

Red, Blue, or Purple? 2024’s Battleground States Tell a Story*

Pennsylvania, Arizona, and Nevada Become Game Changers as Trump Gains Ground [To be Update]

Yingke He Ziheng Zhong

November 3, 2024

This paper analyzes polling data for the 2024 U.S. Presidential election, examining the key candidates notably Donald Trump and Kamala Harris to gain a comprehensive perspective on the race. A Hierarchical Bayesian Model is applied in this study and the findings reveal significant variations across states, with [To be updated...] emerging as particularly influential battlegrounds that ultimately tilt in favor of [To be updated..].By highlighting these trends, this research underscores the dynamic and evolving nature of U.S. political landscapes and the crucial role that swing states play in determining election outcomes.

Table of contents

1	Introduction	2
2	Data	3
2.1	Overview [info need update]	3
2.2	Data Measurement [info need update]	3
2.3	Outcome variables	5
2.4	Predictor variables	5
2.4.1	State	6
2.4.2	Pollster	6
2.4.3	Candidate Name	6
2.4.4	Percentage of Votes	7
2.4.5	Sample Size	7
2.4.6	Pollscore	7

*Code and data are available at: https://github.com/iJustinn/Election_Prediction.

2.4.7	Days to Election	7
3	Model	8
3.1	Model set-up	8
3.1.1	Interpretation of Parameters	10
3.1.2	Prior Distributions	10
3.1.3	Model Justification	10
4	Results	14
4.1	Predicted Electoral Outcomes	15
4.2	Electoral outcome by states	16
5	Discussion	18
5.1	State Effect	18
5.2	Pollster and Pollscore Effect	18
5.3	Days to Election Effet	20
5.4	Weaknesses and Next Steps	20
	Appendix	21
A	Additional data details	21
B	Model details	21
B.1	Posterior predictive check	21
B.2	Diagnostics	21
	References	22

1 Introduction

In the lead-up to the 2024 U.S. Presidential election, understanding voter behavior across states and demographic groups is essential for accurately predicting electoral outcomes. Recent polling data from various sources suggests a close race among major candidates, including Donald Trump, Kamala Harris, and other key contenders. This study applies a Hierarchical Bayesian Model to analyze polling data across all U.S. states, with a focus on identifying critical swing states where even slight shifts in voter sentiment could prove decisive. By exploring trends in voter support across different states and demographic groups, this research aims to highlight the regions most likely to influence the final election outcome.

The primary objective of this study is to estimate state-level vote shares for each candidate while accounting for factors such as regional variations, pollster-specific biases, and the timing of polls relative to Election Day. Predictors including polling percentages, sample sizes, and days remaining until the election are incorporated into the model, which adjusts for state and

pollster effects. The model’s structure enables the estimation of vote shares that reflect the underlying distribution of polling data across diverse regions in the U.S.

Initial findings reveal variations in candidate support across states, with certain swing states emerging as pivotal in determining the Electoral College result. The model’s projections suggest a likely lead for Trump in the Electoral College, underscoring the influence of large-sample, recent polls on forecast accuracy. This study’s results emphasize the importance of swing states in the electoral process, demonstrating how regional dynamics and polling methodologies impact predictions.[To Be updated...]

This research contributes to the field of election forecasting by combining aggregated polling data with a robust modeling approach, offering valuable insights for political analysts, campaign strategists, and policymakers. By identifying critical swing states and accounting for potential polling biases, this study equips stakeholders with the tools to anticipate voter shifts and strategize effectively in an evolving political landscape.

The structure of the paper is organized as follows: following Section 1, Section 2 outlines the data collection and cleaning process, along with a description of the outcome and predictor variables used in the analysis. @sec-model, introduces the forecasting models and discuss the rationale behind choosing these models for election outcomes prediction. Section 4 then presents the main findings, including a breakdown of state-level and pollster-level random effects. Finally, Section 5 interprets the results, highlighting significant trends and predictions, and concludes with a discussion on the reliability of the forecasts and potential limitations of the models.

2 Data

2.1 Overview [info need update]

We use the statistical programming language R (R Core Team 2023).... Our data (Toronto Shelter & Support Services 2024).... Following Alexander (2023), we consider...

Overview text [To be updated...]

2.2 Data Measurement [info need update]

In this study, polling data provides insight into public opinion across different states, measuring support for candidates ahead of the 2024 U.S. Presidential election. The data originates from various reputable polling organizations, each employing distinct methodologies that influence data quality and representativeness. Understanding these methodologies is crucial for interpreting polling data accurately. The primary categories of measurement methods include:

1. **Live Telephone Interviews** Some polling organizations, such as Selzer & Co. and Marist College, conduct live telephone interviews, reaching respondents on both landlines and cell phones. This method often provides more accurate and in-depth data because live interviewers can clarify questions, verify demographic details, and increase respondent engagement. However, it is costlier and time-intensive, leading to smaller sample sizes and potentially fewer surveys. In the real world, live telephone interviews offer insights from a cross-section of the population, though they may over-represent older individuals who are more likely to answer landline calls.
2. **Interactive Voice Response (IVR)** Pollsters like Trafalgar Group and Rasmussen Reports frequently use IVR, or “robo-polling,” where an automated system asks questions over the phone. IVR is more affordable and efficient for reaching large numbers of people. However, because respondents interact with a machine rather than a live person, they may feel less invested in the survey, potentially impacting response rates and depth of engagement. IVR allows pollsters to obtain timely snapshots of voter intentions, though it may miss out on nuanced responses that a live interviewer could capture.
3. **Online Panels** YouGov and Ipsos are known for relying on online panels, which recruit participants via web platforms to answer surveys electronically. This method allows for rapid data collection from a broad audience and is relatively cost-effective. However, online surveys can suffer from self-selection bias, as individuals who are more politically active or internet-savvy are more likely to participate. Despite this, online panels remain popular due to their scalability and cost-efficiency, providing a contemporary approach to measuring public sentiment.
4. **Mixed-Mode Surveys (Telephone and Online)** To balance the benefits and limitations of each mode, organizations like The New York Times/Siena College and Public Policy Polling employ mixed-mode surveys. By combining online and telephone interviews, these pollsters aim to broaden demographic reach and mitigate biases inherent in any single method. Mixed-mode surveys improve representativeness and accommodate diverse respondent preferences, providing data that may better reflect the full spectrum of public opinion.
5. **Specialized In-Person and University-Affiliated Research** Some university-affiliated pollsters, such as Quinnipiac University and University of New Hampshire Survey Center, incorporate face-to-face interviews or demographic-specific focus groups for deeper insights. These methods can yield highly detailed responses and foster a comprehensive understanding of specific voter groups, though they are resource-intensive and typically used for more specialized research rather than general polling.

These varied data collection methods directly affect how public opinion is measured and interpreted. For instance, live interviews may capture a more detailed view of voter sentiment but often exclude younger demographics who primarily use mobile devices or interact online. Meanwhile, IVR and online panels make large-scale data collection affordable and timely,

though they may lack depth and introduce potential self-selection biases. Mixed-mode approaches provide a balanced perspective, catering to both traditional and digital respondents, which may yield more representative data.

The choice of methodology reflects each pollster’s objectives and resources, with implications for accuracy and data interpretation. By categorizing pollsters based on their methodologies, we can contextualize the polling data in this study, recognizing each method’s strengths and limitations. This understanding is essential for modeling real-world voter behavior accurately, as it acknowledges how data collection methods impact polling results.

2.3 Outcome variables

According to Table 1, the outcome variables include critical components for analyzing polling data, such as State, Pollster, Pollscore, Candidate Name, Percentage (pct), Sample Size, Start Date, End Date, and Days to Election. These variables provide the foundation for estimating actual vote shares by capturing regional differences, pollster credibility, candidate-specific support, and timing in relation to Election Day. Together, they ensure a nuanced approach to election forecasting that accounts for variability in poll methodology and voter sentiment shifts over time.

Table 1: Outcome Variables

Outcome Variable	Example
State	Wisconsin
Pollster	RMG Research
Pollscore	0.4
Candidate Name	Donald Trump
Percentage (pct)	48.5%
Sample Size	789
Start Date	2024-9-24
End Date	2024-10-16
Days to Election	20

Above outcome variables Table 1

2.4 Predictor variables

This study includes several key predictor variables that provide insights into regional voting patterns, polling methodologies, and the timing of voter sentiment shifts. Each variable plays

a unique role in forecasting vote shares for the 2024 U.S. Presidential election. Below shown in Table 2 are detailed descriptions of each predictor.

Table 2: Predictor Variables

Predictor Variable	Example
State	Georgia
Pollster	InsiderAdvantage
Pollscore	0.6
Candidate Name	Kamala Harris
Percentage (pct)	47%
Sample Size	800
Days to Election	24

2.4.1 State

State (`state`) is a categorical variable that represents the U.S. state in which each poll was conducted. States have distinct political dynamics, influenced by regional issues, demographics, and historical voting patterns. Including the state variable allows the model to account for these differences, enabling predictions that are sensitive to each state’s unique electoral landscape.

2.4.2 Pollster

Pollster (`pollster`) identifies the organization conducting each poll. Different polling organizations may use varied methodologies, such as sample selection, weighting, and question phrasing, which can introduce systematic biases. By including the pollster as a predictor, the model can adjust for potential biases or methodological differences, providing more accurate and standardized vote share estimates.

2.4.3 Candidate Name

Candidate Name (`candidate_name`) is a categorical variable that identifies the candidate for whom each poll measures support. Including candidate names allows the model to assess each candidate’s baseline popularity and identify variations in support across different demographics and states. This variable enables comparisons between major contenders, such as Donald Trump and Kamala Harris, as well as other candidates, to forecast overall vote shares.

2.4.4 Percentage of Votes

Percentage of Votes (`pct`) represents the percentage of respondents in each poll who indicated support for a specific candidate. As a primary predictor, this variable directly informs the model about the current level of support for each candidate at the time the poll was conducted. By converting these percentages into vote share estimates, the model can project likely election outcomes based on current polling data.

2.4.5 Sample Size

Sample Size (`sample_size`) denotes the number of respondents included in each poll. Larger sample sizes generally increase the reliability of a poll, as they reduce the margin of error and are more likely to represent the broader population. By incorporating sample size as a predictor, the model can weigh polls based on their reliability, giving more importance to larger, more robust samples.

2.4.6 Pollscore

Pollscore (`pollscore`) variable serves as a measure of each pollster's historical accuracy or quality, reflecting how reliable or unbiased the pollster's results have been over time. A higher pollscore likely indicates a pollster with a better track record of accurate predictions, suggesting their polling methodologies or adjustments may be more robust.

Figure 1 shows the distribution of poll scores. The plot shows two primary clusters, with peaks around (-1) and near (0). This bimodal distribution suggests that pollsters' ratings might tend to fall into two categories, potentially reflecting different methodologies, biases, or levels of reliability among pollsters. For example, scores clustered around (-1) could represent pollsters with a more consistent lean or bias in one direction, while scores around (0) may indicate neutral or balanced ratings. These clusters highlight the importance of adjusting for pollster reliability when aggregating data for election predictions, as different pollsters contribute varying levels of accuracy and bias. Incorporating such adjustments can enhance the model's ability to produce a balanced and accurate forecast, providing a more nuanced view of candidate support that takes into account the quality and tendencies of each poll source.

2.4.7 Days to Election

The variable days to election represents the days remaining from the poll's end date to Election Day. It is calculated using the `end_date` variable, by subtracting the `end_date` from the election date, by assuming the election date is November 5 2024.

`days_to_election`

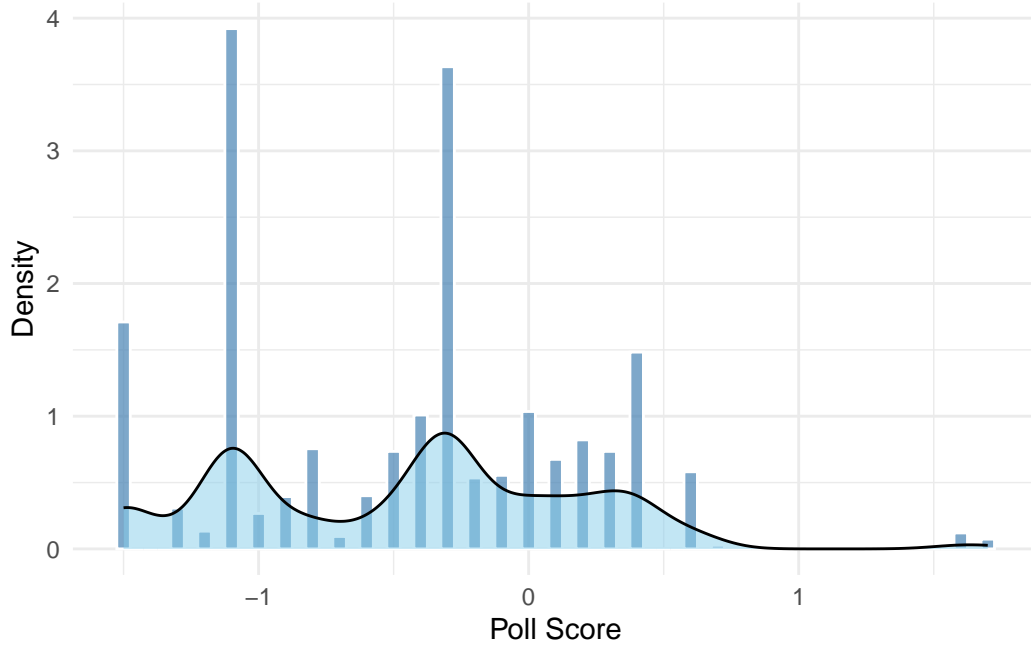


Figure 1

3 Model

To predict the actual election vote share for each candidate in each state while accounting for variations by pollster and other poll-specific factors. Background details and diagnostics are included in Appendix B. [To be updated...]

3.1 Model set-up

A Hierarchical Bayesian Model is utilized to predict the actual election vote for each candidate in each state while accounting for variables by pollster.

The model prediction utilizes the following predictor variables:

- **State** (`state`): Include as a categorical term to capture regional variations.
- **Pollster** (`pollster`): Include as a categorical variable for different polling effects.
- **Pollscore** (`pollscore`): A numerical variable reflecting how reliable the pollster's results have been over time.
- **Candidate Name** (`candidate_name`): Include as a categorical feature or model separately for each candidate.
- **Percentage of Votes by Poll in State** (`pct`): Use as a primary predictor of actual vote share, with a smooth term to allow for non-linear effects.

- **Sample Size** (`sample_size`): Include as a predictor or weight to reflect poll reliability.
- **Days to Election** (`end_date`): capture trends in support leading up to the election.

Let:

- y_{ijk} : The target variable, representing the actual election vote share for candidate k in state i with pollster j .
- pct_{ijk} : The observed polling percentage for candidate k in state i by pollster j .
- sample_size_{ijk} : The sample size of the poll, which helps in weighing the poll reliability.
- $\text{days_to_election}_{ijk}$: Derived as the days remaining until the election from the poll's end date, capturing the effect of recency.
- pollscore_{ijk} : The reliability score for each pollster j , reflecting the accuracy of the pollster's historical results.

The model takes the form of the following equation:

$$y_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta \cdot \text{pct}_{ijk} + \eta \cdot \text{sample_size}_{ijk} + \theta \cdot \text{days_to_election}_{ijk} + \zeta \cdot \text{pollscore}_{ijk} + \epsilon_{ijk} \quad (1)$$

where:

- $\alpha_i \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$: State-level random effect for each state i , capturing regional variations in voting patterns
- $\beta_j \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$: Pollster-level random effect for each pollster j , accounting for systematic biases or differences in polling methods.
- $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$: Error term, accounting for random noise.
- γ_k : Candidate fixed effect for each candidate k , representing baseline support across states and pollsters.
- δ : Coefficient for Percentage of Votes by Poll (pct_{ijk}), reflecting how poll support translates to actual vote share.
- η : Coefficient for Sample Size (sample_size_{ijk}), weighing polls based on their reliability.
- θ : Coefficient for Days to Election ($\text{days_to_election}_{ijk}$), capturing the trend in support as the election date approaches.
- ζ : Coefficient for Pollscore (pollscore_{ijk}), accounting for the reliability of the pollster's historical performance.

3.1.1 Interpretation of Parameters

- α_i : Captures state-specific effects, adjusting the baseline vote share prediction based on regional differences.
- β_j : Accounts for systematic biases or differences in methodologies across pollsters.
- γ_k : Provides an overall baseline effect for each candidate, independent of state or pollster.
- δ : Measures how closely polling support translates to actual vote share.
- η : Adjusts the model's sensitivity to polls based on sample size, giving more weight to larger polls.
- θ : Captures how support trends change as the election date approaches.
- ζ : Incorporates the reliability of the pollster based on historical performance, weighing polls by the pollster's accuracy.

3.1.2 Prior Distributions

- $\alpha_i \sim \mathcal{N}(0, 2.5)$: State-level random effect prior for each state i .
- $\beta_j \sim \mathcal{N}(0, 2.5)$: Pollster-level random effect prior for each pollster j .
- $\gamma_k \sim \mathcal{N}(0, 2.5)$: Candidate fixed effect prior for each candidate k .
- $\delta, \eta, \theta, \zeta \sim \mathcal{N}(0, 1)$: Coefficients for polling percentage, sample size, days to election, and pollscore.
- $\sigma \sim \text{Exponential}(1)$: Prior for the standard deviation of the error term.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.3 Model Justification

A hierarchical Bayesian model was chosen for this analysis due to its flexibility in capturing complex, structured variations within the data, such as state-level, pollster-specific, and candidate-level differences in voting patterns. Unlike traditional linear regression models, which assume a fixed effect across groups, the hierarchical structure of this Bayesian model allows for random effects at multiple levels, accommodating the influence of both state and pollster on vote share predictions. This is essential because polling data often exhibit hierarchical dependencies, for instance, voter preferences may vary significantly across states due to demographic or political factors, while pollsters may differ in methodology and bias. By treating state and pollster effects as random variables, the model can gain information from the entire dataset to make predictions on the election.

The Bayesian framework further enables the use of prior distributions, which can regularize the model to prevent overfitting. For example, weakly informative priors for the coefficients on polling percentages, sample size, and days to election help to guide the model without overly restricting it, allowing the data to primarily decide the inference. Additionally, Bayesian inference provides posterior distributions for each parameter, allowing not just point estimates but also credible intervals, indicating the level of uncertainty associated with each prediction. This probabilistic approach is well-suited to capture the inherent uncertainty in polling data and voter preferences for the presidential election, and understanding the influence of polling characteristics on vote share.

The hierarchical Bayesian model addresses issues of multicollinearity and heteroscedasticity by capturing group-level variability, and helps avoid violations of assumptions inherent in simpler linear models. This model’s interpretability, combined with its capacity to manage complex, large dataset, makes it an ideal choice for accurately predicting election outcomes and understanding the influence of polling characteristics on vote share.

1. State Effects (α_i)

Table 3: State Effects

State	Pollster Rating Name	State Effect
Alabama	John Zogby Strategies	55.6
Alabama	WPA Intelligence	57.0
Alaska	Alaska Survey Research	45.6
Alaska	Cygnal	48.0
Alaska	Data for Progress	53.0

Table 3 illustrates the state effect. The state effects in this Hierarchical Bayesian Model provide adjustments to account for unique voting patterns across different states, enhancing the accuracy of election predictions. For instance, Alabama shows relatively high state effect values, like 55.6 from John Zogby Strategies, suggesting an upward adjustment that aligns with Alabama’s regional voting tendencies. In contrast, Alaska displays lower effect values, such as 45.6 from Alaska Survey Research, indicating a possible regional leaning away from certain candidates.

Arizona’s effect estimates vary across pollsters, with higher values like 49.2 from AtlasIntel and lower values around 39.5 from Blueprint Polling. These adjustments reflect Arizona’s dynamic political environment, where regional preferences can sway support for candidates.

By incorporating these state-specific adjustments, the model better aligns with local voting behaviors, capturing regional nuances and improving the reliability of its predictions.

2. Pollster Effect (β_j)

Table 4: Pollster Effects

Pollster Rating Name	Pollster Effect
1892 Polling	0.1
ABC News/The Washington Post	-1.2
Alaska Survey Research	-0.1
AtlasIntel	-0.8
BK Strategies	0.3

The pollster effects in Table 4 in this model adjust for systematic biases or methodological tendencies in polling data from specific organizations, standardizing vote share predictions across diverse sources. For example, ABC News/The Washington Post has a notable negative adjustment of -1.2, suggesting that this pollster’s data tends to overestimate support for certain candidates, prompting a downward correction. In contrast, BK Strategies has a positive effect of 0.3, indicating a slight upward adjustment, which may be due to an observed underestimation in their polling results. Other pollsters, such as 1892 Polling with a minor positive effect of 0.1, and Alaska Survey Research with a small negative effect of -0.1, receive subtle adjustments to bring their estimates in line with the model’s overall predictions. By incorporating these pollster-specific effects, the model accounts for known polling biases, enhancing its predictive reliability.

3. Candidate Fixed Effect (γ_k)

Table 5: Candidate Fixed Effects

Candidate Name	Candidate Fixed Effect (γ_k)
Donald Trump	45.3
Kamala Harris	47.2

Shown in above Table 5, the candidate fixed effects in this model capture the baseline level of support for each candidate, reflecting their overall popularity or favorability across the polls. For example, Kamala Harris has a candidate fixed effect of 47.22, while Donald Trump has a slightly lower effect of 45.28. These values suggest that, on average, Harris may have a marginally higher baseline level of support in the polling data compared to Trump. These fixed effects allow the model to adjust for inherent differences in candidate popularity, providing a more accurate prediction by factoring in each candidate’s overall favorability independent of state, pollster, or specific poll characteristics. This adjustment is crucial for capturing the candidates’ general standing in the electorate, contributing to more reliable forecasting.

4. Adjust predictions by poll score and sample size ($\delta \cdot \text{pct}_{ijk} + \eta \cdot \text{sample_size}_{ijk}$)

Table 6: Weighted Predictions by Poll Score and Sample Size

State	Candidate Name	Weighted Percentage	Total Weight
Alabama	Donald Trump	57	-50
Alaska	Donald Trump	52	-4031
Alaska	Kamala Harris	43	-934
Arizona	Donald Trump	47	-76820
Arizona	Kamala Harris	46	-34822

Table 6 shows the adjusted predictions by poll score and sample size. The table of weighted predictions shows adjusted support percentages for each candidate by state, incorporating both poll score and sample size. The “Weighted Percentage” column represents the predicted support level for each candidate after applying these weights, while the “Total Weight” column indicates the cumulative influence of poll score and sample size. For instance, in Alaska, Donald Trump has a weighted percentage of 51.57 with a total weight of -4031.0, while Kamala Harris has a lower weighted percentage of 43.32 with a weight of -934.4. In Arizona, the weights are significantly higher in absolute value, such as -76819.6 for Trump, reflecting the model’s prioritization of more reliable and larger polls in this state. These adjustments ensure the model’s predictions are robust, accurately reflecting candidate support by accounting for the quality and scale of each poll.

5. Adjustment of weighted predictions by accounting for recency ($\theta_{\text{days_to_election_ijk}}$)

Table 7: Weighted Predictions with Recency Adjustment

State	Candidate Name	Weighted Percentage	Total Weight
Alabama	Donald Trump	57.0	-0.1
Alaska	Donald Trump	51.9	-33.4
Alaska	Kamala Harris	43.6	-21.6
Arizona	Donald Trump	48.7	-915.1
Arizona	Kamala Harris	46.7	-716.5

Table 7 illustrates the model’s adjustment of weighted predictions by accounting for recency. This adjustment decreases the influence of older polls while giving more recent polls greater weight, enhancing prediction relevance as Election Day approaches. Here, the “Total Weight” column reflects the combined impact of poll score, sample size, and recency. For example, in Arizona, Donald Trump’s weighted percentage is 48.66 with a significant total weight of -915.07, indicating a substantial recent adjustment due to proximity to the election. Kamala

Harris’s weighted percentage in the same state is 46.66 with a weight of -716.47, reflecting similar adjustments. This recency factor ensures that the model prioritizes fresher data, which better captures evolving voter sentiment, thereby making predictions more accurate as the election nears.

6. Pollscore Reliability (ζ)

Pollster reliability represents an adjustment factor that accounts for the credibility and historical accuracy of each pollster’s results. Pollster reliability scores are typically calculated based on a pollster’s past performance, including how closely their predictions align with actual election outcomes. This score reflects the likelihood that a given poll’s results accurately represent voter sentiment. By either increasing or decreasing it based on the reliability score; a higher score would yield a more positive adjustment, while a lower score would lead to a reduction. This adjustment aims to reduce the influence of less reliable polls, thereby improving the model’s overall accuracy by weighting predictions according to the quality of the polling data.

4 Results

The final predictions for the 2024 presidential election indicate a close race between the two candidates. The model results are summarized in **?@tbl-modelresults.Based** on the weighted average calculations, which incorporate factors such as state effects, pollster reliability, candidate-specific adjustments, and recent polling trends, Donald Trump is predicted to receive approximately 46.56% of the vote, while Kamala Harris is predicted to receive around 45.65%. These predictions account for adjustments based on sample size, recency of the polls, and reliability scores, ensuring that more recent and reliable polls have a greater influence on the final outcome. The model also applies a slight adjustment to account for known biases, providing a robust estimation of each candidate’s standing. Although these predictions offer a snapshot of the current polling landscape, they should be interpreted with caution, as last-minute changes in voter sentiment could still impact the final outcome.

Table 8: Final Predictions for Trump and Harris

Candidate	Average Prediction (%)
Donald Trump	46.56
Kamala Harris	45.65

4.1 Predicted Electoral Outcomes

The outcome pie chart Figure 2 below shows the predicted vote share distribution between Donald Trump and Kamala Harris, along with a comparison that includes votes projected for other candidates or undecided voters.

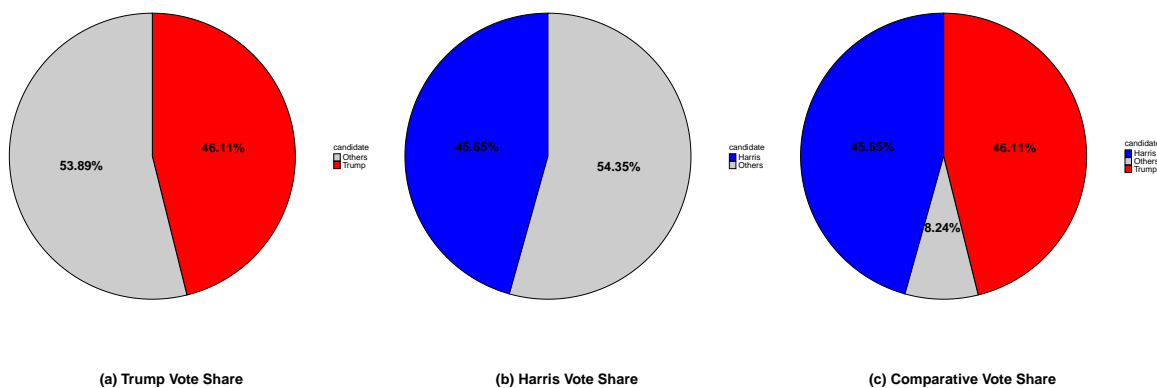


Figure 2: 2024 Presidential Predictions

The first pie chart, labeled (a) Trump Vote Share, displays Donald Trump’s projected support level if we consider his vote share against all other potential candidates. According to the model, Trump is estimated to hold 46.11% of the vote, while 53.89% falls under the “Others” category. This result suggests that although Trump has a strong share of the projected votes, he does not exceed the 50% threshold in isolation, indicating there remains a sizable portion of voters who may support other candidates or remain undecided.

The second pie chart, (b) Harris Vote Share, presents Kamala Harris’s estimated support level against all other candidates. Harris is projected to receive 45.65% of the vote, with 54.35% categorized as “Others.” Similar to Trump’s results, Harris shows a solid portion of support, though she too falls short of the majority. This chart highlights that Harris’s vote share is almost on par with Trump’s but leaves a considerable portion of votes uncommitted to either of the main candidates.

The third pie chart, (c) Comparative Vote Share, illustrates the direct comparison between Trump, Harris, and the “Others” category. In this model, Trump holds a slight lead with 46.11%, while Harris closely follows with 45.65%. Meanwhile, 8.24% of the vote remains attributed to “Others,” representing voters who may support third-party candidates or are undecided. This small lead for Trump, amounting to just a 0.46% difference, emphasizes the closeness of the race. The presence of a significant “Others” portion could prove crucial if those votes shift toward one of the primary candidates, potentially altering the predicted outcome.

Overall, these results suggest a highly competitive election, with Trump holding a marginal

edge over Harris. The close percentages indicate that even slight changes in voter preferences could impact the election result, and the 8.24% of “Others” reflects the potential influence of undecided or third-party voters. This tight margin underscores the importance of swing votes in determining the final outcome, with neither candidate showing a decisive lead at this stage.

4.2 Electoral outcome by states

Model predictions for key states reveals a competitive landscape for the 2024 presidential election, with slight advantages for each candidate in various swing states. Donald Trump is projected to lead in most states, including Arizona, Florida, Michigan, Nevada, Ohio, Texas, and Wisconsin. However, Kamala Harris is predicted to secure a narrow edge in Georgia, North Carolina, and Pennsylvania, reflecting her competitiveness in these critical battlegrounds. This distribution indicates that while Trump may maintain a slight overall advantage in the swing states, Harris still shows potential to contest and even lead in pivotal regions, underscoring the unpredictable nature of the election. The relatively close percentages in each state illustrate the highly competitive environment and the significance of voter turnout and campaign efforts in determining the final outcome.

Table 9: Model Prediction Results for Key States

State	Trump Predicted (%)	Harris Predicted (%)	Predicted Winner
Arizona	47.67	46.50	Trump
Florida	49.16	47.55	Trump
Georgia	46.52	47.53	Harris
Michigan	47.22	45.72	Trump
Nevada	47.36	47.22	Trump
North Carolina	47.17	47.40	Harris
Ohio	49.02	48.50	Trump
Pennsylvania	45.67	45.85	Harris
Texas	47.71	46.22	Trump
Wisconsin	47.19	47.06	Trump

Table 9 indicates a close race for the 2024 presidential election, with Donald Trump slightly leading in several key swing states. Trump is projected to hold a narrow lead in states such as Arizona (47.67% vs. 46.50%), Florida (49.16% vs. 47.55%), Michigan (47.22% vs. 45.72%), and Ohio (49.02% vs. 48.50%), highlighting his advantage in these critical regions. Harris, on the other hand, is estimated to lead in Georgia (47.53% vs. 46.52%), North Carolina (47.40% vs. 47.17%), and Pennsylvania (45.85% vs. 45.67%), demonstrating her strength in these areas.

This pattern underscores the competitive nature of the election, where both candidates have strongholds, but the overall lead may still be within reach for either. The close percentages—often within 1-2% margins—highlight the importance of final campaign efforts and voter turnout in swaying these crucial states.

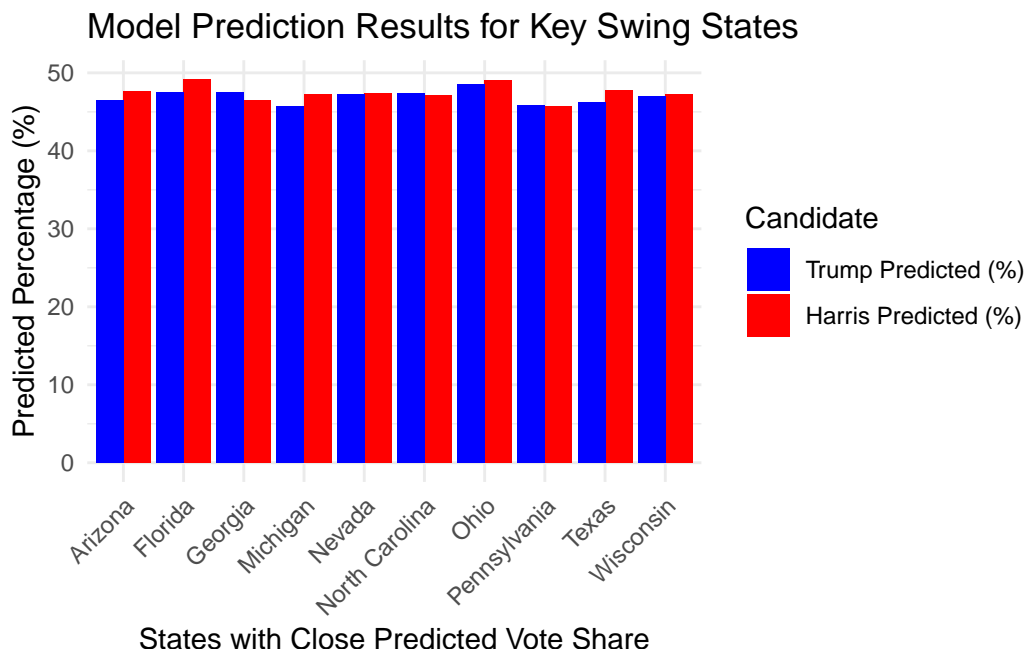


Figure 3

The bar plot Figure 3 shows the model’s predicted vote percentages for Trump and Harris in several competitive states for the 2024 U.S. Presidential Election, with each state represented by bars for each candidate’s projected support. The close alignment of the bars across many states reflects the high level of competition in these regions. Georgia, North Carolina, Pennsylvania, Ohio, and Nevada are identified as key swing states this year due to the narrow predicted vote margins between Trump and Harris, indicating that the outcome could go either way in these areas. This small difference in projected support makes these states pivotal battlegrounds, as any shifts in voter sentiment could significantly impact the election’s overall result. As a result, both campaigns are likely to concentrate their resources and efforts on these swing states to secure a decisive advantage.

The updated map Figure 4 above provides a visual forecast of the 2024 U.S. Presidential Election by state, using weighted polling data and recent trend adjustments. Each state is color-coded to show the projected winner: red for states likely to support Trump, blue for those leaning towards Harris, and gray for swing states where the predicted vote margins are extremely close. This map shows a competitive electoral landscape. Harris is projected to perform well in traditionally Democratic regions on the West Coast and in the Northeast,

2024 U.S. Presidential Election Prediction by State Predicted Winner by State with Key Swing States Highlighted

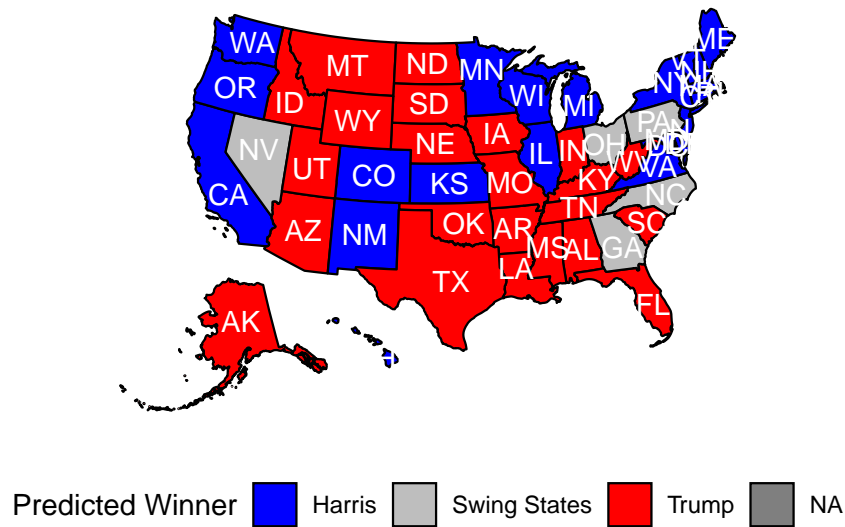


Figure 4

including states like California, New York, and Illinois. Trump demonstrates strong support across the South, Midwest, and several central states, such as Texas, Florida, and the Dakotas.

The swing states highlighted in gray—include states like Georgia, North Carolina, Pennsylvania, Ohio, and Nevada are classified as swing because the polling indicates very narrow vote margins, making them highly competitive. Given the closeness of the race in these states, they could feasibly shift in favor of either candidate, positioning them as key battlegrounds in the election. The outcomes in these swing states are likely to be decisive in determining the overall result, highlighting their critical role in shaping the candidates' campaign strategies.

5 Discussion

5.1 State Effect

5.2 Pollster and Pollscore Effect

Figure 5 illustrates the relationship between poll score reliability and percentage support for Kamala Harris, showing that her support levels are fairly consistent across varying levels of poll reliability. The points are widely spread, with support percentages ranging from around

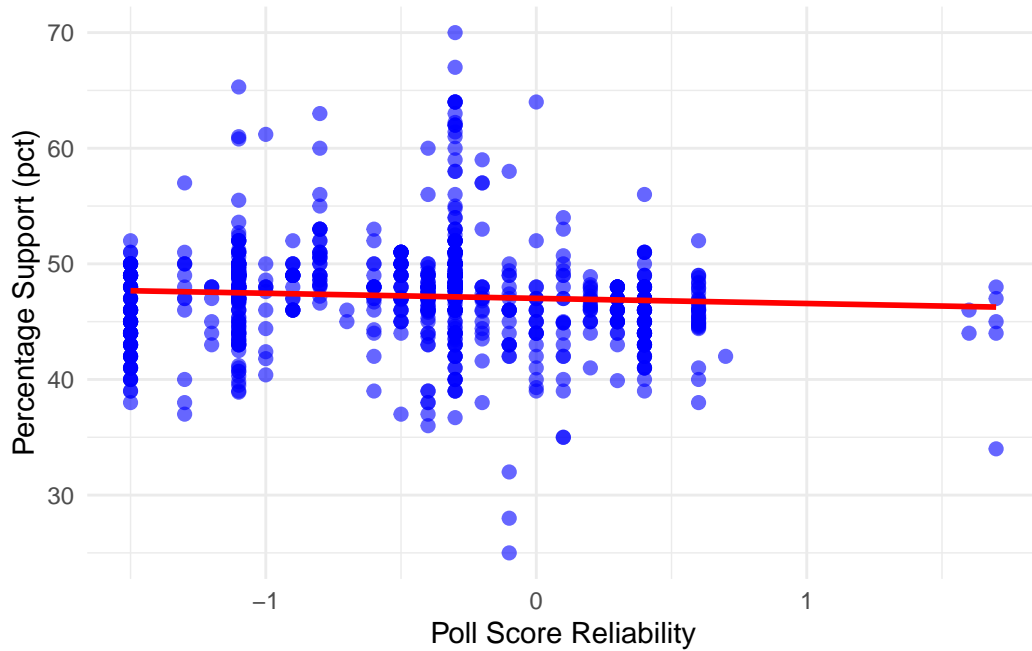


Figure 5: Relationship Between Poll Score Reliability and Support Percentage for Kamala Harris

30% to over 60%, though most are concentrated around the 45-55% range. The trend line is nearly flat, indicating a negligible correlation between poll score reliability and support for Harris. This suggests that the reliability of the pollster has little influence on Harris's reported support levels, implying that her support is relatively stable across both high- and low-reliability polls. This stability may indicate that adjustments for poll score reliability might have minimal impact on the aggregated support predictions for her.

Figure 6 shows that Donald Trump's percentage support is widely distributed across different poll score reliability levels, with support percentages spanning from around 20% to above 60%, though most values center around 40-50%. The slight negative slope of the trend line suggests a minimal decrease in support as poll reliability increases, but this effect is weak, indicating that pollster reliability does not significantly influence the reported support for Trump. This suggests that Trump's support remains relatively stable across both high- and low-reliability polls, implying that adjustments for poll score reliability may not substantially impact predictions for him in election models.

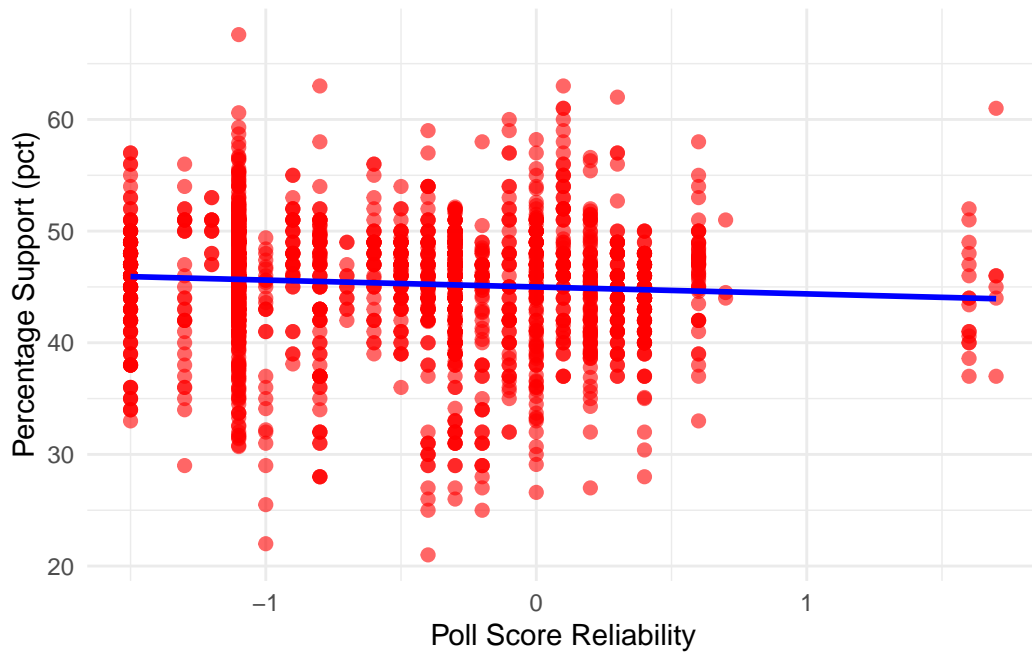


Figure 6: Relationship Between Poll Score Reliability and Support Percentage for Donald Trump

5.3 Days to Election Effet

5.4 Weaknesses and Next Steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. <https://open.toronto.ca/dataset/deaths-of-shelter-residents/>.