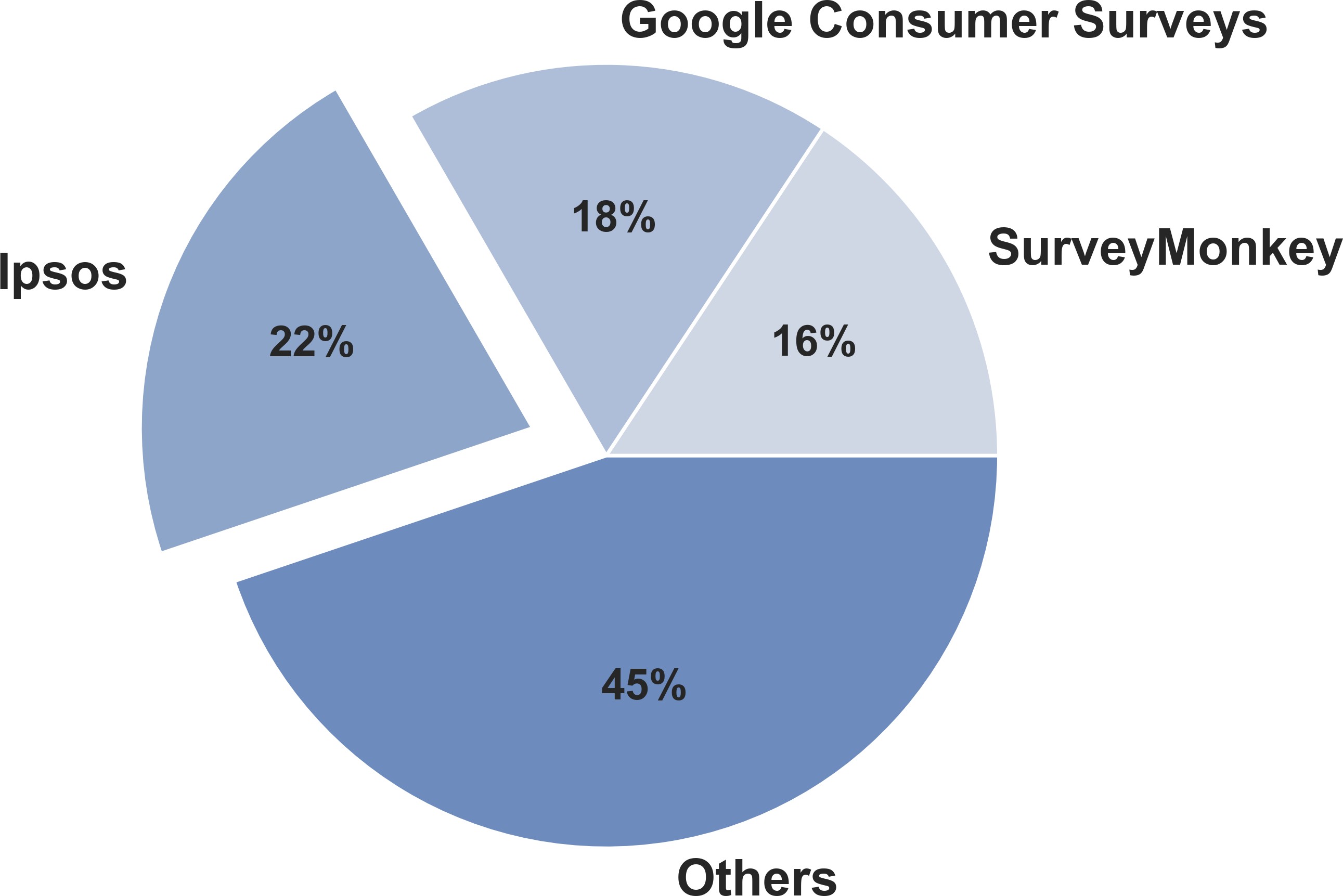
from 41.3% to 43.4%.

1Google will close Google Consumer Surveys on November 1st, 2022.

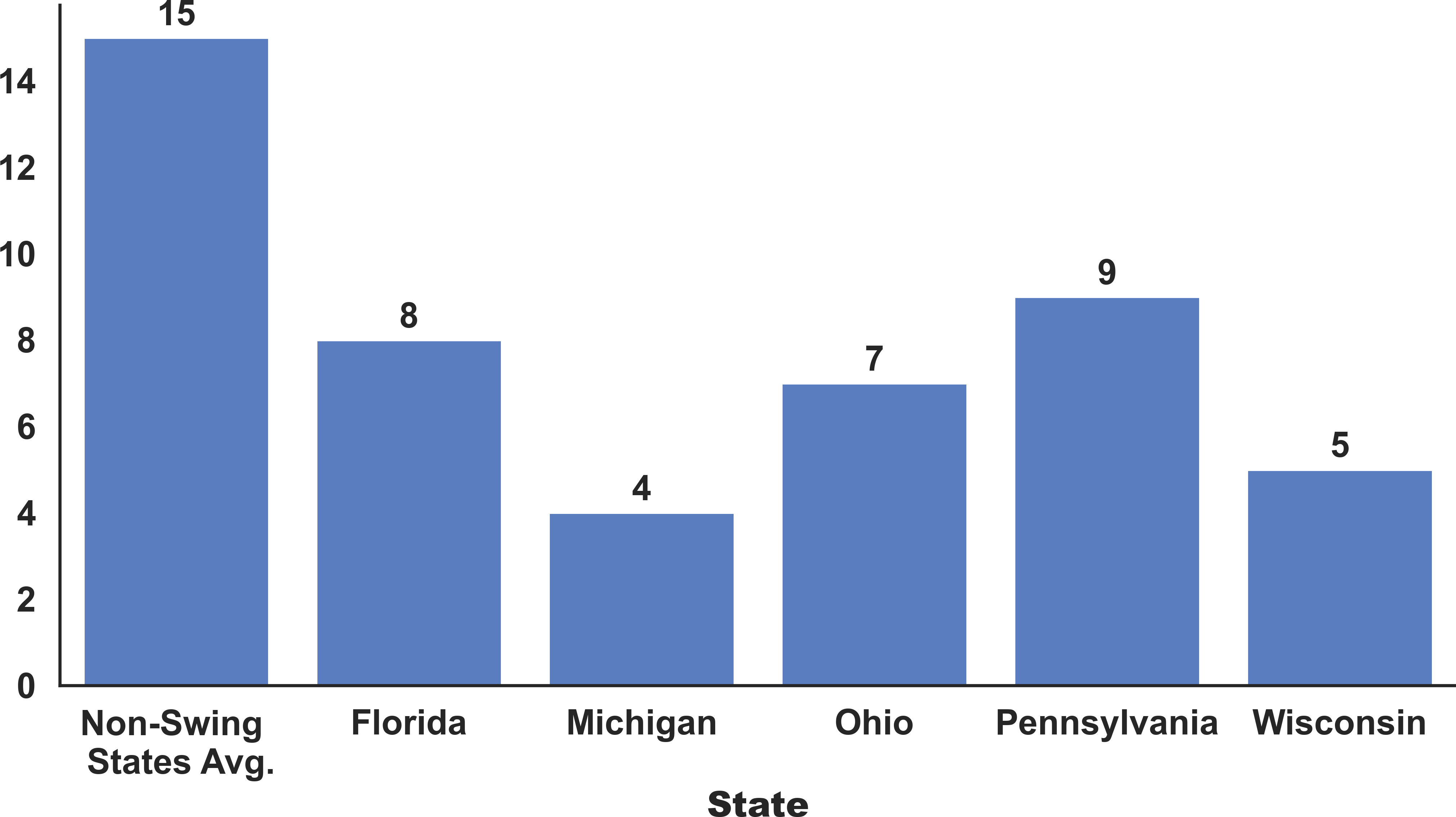
1. **Grade-A Pollsters**



1. **All Pollsters**

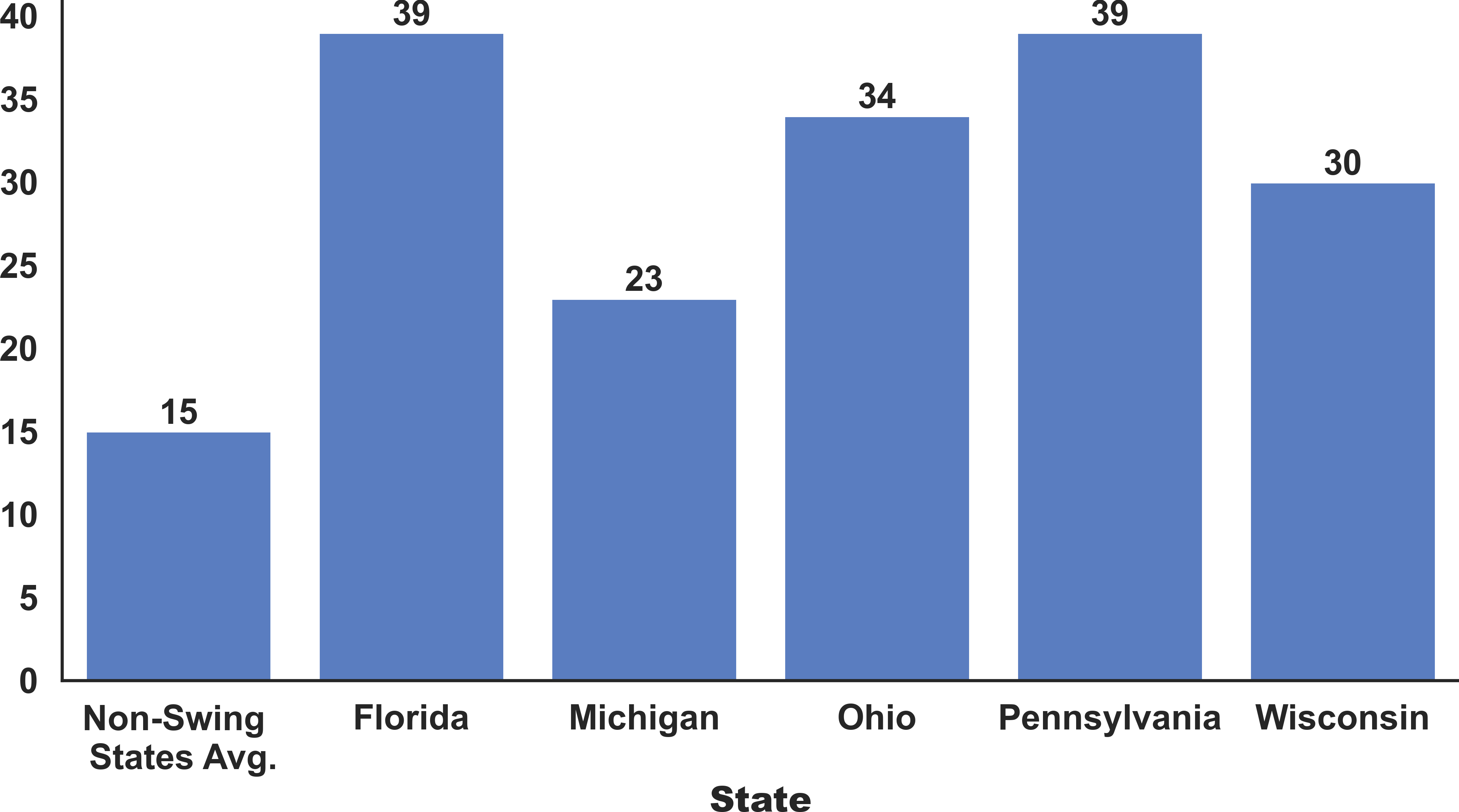


**Figure 4.** The Share of Polls from the Three Pollsters with Most Polls, Per Cent



**Figure 5.** The Number of Grade-A Pollsters, Swing States Vs. State-level Average Excluding Swing States, November 12, 2015 to November 6, 2016

are more than double the non-swing-state average (39 vs 15).



**Figure 6.** The Number of Polls Conducted by Grade-A Pollsters, Swing States Vs. Non-Swing State Average, November 12, 2015 to November 6, 2016

### Remark 2

Small pollsters attempt to compensate for their lack of adaptability to changes in the election through more fre- quent polls. Indeed, the two swing states with the fewest pollsters (Michigan and Wisconsin) have the highest poll- to-pollster ratio (5.75 and 6, respectively) compared with other swing states2 and the non-swing state average of 1. However, more frequent polls by small pollsters further tightened their budget and further weakened their ability to adapt to the unique evolution of this election.

* 1. **Pollsters Attempted to Use Larger Sample Size to Cor- rect for Methodological Weakness.**

Figure 7 demonstrates two tendencies regarding whether

the polls are for swing states and pollsters’ grades.

 Polls conducted in the swing states have larger sample sizes on average.

 Polls conducted by lower-grade pollsters have larger sample sizes on average.

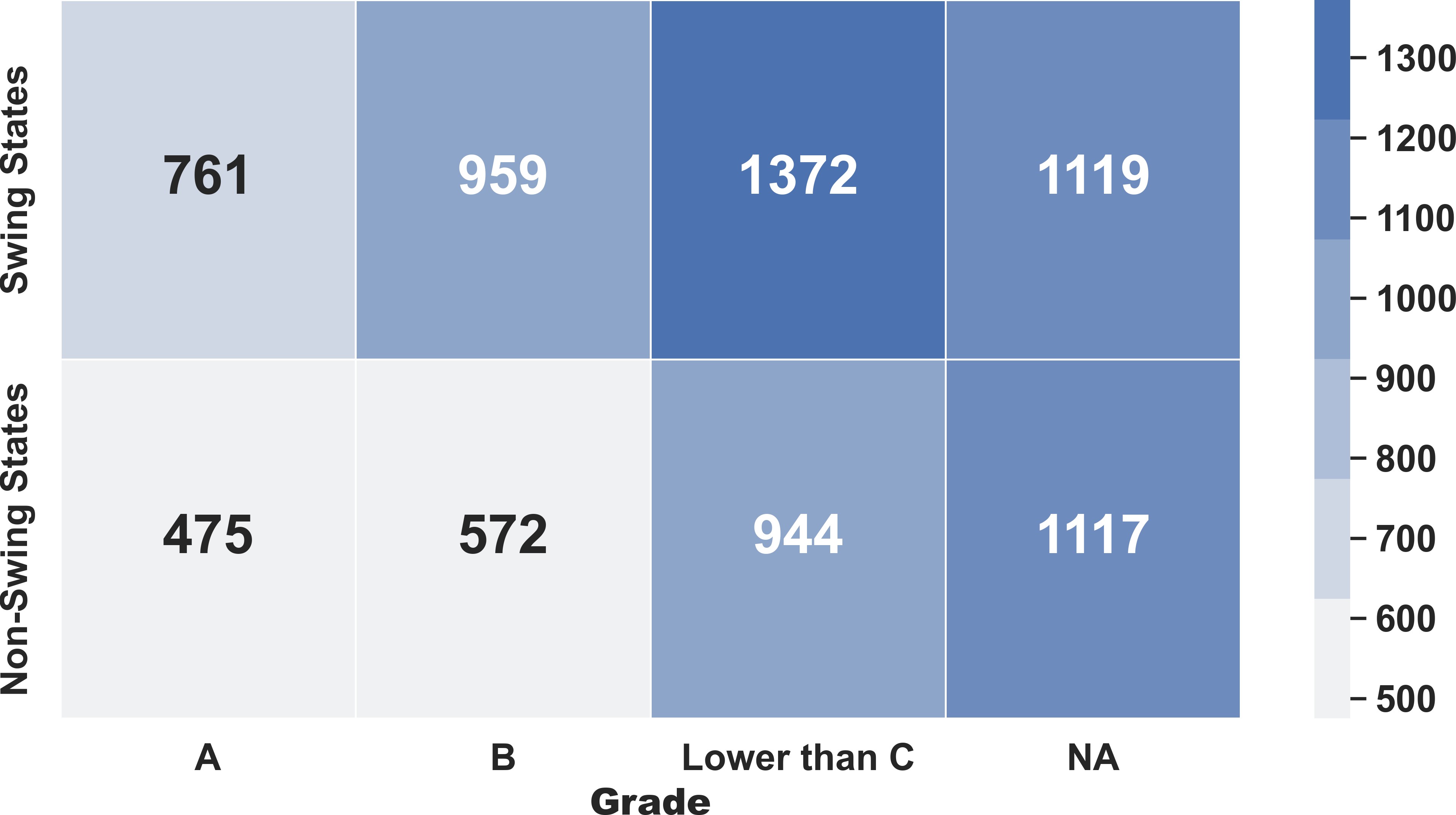
These two trends are not incidental because, generally, larger sample sizes can reduce typical errors in estimates from a poll and possibly mitigate some problems arising from the pollsters’ weaker methodologies, such as the ill- definition of "likely voters" discussed below.

* 1. **Most Polls are Targeted to Likely Voters.**

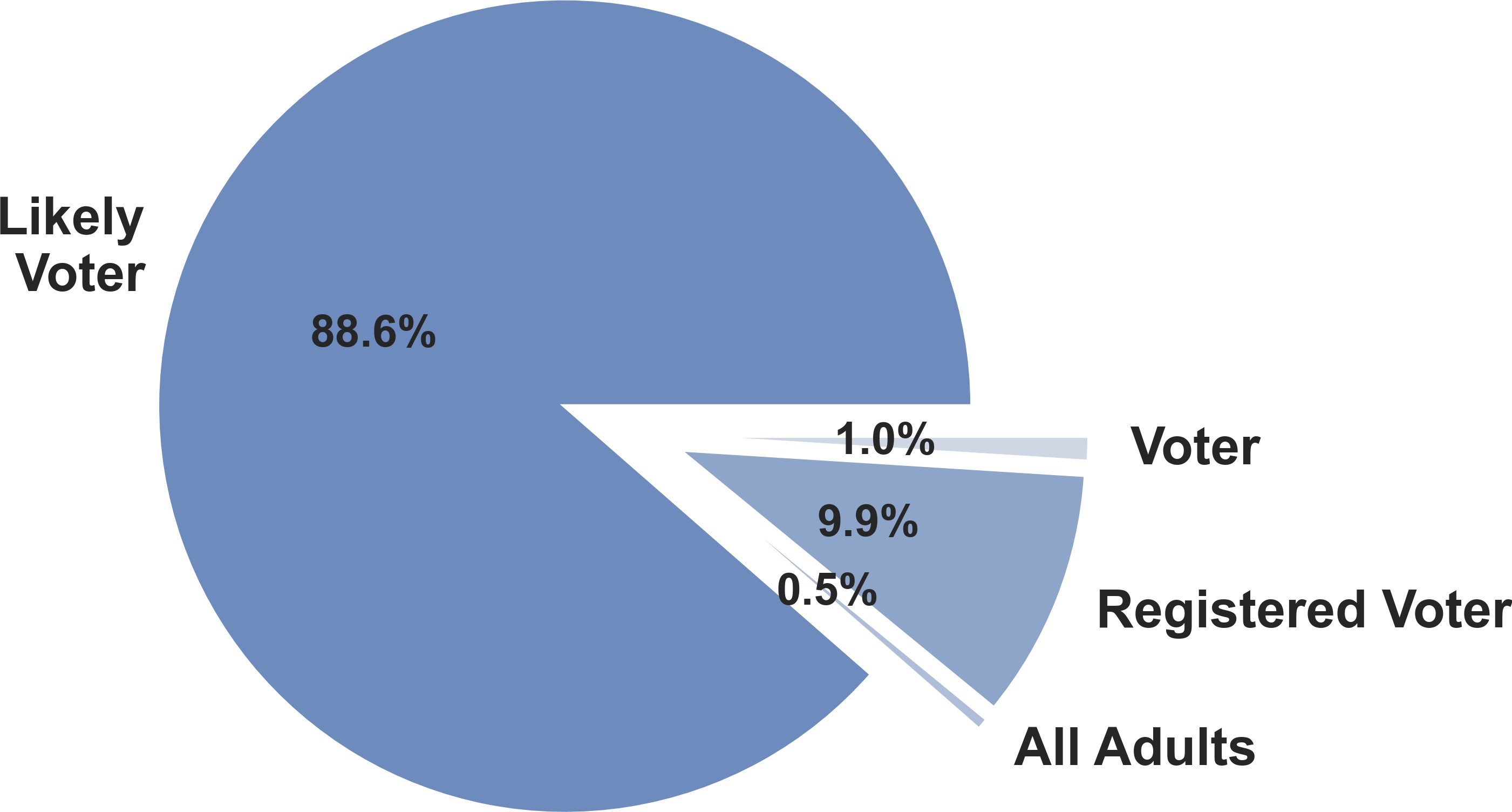
An overwhelming majority of the polls (88.6%) were tar- geted towards "likely voters" when compared to other pop- ulation types (Figure 8). This situation is problematic be- cause there is no universally accepted definition of a likely voter. The use of 2012 election turnout data by pollsters to build likely voter models did not fit the uniqueness of the 2016 election. Therefore, the significant deviation between the actual election result and the forecasts is understand- able. Section 4 of this report further discusses this issue.

2The poll-to-pollster ratios for Florida, Ohio, and Pennsylvania are 4.9,

4.9 and 4.3 respectively.



**Figure 7.** Average Sample Size by Pollster Grade, Swing States Vs. Non-Swing States, November 12, 2015 to November 6, 2016



**Figure 8.** Percentage of Polls by Population Type, November 12, 2015 to November 6, 2016

sensed the divisive political environment and lack of trust in government and portrayed himself as an outsider with an anti-establishment narrative catered to the "forgotten people". Many pollsters failed to fully comprehend this and underestimated Trump’s chances, by choosing a poll sample which underrep- resented Trump supporters.

**Polarization and Sowing Division**: Trump’s con- troversial comments regarding Mexican immigrants, Muslim ban, and border wall with Mexico among others divided people to the extent that many shy Trump voters did not wanted to openly mention their support of Trump to others. This resulted in lower engagement and under representation of Trump sup- porters in many polls, especially in the battleground states.

**Voters’ Distaste for Both Candidates**: Both Trump and Clinton were strongly disliked by voters. This made them the most unpopular presidential candi- dates since 1984 (ABC/Post poll). As a consequence, it led to an unusually large number of undecided vot- ers up till the last days of the campaign, especially in battleground states (Figure 9). Most polls were not able to take into account whom these undecided vot- ers ultimately chose to support due to them deciding right before the election.

### Remark 3

Rather than improving methodological weaknesses, poll- sters resorted to significantly larger sample sizes of swing- state polls to obtain more accurate estimates at the state level. It was an understandable compromise, given the limited resources of these pollsters in the swing states. Unfortunately, this seemingly unavoidable move was one of the causes of the inaccurate pre-election forecasts.

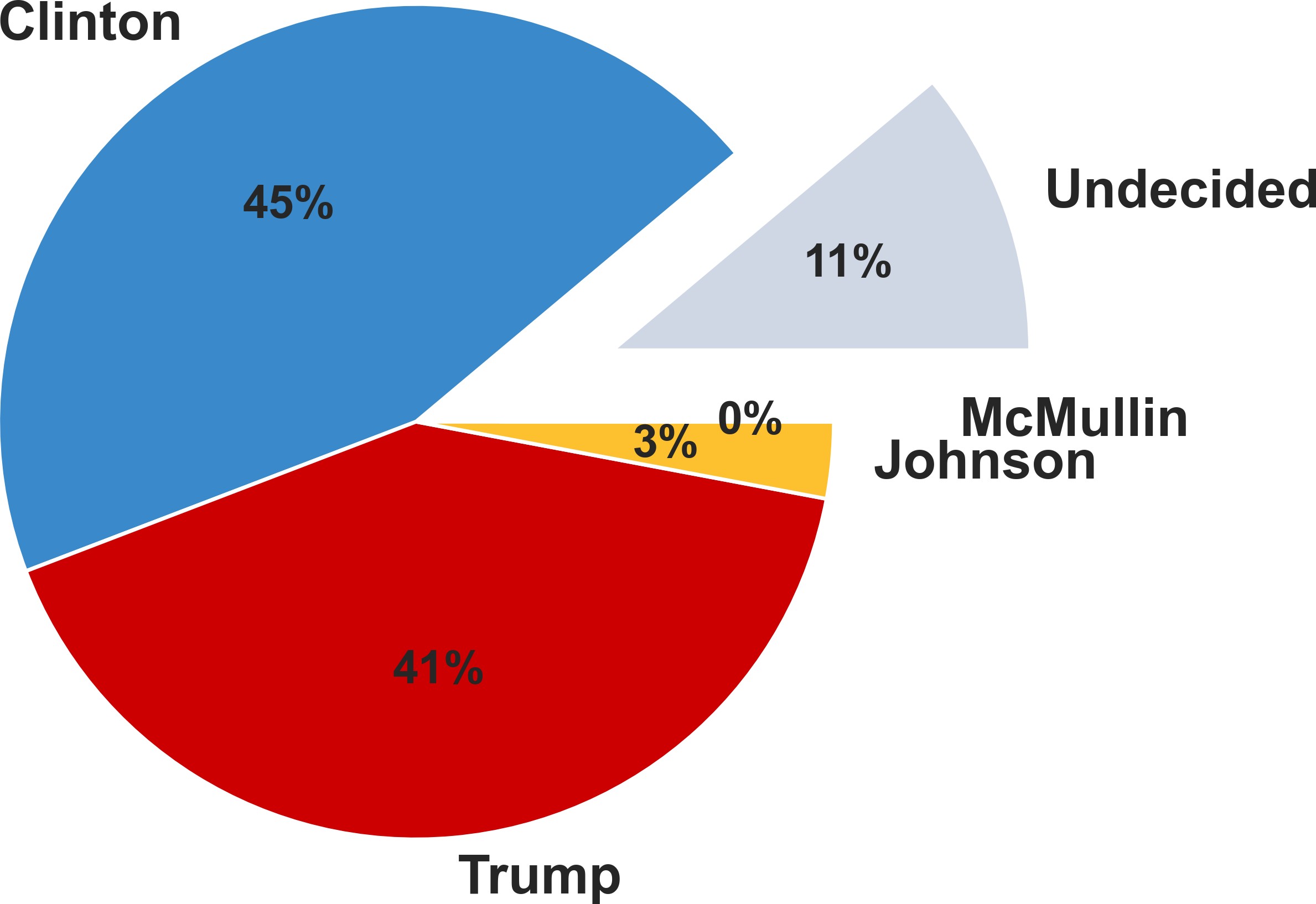
# 4. Problems, Lessons and Recommendations

This section expands on problematic characteristics of pre- election polls identified in Section 3 through the FiveThir- tyEight data set and discusses the reasons behind them.

* 1. **Problem 1: An Unprecedented Election**

Polling for the 2016 U.S. Presidential Election had the dif- ficult task of forecasting the winner of an election which was unprecedented in many ways.

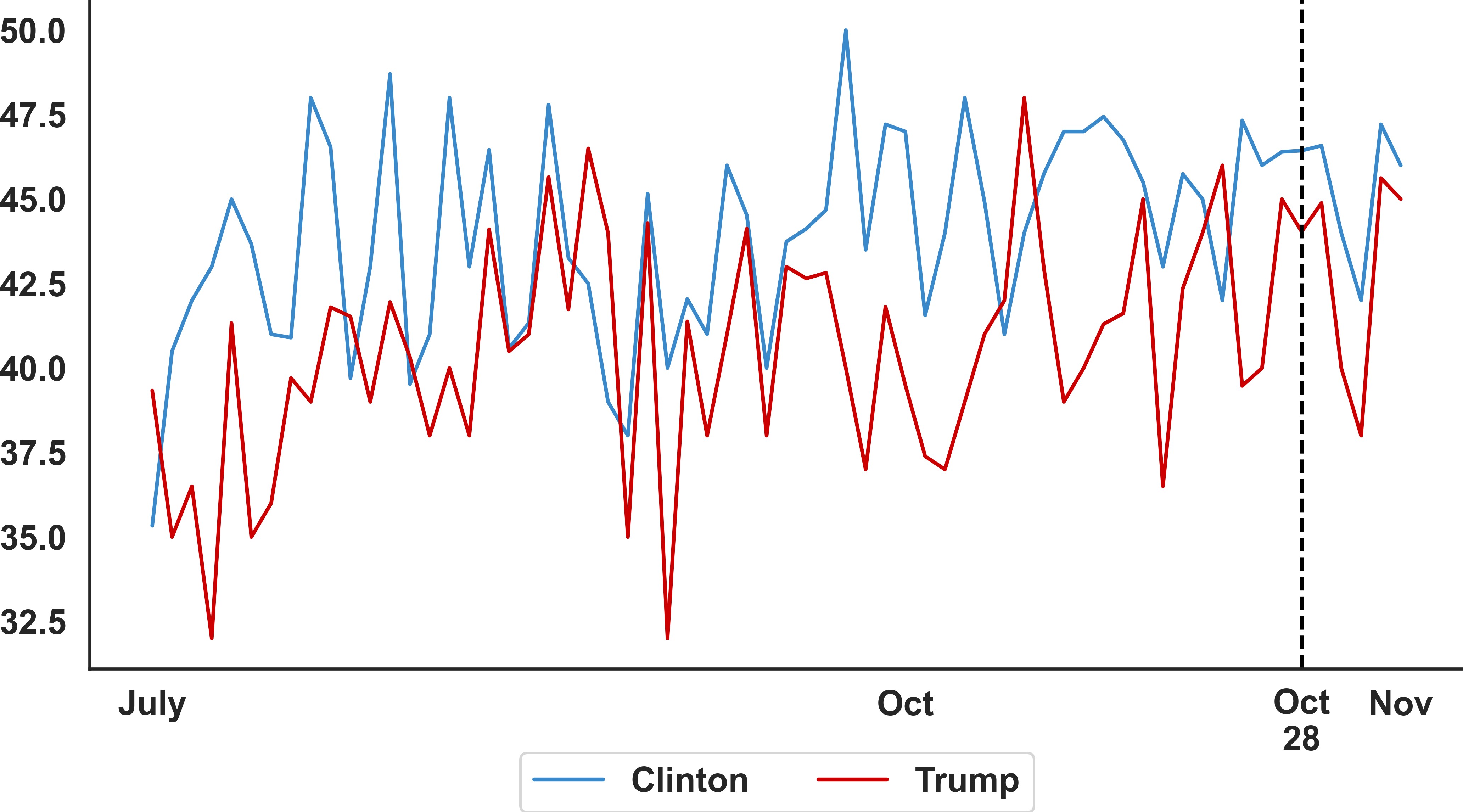
 **TV Star turned Presidential Candidate**: Trump had no previous political experience, leading to many peo- ple dismissing his candidacy as a "joke" on the premise that he was using it to garner fame and promote his businesses. They failed to notice that Trump had



**Figure 9.** Percentage of Undecided Voters, Grade-A Polls, Swing States, November 12, 2015 to November 6, 2016

 **Re-opening of Clinton’s Investigation**: On October 28, 2016, just eleven days before the election, the FBI director took the unprecedented step of announcing that upon discovering new emails, the investigation into Hillary Clinton’s email server was re-opening. This had a devastating effect on Clinton’s polling, though most of the polls were not able to take this event into account due to it being so close to the election (Figure 10). Despite no charges being file against Clinton, the topic dominated news headlines for the last days leading up to the election and may

* + 1. **Raw Polls**



* + 1. **Adjusted Polls - Raw Polls**



Note: On October 28th, 2016, the FBI re-opened the email server

investigation against Clinton.

**Figure 10.** The Evolution of the Average Percentages of Support for Clinton and Trump, Grade-A Polls Only, July 1st, 2016 to November 6th, 2016

have played a significant role in shifting undecided voters, especially in swing states, towards Trump.

**Lessons**

Failure to take Trump seriously and understanding who his true voter base was caused pollsters to design polls with a sample that underrepresented Trump supporters and caused polls in the battleground states to favour Clinton. Polling was also made unreliable due to a large number of undecided voters and major announcements very near the election, both factors which most polls were never able to take fully into account.

**Recommendation**

 In future elections with a significant number of un- decided voters, more polls should be conducted in the last few weeks of the campaign to capture the preference of undecided voters. Polling methodol- ogy should be designed to assign more weight to polls conducted after major announcements like an investigation related to one of the candidates being launched.

* 1. **Problem 2: Traditional Polling Method Unsuitable**

The polls for the 2016 U.S. Presidential Election suffered the problems of coverage, non-responses, the ill-identification of likely voters and the lack of pollsters in battleground states.

 **Coverage Problem**: There was a under-coverage of non-college educated voters and hence a over- coverage of college educated voters among all pre- election polls.

 **Partisian Non-response**: Trump voters did not re- spond to telephone surveys because of their lower level of engagement.

 **Identification of likely voters**: There are various ways to identify a likely voter, including using the 2012 voter files and asking the respondent directly for their intention to vote. Because most votes at the state level are targeted to likely voters (Figure 8), a problematic identification of the likely voters will distort pre-election forecasts.

While identifying likely voters by the 2012 voter files seems more trustworthy than other means, this could have easily inflated support for Clinton while un- derestimating the support for Trump. For example, these models would have expected solid support from African American voters for Clinton in 2016, as was the case for Obama in 2012. In reality, African Amer- ican voter support for Clinton was much lower than anticipated. On the other hand, groups of voters ex- pected to support Trump turned out in much higher numbers than expected. The voting patterns were not the same in 2012 and 2016. Pollsters who created "likely voter" models using 2012 data caused their polls to be unreliable are the people who are most expected to show up at the ballot box based on past voting behaviour and a slate of other characteristics, and polls of this group would more closely illustrate how the electorate will act on Election Day, compared to polls of the broader population at large.

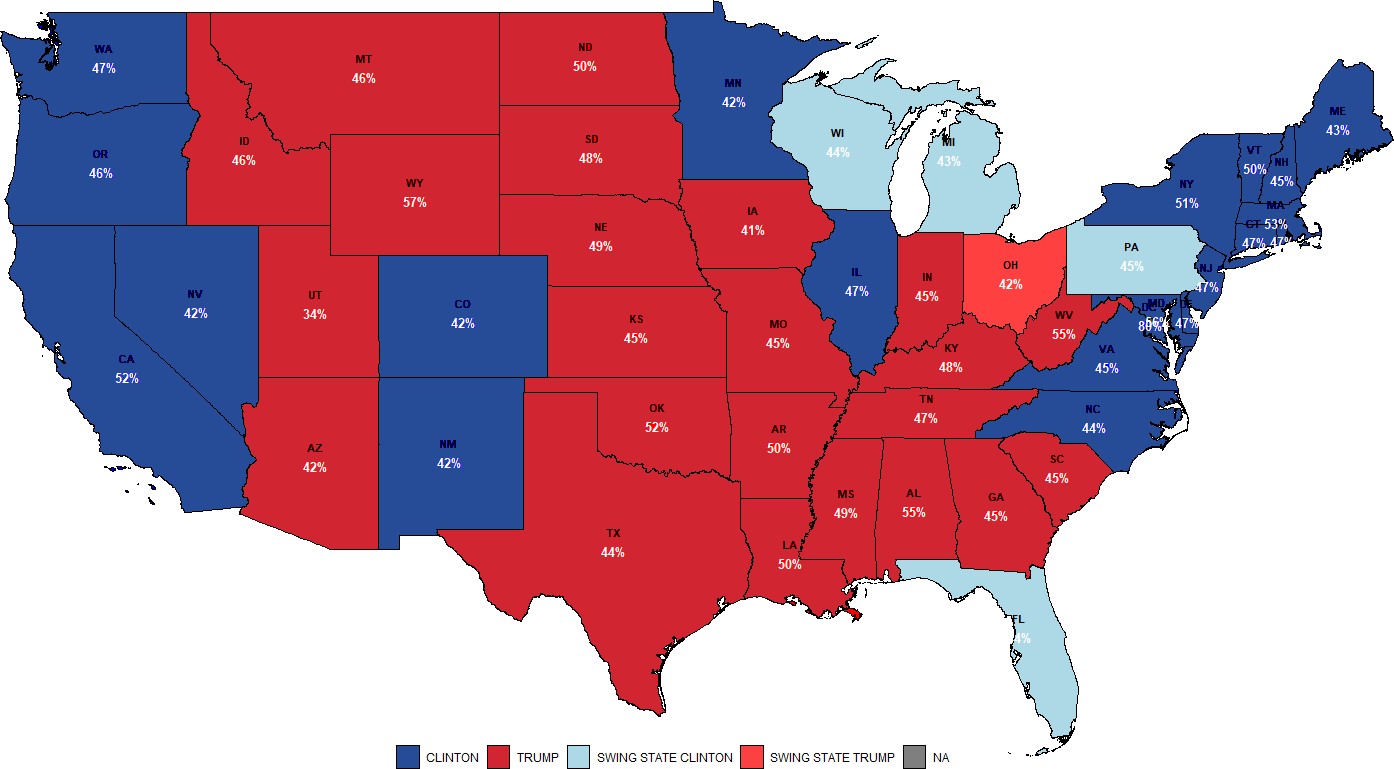
 **The Lack of Pollsters in Battleground States**: The number of pollsters in swing states were significantly fewer than other states (Figure 5) because of bud- getary constraints.

**Lessons**

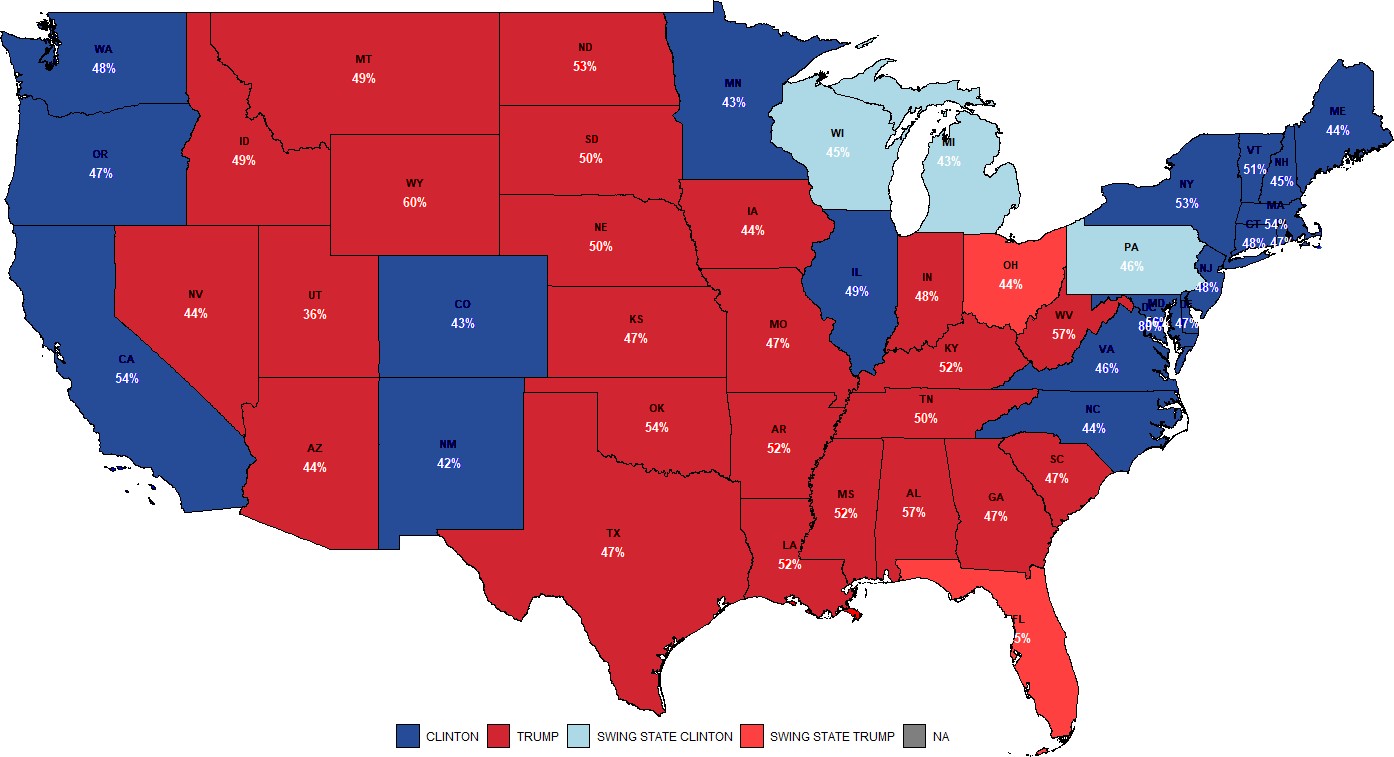
These four issues caused the failure to capture voter habits and characteristics during the pre-election period. There- fore, pre-election polls, on average, cannot capture Trump’s lead in swing states (panel (a) of Figure 11). These issues, especially in this tight race in terms of popularity votes, hampers the accuracy of the forecasts. The situation did not improve even after FiveThirtyEight’s adjustment (panel

(b) of Figure 11). Indeed, the differences between the two panels on Figure 11 are not material.

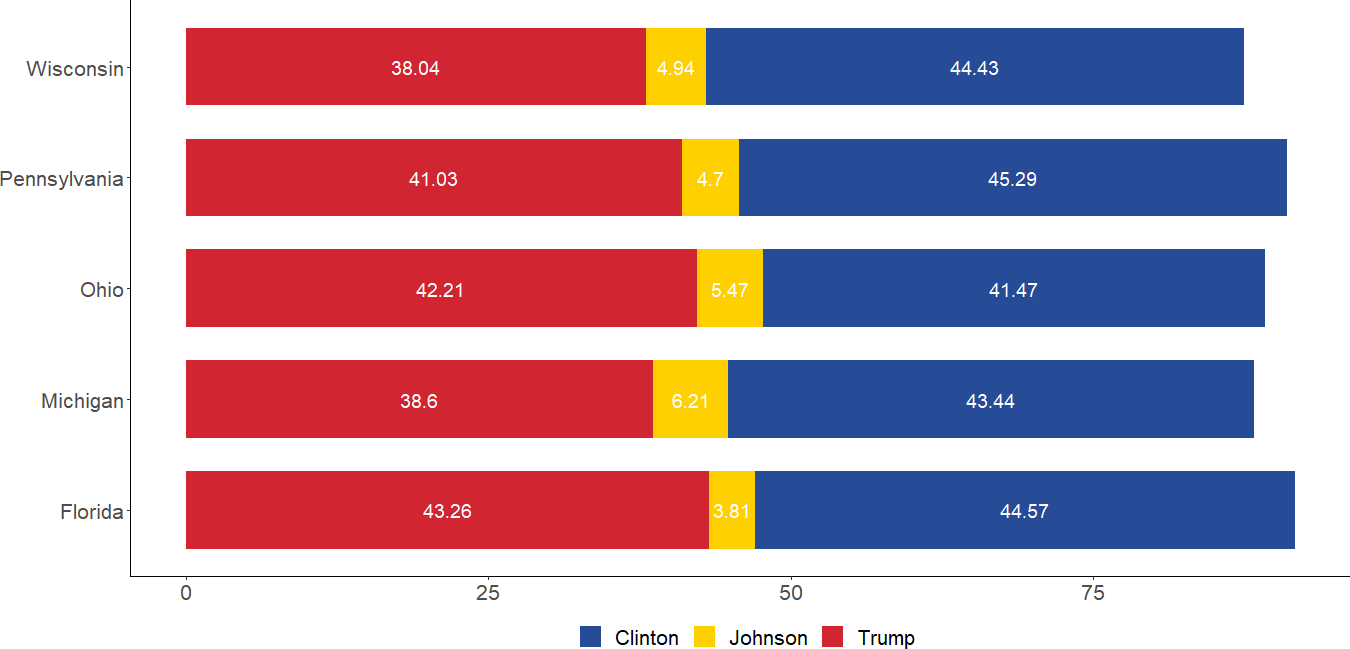
* + 1. **Raw Poll**

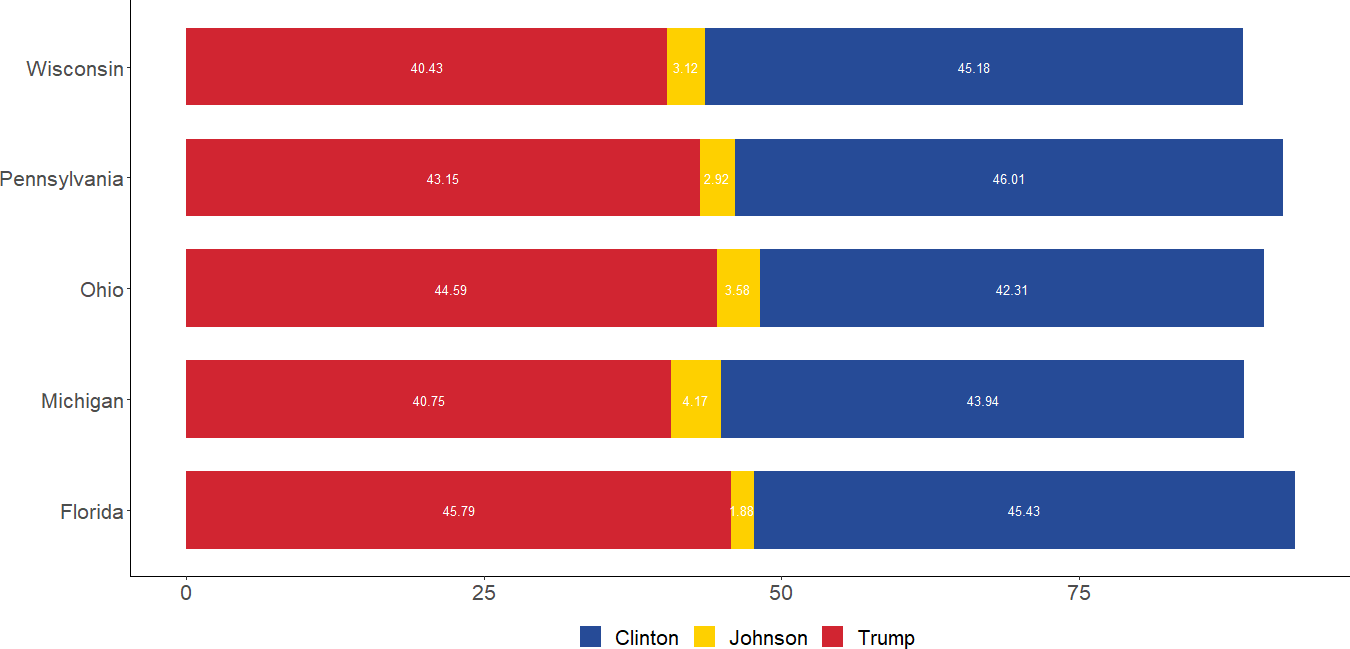


* + 1. **Adjusted Poll**



**Figure 11.** Average Grade-A Polls in Major Swing States (Weighted by Sample Size)

1. **Raw Poll**
2. **Adjusted Poll**



**Figure 12.** Average Grade-A Polls in Major Swing States (Weighted by Sample Size)

**Recommendation**

 As discussed in Remark 3 of Section 3, many poll- sters in the swing states are run on a tight budget. Therefore, we suggest that well-resourced survey in- stitutions can collectively finance high-quality state- level polls and encourage more pollsters to carry out state-level polls.

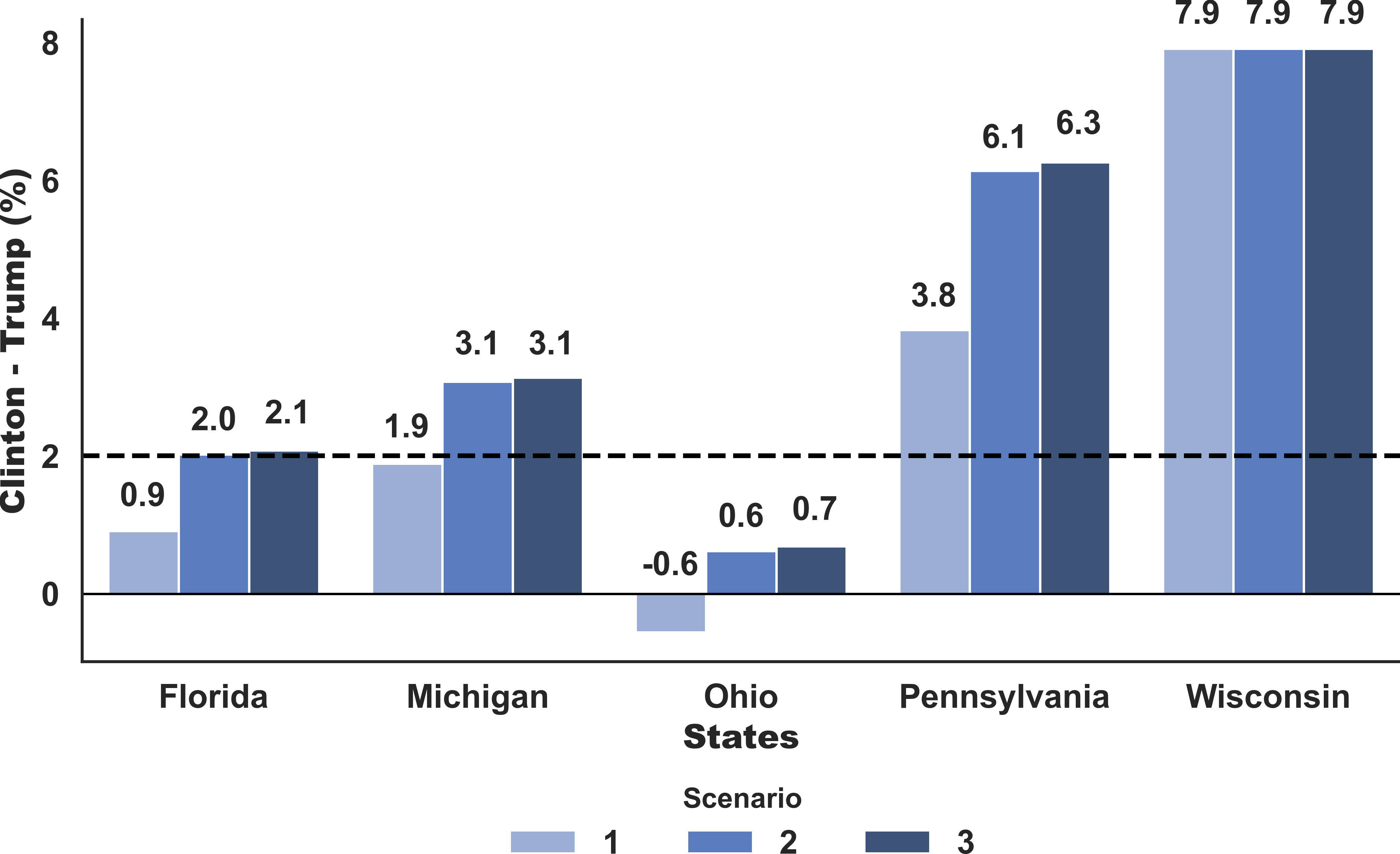
* 1. **Problem 3: The Johnson Effect**

Our discussion up to this point has not mentioned the third candidate of the election – Gary Johnson. However, by no means would we like to undermine this candidate’s impor- tance in this election, especially when it comes to this partic- ularly tight race. Figure 12 show that, on average, if Trump obtained all of Johnson’s support, the difference between Trump and Clinton would have been within the margin of error. Without Johnson, most pre-election polls may not

have predicted a landslide victory for Clinton. Therefore, the role of Johnson in pre-election polls could be important to the forecast.

We use a simple exercise to quantify the effect of a third popular candidate in the election. Specifically, we compare the November average grade-A raw polls in each swing state with the three hypothetical scenarios (Figure 13). These three scenarios assume that Clinton obtained 60% (Scenario 1), 95% (Scenario 2) and 100% (Scenario 3) of Johnson’s support in each poll, holding the Trump’s support constant. Then we compare the resultant average percentage for Clinton with that for Trump to see if forecasts change.

Assuming a 2% margin of error, we observe that Clinton would need to obtain 100% of Johnson’s support to safely win the three most important swing states at the same time:



**Figure 13.** The Average Percentage Point Difference between Clinton and Trump for Each Hypothetical Scenario, Raw Polls of Swing States, Grade-A Pollsters Only, November 2016

(1) Wisconsin, (2) Pennsylvania and (3) Florida. Despite the simplicity of this exercise, this exercise shows the possibility of Johnson’s influence on the election outcome. Of course, further analysis would be required to look into how Johnson affected the election forecast.

natesilver538. (2016, June 29). *A user’s guide to fivethir- tyeight’s 2016 general election forecast*. FiveThirtyEight. Re- trieved October 11, 2022.

Watkins, E. (2016, November 26). H*ow Gary Johnson and Jill Stein helped elect Donald Trump* | CNN politics. CNN. Retrieved October 11, 2022.

Wikimedia Foundation. (2022, September 28). Gary John- son 2016 presidential campaign. Wikipedia. Retrieved October 11, 2022.

Zhou, L. (2018, October 24). *How pollsters are responding to problems that came up in 2016*. Vox. Retrieved October 11, 2022.

# 5. Conclusion

In conclusion, problems associated with swing state polls caused the inaccurate election forecast of a landslide vic- tory for Clinton. Because of Trump’s unconventional nature, samples of polls under-represent Trump’s supporters. Mis- leading estimates from these polls seriously overestimated Clinton’s magnitude of support. This problem was worse in swing states because of the lack of pollsters and their lack of resources. To alleviate this issue, we suggest well- resourced pollsters assist these pollsters financially. The impact of the third candidate was possible and is worthy of further investigation.

# 6. Bibliography

CBS Interactive. (n.d.). *What went wrong with polls in 2016?* can we trust them now? CBS News. Retrieved Octo- ber 11, 2022.

Cohn, N. (2017, May 31). *A 2016 review: Why key state polls were wrong about trump.* The New York Times. Re- trieved October 11, 2022.

Kimmell, A. (2016, November 22). *Donald Trump and Hillary Clinton were both "unlikeable," but that only mattered for one candidate*. Presidential Gender Watch. Retrieved October 11, 2022.

Krieg, G. (2016, December 22). *It’s official: Clinton Swamps Trump in popular vote* | CNN politics. CNN. Retrieved Octo- ber 11, 2022.