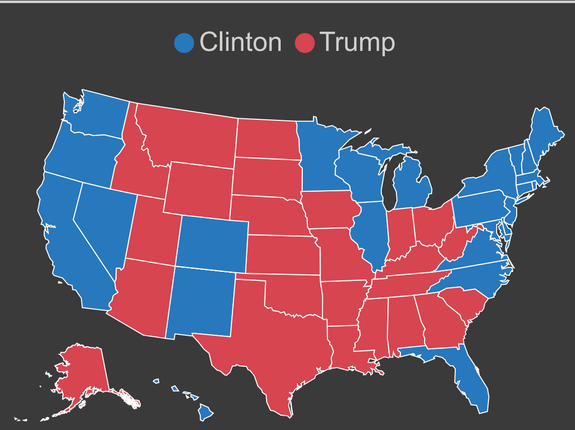
THE 2016 POLLS: WHERE THEY WENT WRONG
A VISUAL REVIEW

During presidential elections, polls and surveys are essential tool that help gauge candidates popularity and predict potential outcomes. Although, they often are accurate tools of estimation, but can sometimes prove unreliable. The electoral polls of 2016 serve as a prime example, as they have missed their projection margins by 2% to 5%.

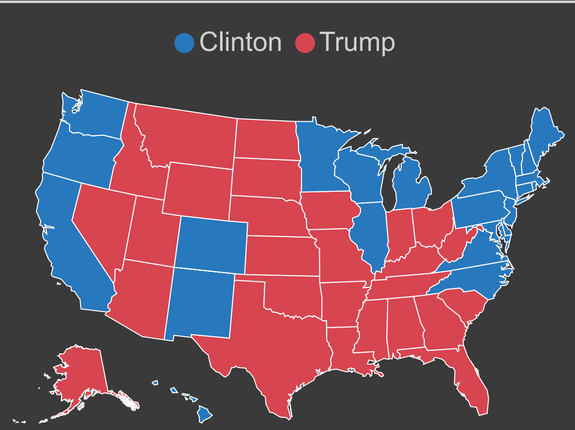
While the results consistently predicted Hilary Clinton's victory, with a popular vote in her favour and an evenly split electoral map, the actual results have been telling a different story. Which prompt us to question the credibility of the polls and whether the prediction mistakes were avoidable. However, many factors can influence their precision. In the case of presidential elections, we must take into account the effect of swinging states, debates and polling demographic, such as the education level among the surveyed populace.

Within the context of our techniques of data analysis course (MAT4376), we employ data visualization to explore and understand the disparities between the 2016 election polls and realized outcomes. We discover that beyond numbers and graphs, there are more to using surveys for prediction, as faulty answers and lack of representation can become hindrances to our results.

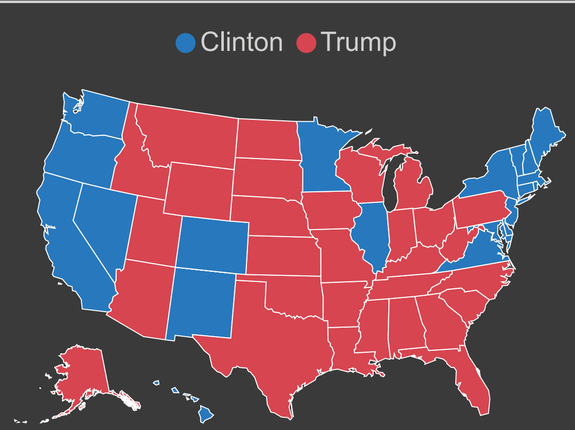
Majority per State Raw Poll



Majority per State Adjusted Poll



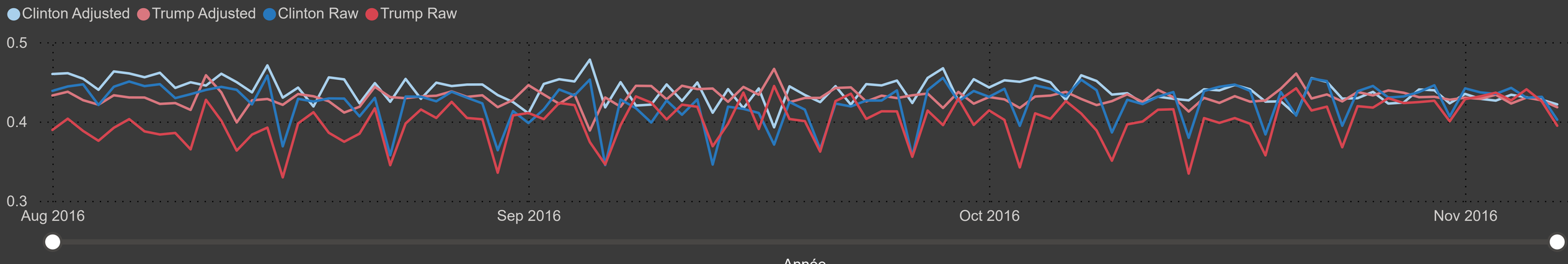
Actual Results



Electoral Maps

Before delving into the underlying causes of the disparities within the poll results, it is essential to grasp the broader context. Polls are often adjusted to represent missing demographic, or correct possible erroneous votes. In the maps above, we notice that adjusted polls have narrowed the gap between predictions and realized outcomes when compared to raw poll data. Additionally, from the time series graph seen below, adjusted results are less volatile. That being said, differences between the adjusted and actual results remain significant, as we see many states shifting colours, with some situated in the Rust Belt region.

Adjusted and Raw Polls Over Time



THE RUST BELTERS

The Rust Belt is a term used to refer to a region in the northeastern and midwestern United States. It was formerly home to America's most important and core industries and once comprised the nation's industrial heartland. However, it has seen a sharp decline in recent decades due to the increased cost of domestic labor and capital-intensive manufacturing. Several political swing states such as Pennsylvania, Ohio, Wisconsin and Michigan are part of the Rust Belt. Since these states are often the deciding factor in presidential elections, the Rust Belt plays a crucial role in US politics.

The population structure, which is characterized by white industrial workers, used to be considered a stronghold of the Democrats and part of the so-called "blue wall." Donald Trump shattered the blue wall in 2016 which has not happened since 1988, thus made his election possible. On the right hand side we can see a comparison of the rust belters states in the 2012 elections with the actual 2016 election results and the predictions polls results. We can see that for three particular states: Michigan, Pennsylvania and Wisconsin, the predicted winning candidate are different from what the actual results tells us. Below are some key reasons why the Republican Party representative was able to beat the odds :

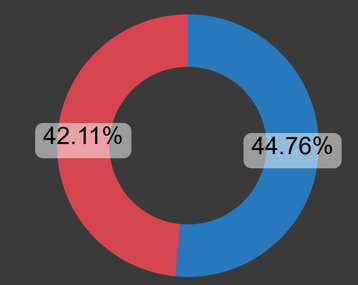
- Michigan has a significant manufacturing and automotive industry presence, making economic concerns a significant factor in its voting patterns. Donald Trump clearly had an advantage by winning the 2016 ORC poll, on who would *"better handle economy?"*, by 15% in Midwest and 5% nationwide.
- Voters who were dissatisfied with the government and assumed the corruption of Washington, overwhelmingly went for Trump by default.
- Additionally, Hilary Clinton mistakenly took the Rust Belt state for granted during presidential tour, leading the a preferential voting result for Trump.

In the next page we notice that although the polls predict the win of the democrates, in the above mentioned stated, their actual county map are significantly republican dominant. While the candidates popularity overtime shows Clinton leading the race until the last day before election, Trump won by a less than a one percent margin in each of these states. This outcome might be attributed to Trump's compelling arguments during debates.

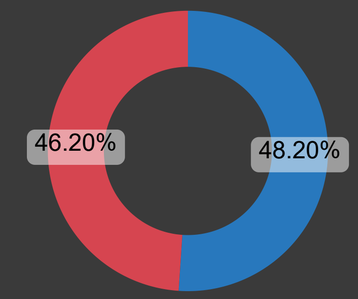
State

- Colorado
- Florida
- Iowa
- Michigan
- Nevada
- New Hampshire
- North Carolina
- Ohio
- Pennsylvania
- U.S.
- Virginia
- Wisconsin

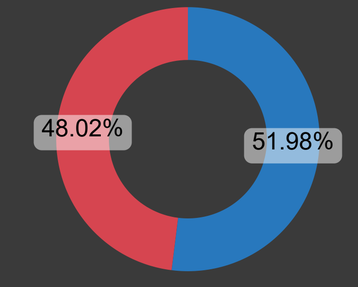
2016 Poll Results



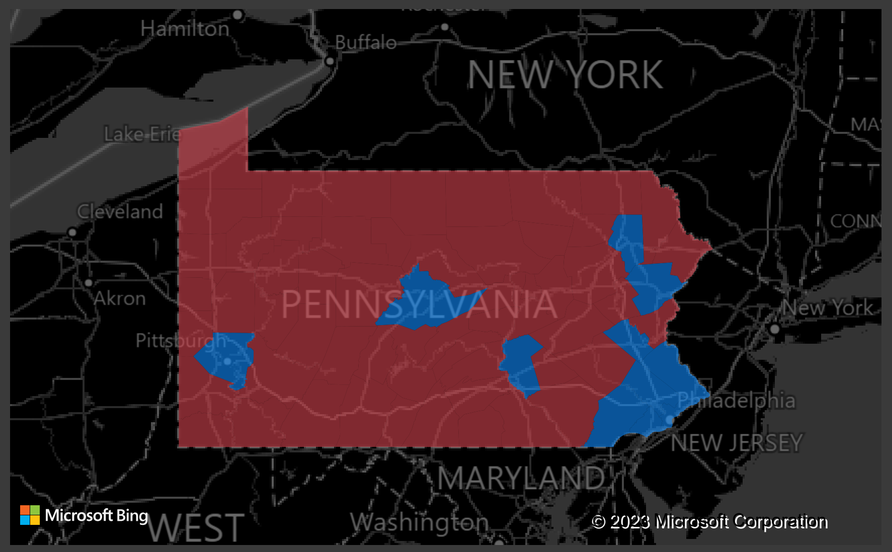
2016 Actual



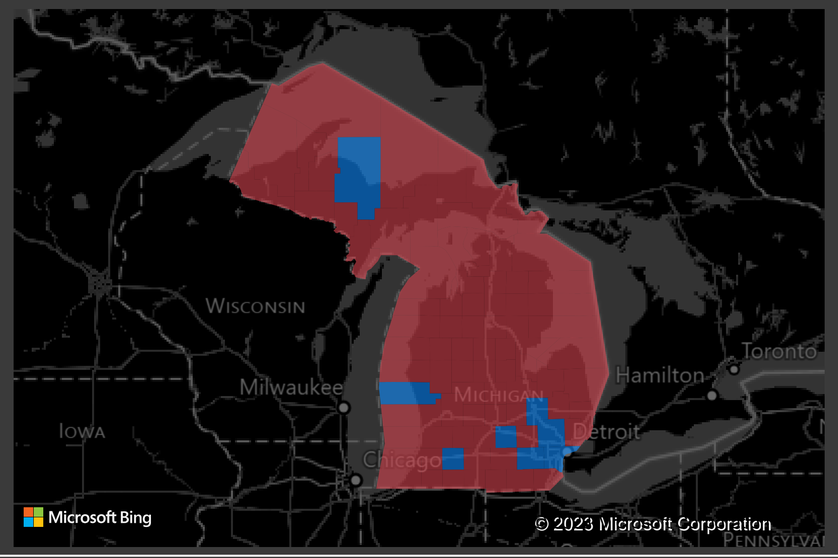
2012



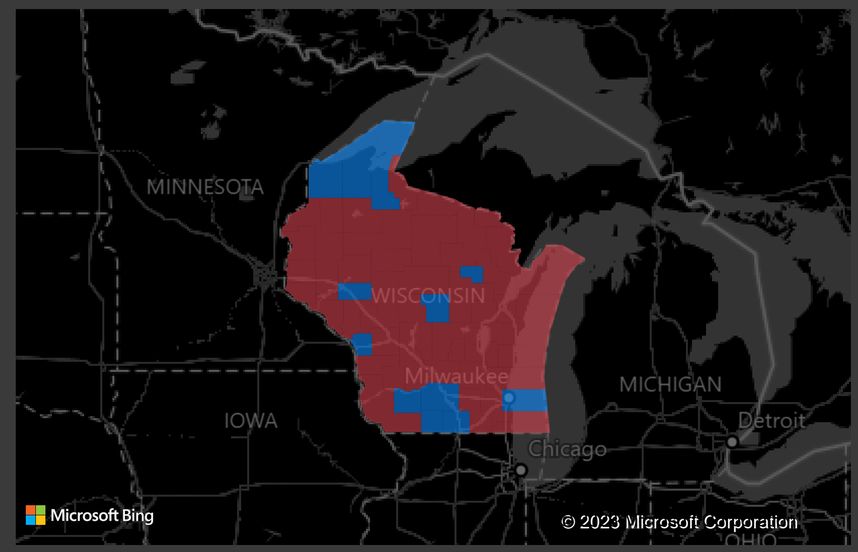
Pennsylvania



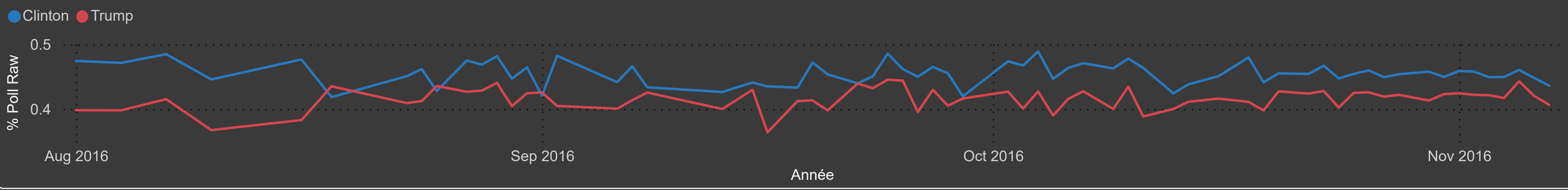
Michigan



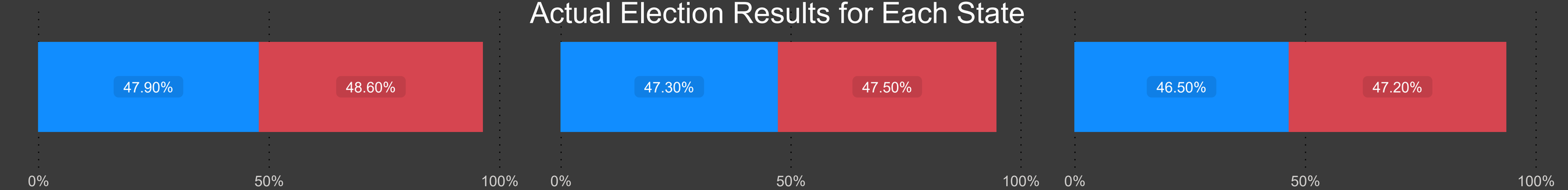
Wisconsin



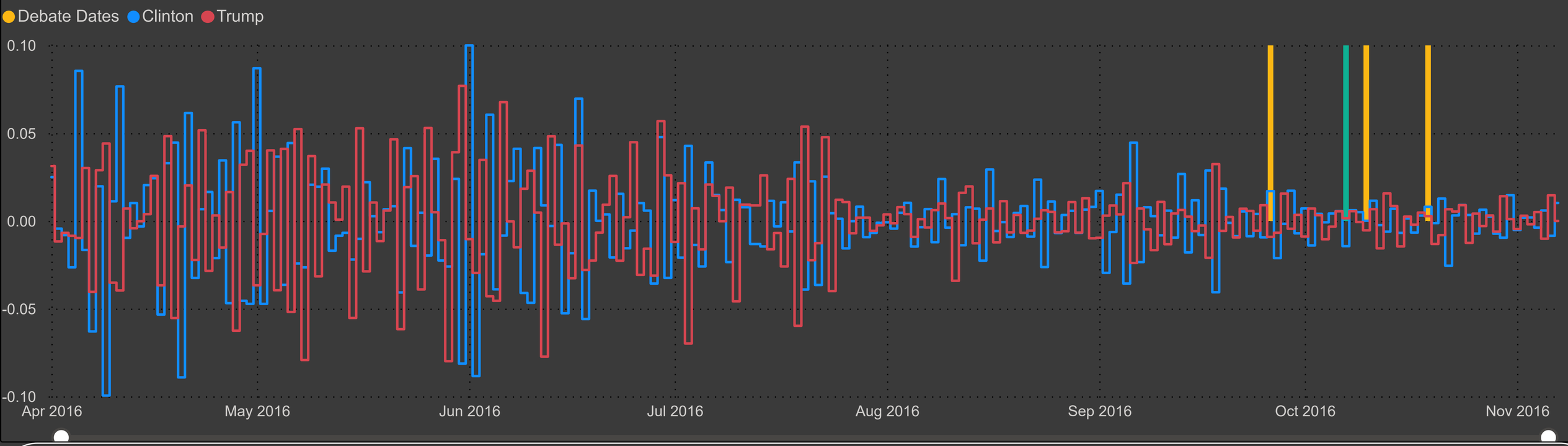
Poll Variation Over Time



Actual Election Results for Each State



Voting Variation Over Time



Do Presidential Debates affect election results?

The final stretch of elections witnesses the crucial occurrence of debates, allowing candidates to increase their popularity and defend their policies. By looking at the graph thoroughly around the period when debates occurred, we notice a decline in Hilary's votes, leading us to speculate that these debates could have had a negative impact on her presidential ambitions. However a definite cause for the events remain elusive.

According to Pew surveys from 1988, the voters see debates as "useful" but not essential for making their decisions. In other words, voters who change their opinions about a politician may do so based on new information about this candidate or their stance on major issues, rather than the debates.

In 7 October 2016, an important event related to Hillary Clinton email controversy, shifted the rest of the 2016 Presidential campaign in Trump's favor. Its incessant media coverage is was played a significant role in tipping the election.

Despite Clinton's noteworthy ability to exhibit her informed leadership with the temperament to be president, during the three debates, her impact on the voters' primary concern, the economy, fell short of Trump's unparalleled effectiveness in that domain.

Despite these significant shifts in votes as the election neared its conclusion, the polls remained inaccurate on multiple fronts.

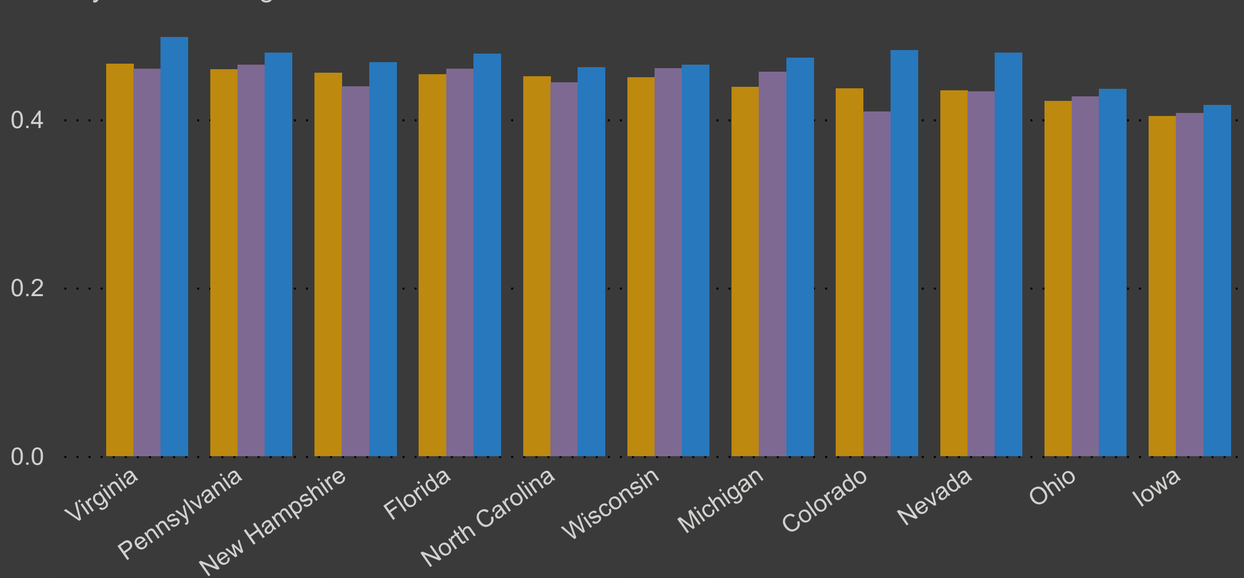
WEIGHTING THE DATA

Poll accuracy are heavily influenced by the surveyed demographic and nonresponse biases. This has affected the polling accuracy as some groups, such as less educated voters who supported Trump, were not well represented. For instance, in Pennsylvania, state-level polling also overestimated democratic support. The polls failed to account for education level as a predictor of candidate preferences, as there was a higher voter turnout among those without college degrees.

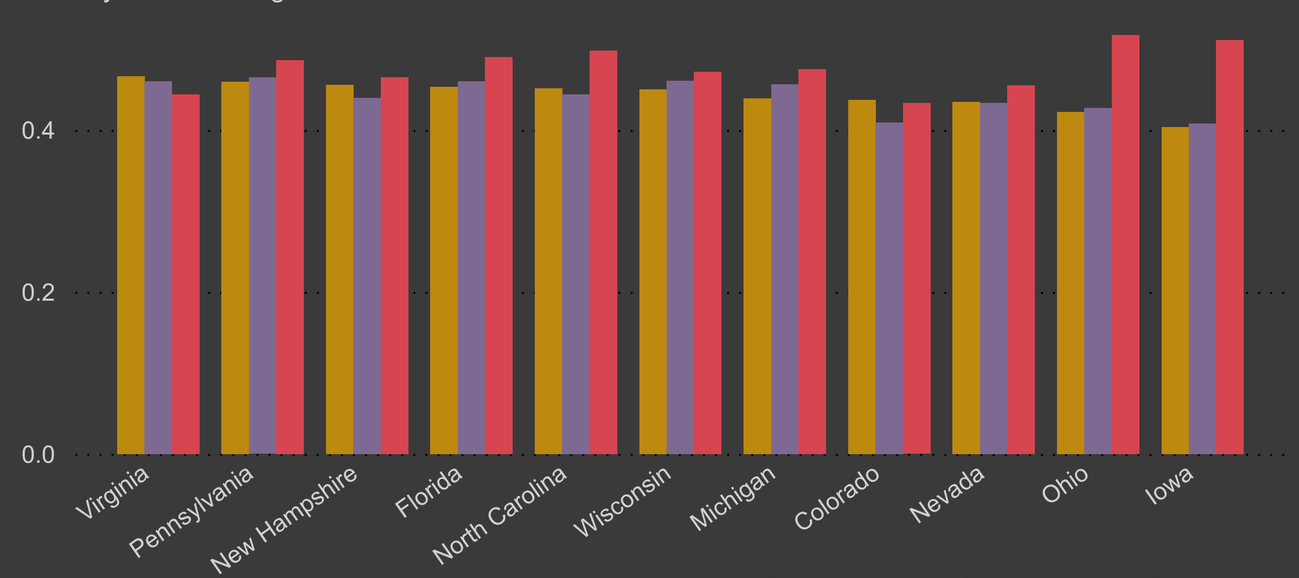
Another suggested reason is the existence of "shy Trumpers," individuals who were reluctant to express their support to Trump due to social pressure and perceptions. This phenomenon could have led to an underestimation of Trump's support in the polls.

The third possibility relates to the challenges pollsters face in identifying likely voters. Below we observe a higher error rate in term of average and standard deviation for registered voters in comparison to likely voters. Predicting who will actually vote and what the electorate will look like on Election Day is a complex task. Small variations in assumptions can significantly impact the accuracy of election predictions. It is suggested that the voters pollsters expected, particularly in the Midwestern and Rust Belt states, were not the ones who showed up to vote, which could have contributed to the surprising outcomes.

Clinton - Repondents Comparison



Trump - Repondent Comparison



RESIDUALS STATS

0.02	0.03	0.03	0.04		0.06	0.03	0.05	0.04
Average	Average	Average	Deviation		Average	Deviation	Average	Deviation

Conclusion

In this report, we have conducted an analysis of the 2016 US election polls, which revealed significant inaccuracies in their predictions. Despite consistently favoring Hillary Clinton with a popular voting advantage, the actual results told a different story. By investigating the effect of various factors, such as swinging states, impactful events, and polling demographics, we aimed to understand the reasons behind these prediction mistakes and whether they could have been avoided.

We therefore come to the conclusion that, polls cannot provide perfectly accurate prediction when significant variables are not taken into account. That is, we have observed the great impact that considering each demographics historical and educational background as well as properly weighting the data, can have on our predictions. In the case of the 2016 polls, for instance, the Rust Belt regions which encompasses the northeastern and midwestern states, played a pivotal role in the election. Additionally, not taking into account possible nonresponse biases, the presence of "shy Trumpers" and many other factors did in fact disturb the polling.

That being said, it is not impossible for polls to be 100% correct, and although the 2016 polls were particularly off, is wasn't without unpredictable candidates and events.

We will still emphasized the importance of considering these factors in future election predictions to improve accuracy with the aim of a healthier democratic process, increasing transparency, accountability, and trust in the electoral system.

Dictionnary

Field Name	Field Size	Data Type	Data Format	Description	Example
Population	2	Boolean	--	Voters nature	lv
Adjpoll_clinton	8	float	NN.NNNNN	% of trump voters adjusted	43.09876
Adjpoll_trump	8	float	NN.NNNNN	% of clinton voters adjusted	47.98765
Grade	2	string	--	Relevancy of the pollster	A+
Rawpoll_clinton	5	float	NN.NN	% of clinton voters	45.54
Rawpoll_trump	5	float	NN.NN	% of trump voters	48.98
Sample Size	5	Integer	NNNNN	# of People in a sample	22456
State	13	string	----------	Location of the voters	Virginia

: number
% : pourcentage

- Voters nature includes likely voters (lv) indicating people who pollsters think are more likely to vote, registered voters (rv), all adults (a) indicating every adults in the location combining registered and non-registered voters, and voters (v) used by Public Policy Polling to indicate who they think are actual voters.
- The adjpoll determines Polls adjusted by the pollster with different weights according to several factors including historical information on the population in question.
- Grades are affected to pollsters dependant on the proximity of their previous reports to the actual electionresults (their relevancy)

References

https://apl.wisc.edu/shared/tad/how-and-where-trump

https://projects.fivethirtyeight.com/2016-election-forecast/wisconsin/

https://perspective.usherbrooke.ca/bilan/servlet/BMAnalyse/2338

https://www.cbc.ca/news/politics/grenier-uspolls-trust-1.5765695

https://whyy.org/articles/what-pa-pollsters-got-wrong-in-2016-and-why-2020-might-be-different/

https://www.cbc.ca/news/politics/grenier-uspolls-trust-1.5765695

https://www.insideover.com/politics/what-is-the-rust-belt-and-why-it-matters-on-election-day.html

Contributors

Qualitative Analysis and Storyline:
Mohamed Diakhate
Salma Chiheb
Alphonse Marie Diatta

Visualization and Analysis:
Amina Al-Nadjib