

Forecasting the 2024 US Presidential Election*

Analyzing Demographic Patterns and Predicting Swing States

Tina Kim

David Flores

Kevin Shao

October 31, 2024

Our analysis of swing state polls for the 2024 U.S. Presidential Election shows a close race. Republicans lead in most key states like Michigan, Georgia, Nevada, Wisconsin, and Pennsylvania, while Democrats have a slight edge in Arizona. The national average also leans Republican, suggesting a Republican win in the 2024 Election.

1 Introduction

Overview paragraph: Provide a brief overview of the context of the upcoming US presidential election and the importance of forecasting its results.

Estimand paragraph: Define the estimand clearly (e.g., predicting the probability of a candidate winning the election based on poll data).

Results paragraph: Summarize the key findings of the model, highlighting its predictive accuracy and implications.

Why it matters paragraph: Explain the broader significance of accurately forecasting election results for politics, society, and policymaking.

Telegraphing paragraph: The remainder of this paper is structured as follows: Section 2 details the data and measurement process; Section 3 covers model development and results; Section 4 discusses implications and future steps. The remainder of this paper is structured as follows. Section 2....

*Code and data are available at: <https://github.com/DavidFJ207/USPresidentialForecast>

2 Data

2.1 Overview

State	Democrat Pct	Republican Pct	Voted Biden 2020	Voted Trump 2020
Arizona	48.2	50.5	46.0	45.2
Georgia	49.8	49.7	44.5	44.3
Michigan	49.7	49.6	57.9	26.8
Nevada	49.2	48.1	57.9	26.8
New Hampshire	50.9	47.2	51.8	35.8
North Carolina	48.2	50.2	46.0	45.2
Pennsylvania	49.0	50.7	45.3	44.9
Wisconsin	49.2	49.9	58.2	28.7

Figure 1

State	High Educated	Low Educated	Female	Male	Nonbinary	Caucasian	Minority Ethnicity
Arizona	22.1	24.3	51.8	46.4	1.8	80.2	19.8
Georgia	23.8	24.1	49.8	49.5	0.7	86.5	13.5
Michigan	27.2	22.6	53.0	46.2	0.8	60.3	39.7
Nevada	27.2	22.6	53.0	46.2	0.8	60.3	39.7
New Hampshire	30.1	17.7	52.5	46.7	0.8	68.3	31.7
North Carolina	22.1	24.3	51.8	46.4	1.8	80.2	19.8
Pennsylvania	25.8	21.8	53.7	44.6	1.7	60.9	39.1
Wisconsin	28.7	19.5	51.9	47.1	1.0	49.1	50.9

Figure 2

Latest Poll Points by Party

We sourced the “Presidential General Election Polls” dataset from FiveThirtyEight (FiveThirtyEight 2024) and analyzed it using the statistical programming language R (R Core Team 2023).

Our main goal was to clean, organize, and analyze U.S. presidential election polling data to understand voter preferences by state. We selected data from credible pollsters, sorted it by state, and resolved missing or incomplete entries. Where necessary, we merged additional datasets, ensuring that we carefully handled any gaps in the data. The result was a refined dataset ready for analysis. The table below provides a summary of polling data by state, showing voter preferences by party and various demographics. It highlights key trends and differences across states, focusing on party leanings and demographic splits, as illustrated in [?@fig-predictors-pct-summary](#).

2.2 Measurement

We chose Emerson as our primary pollster, and the reasoning for this is detailed in Section 3. Emerson’s polls offered crucial information, such as polling dates, party affiliations, sample sizes, and the percentage of support for each political party.

Our analysis centered on voter preferences by party rather than individual candidates. This approach was more relevant for identifying overall trends and helped us forecast which states might swing in future elections, especially swing states. We will discuss this focus further in our results section, Section 5.

The Emerson dataset also included valuable demographic data, voting history, and approval ratings. To ensure a consistent analysis, we concentrated on questions that were uniformly asked across all states. This approach allowed us to aggregate and analyze data at the state level without being limited by state-specific polling variations.

2.3 Outcome Variables

Our initial analysis of the dataset revealed important trends in party preferences across states. In Figure 3, we show which states have the strongest leanings toward the Democratic or Republican parties. This visualization is crucial for identifying key swing states, which lie near the center of the graph and show nearly equal support for both parties. These states are especially important for predicting the 2024 election outcome.

From this analysis, we identified eight critical swing states: Michigan, Georgia, Nevada, North Carolina, New Hampshire, Wisconsin, Pennsylvania, and Arizona. These states will be central to our election prediction. Although Minnesota emerged as a potential swing state, we excluded it due to the wide range of polling data, as explained in Table 3. Additionally, we will consider the national average in our overall prediction.

Table 3: Final Summary Table of 95% Confidence Intervals for Common States

State	DEM Mean (%)	DEM Lower 95%	DEM Upper 95%	REP Mean (%)	REP Lower 95%	REP Upper 95%
Arizona	42.5	40.8	44.2	47.1	45.7	48.5
Georgia	42.8	41.0	44.7	47.8	46.7	48.9
Michigan	44.9	43.4	46.4	45.6	44.3	46.9
National	42.4	41.5	43.3	43.7	42.5	44.9
Nevada	41.8	39.6	44.0	45.9	44.6	47.2
New Hampshire	44.8	42.6	47.0	41.6	39.9	43.4
North Carolina	43.2	40.4	46.0	48.6	47.7	49.6

Table 3: Final Summary Table of 95% Confidence Intervals for Common States

State	DEM Mean (%)	DEM Lower 95%	DEM Upper 95%	REP Mean (%)	REP Lower 95%	REP Upper 95%
Pennsylvania	44.2	42.5	45.9	47.6	46.6	48.6
Wisconsin	44.4	42.7	46.0	46.5	45.2	47.8

Beyond identifying swing states, it’s also important to understand the differences between the strongest Democratic (“blue”) and Republican (“red”) states. The bar graph in Figure 4 highlights these differences. It focuses on key factors that create the biggest contrasts between Democratic-leaning and Republican-leaning states, helping to show how these factors differ between the most extreme red and blue states.

To start, the graph features two bars, one for Democratic points and one for Republican points, reflecting the responses to questions asked by Emerson Pollster about voting intentions. These responses closely align with the percentages shown for each state, giving us a clear picture of voter preferences.

The Figure 4 visualization gives us valuable insight into how people’s poll responses may influence their political leanings.

Through these visualizations, we not only identify which states are most likely to swing in future elections but also better understand the public opinions driving these partisan divides.

2.4 Predictor Variables

We then take a closer look at the demographic and social factors that shape voter preferences. The predictors we analyze include education levels, gender, and past voting behavior. These variables are important because they significantly impact political leanings and can influence how a state is likely to vote.

In Figure 5, we examine how these factors correlate with party lean across U.S. states. This analysis helps identify which factors have the greatest influence in each state, providing a clearer understanding of what drives voter behavior.

The correlation matrix in Figure 5 shows how different predictors relate to party preferences. We can see which factors, like education or gender, are more strongly tied to Democratic or Republican leanings. Blue shading indicates stronger Democratic support, while red shows stronger Republican support. Key observations include higher education being linked to Democratic support and lower education favoring Trump in 2020. Minority groups generally lean Democratic, while Caucasian/White voters tend to support Trump. Past voter behavior shows Biden supporters remain Democratic, and Trump supporters disapprove of Biden. Additionally, women lean slightly more Democratic compared to men, who often lean Republican.

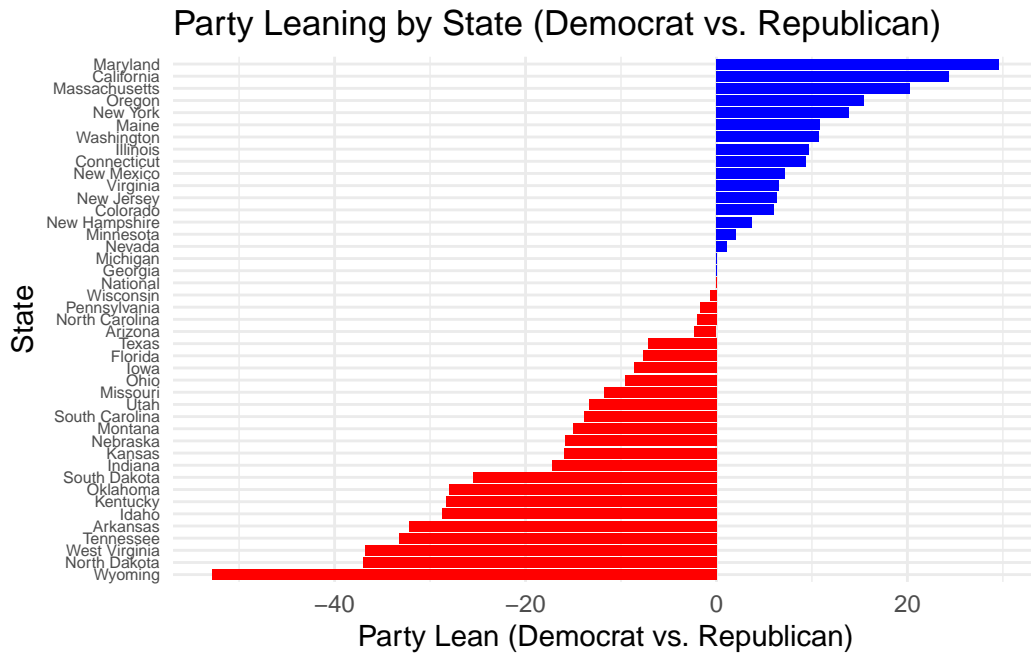


Figure 3: Example Predictor Variable Visualization

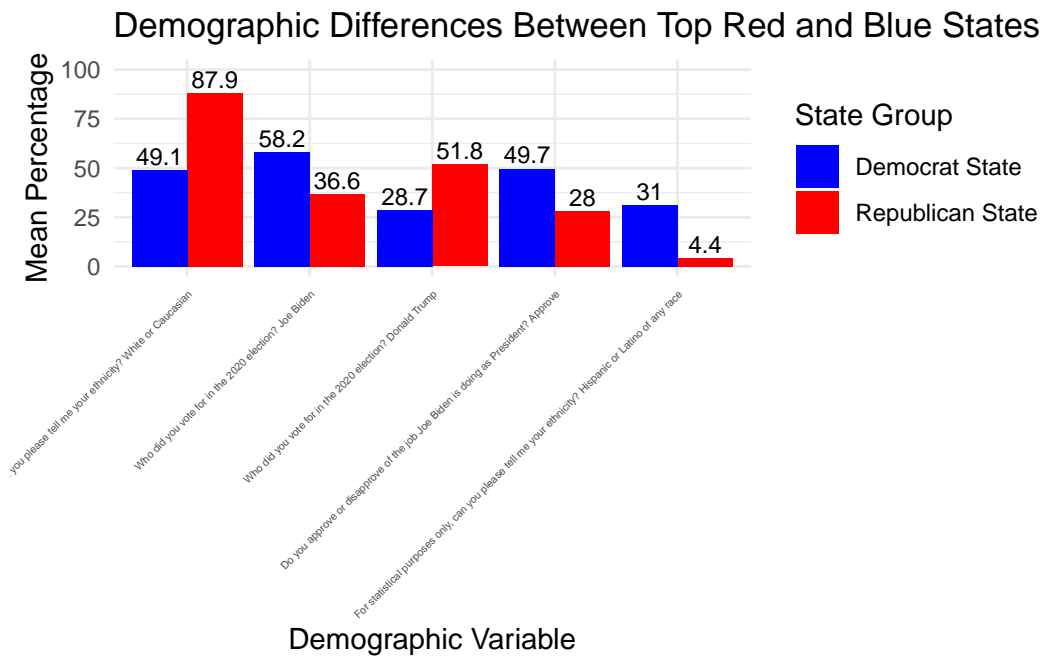


Figure 4: Demographic Differences Between Top Red and Blue States

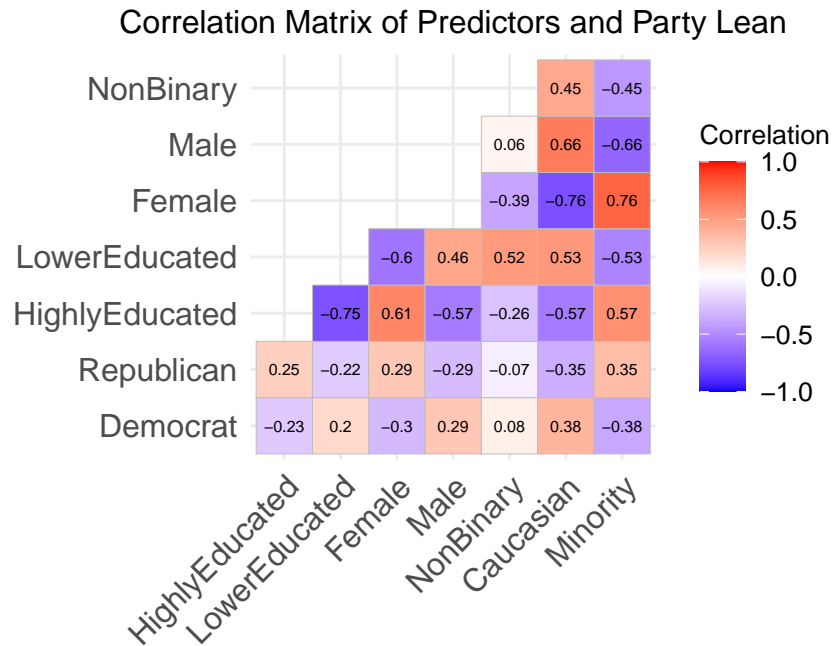


Figure 5: Example Predictor Variable Visualization

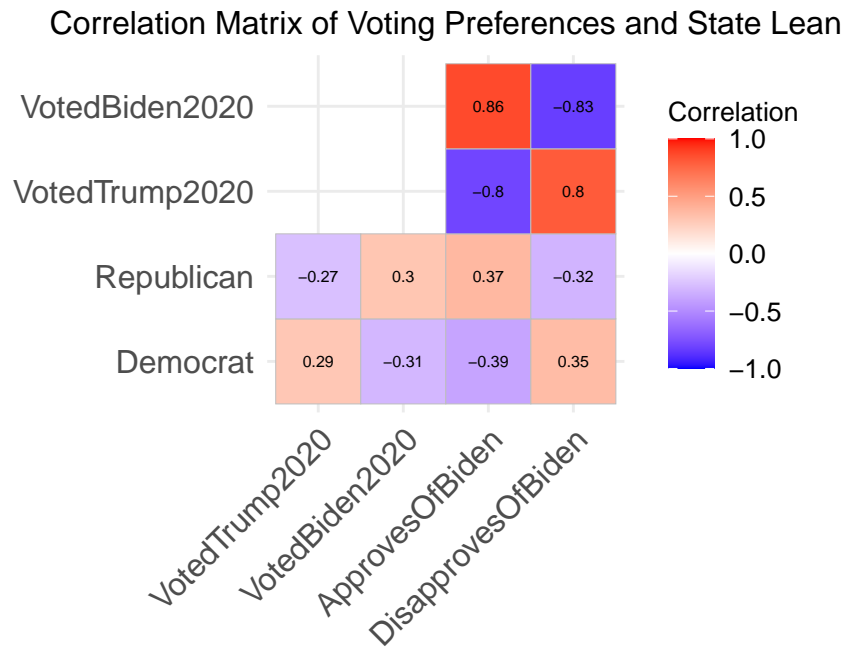


Figure 6: Example Voting and Approval Predictor Variable Visualization

Correlation Matrix of Education, Race, and Past Voting Behavior

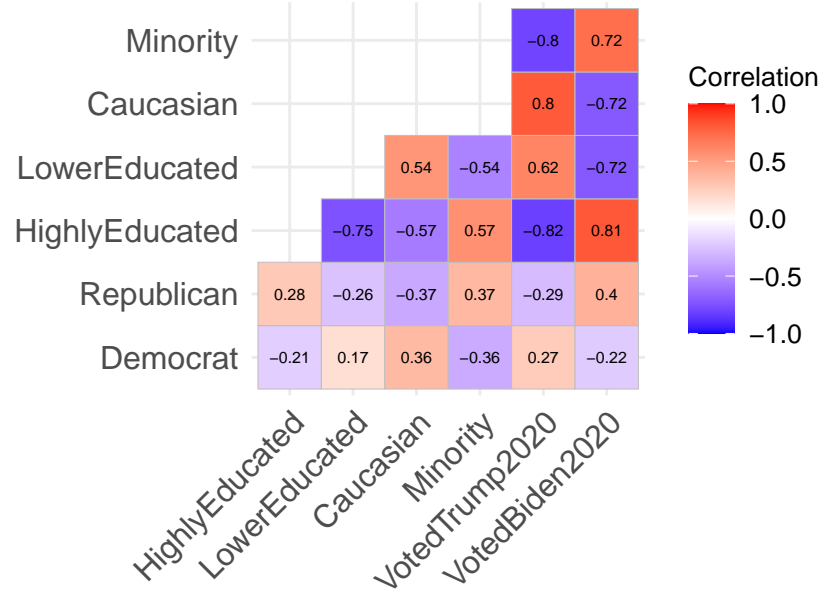


Figure 7: Correlation Matrix of Education, Race, and Past Voting Behavior

Democrat vs. Republican and Biden Approval Over Time

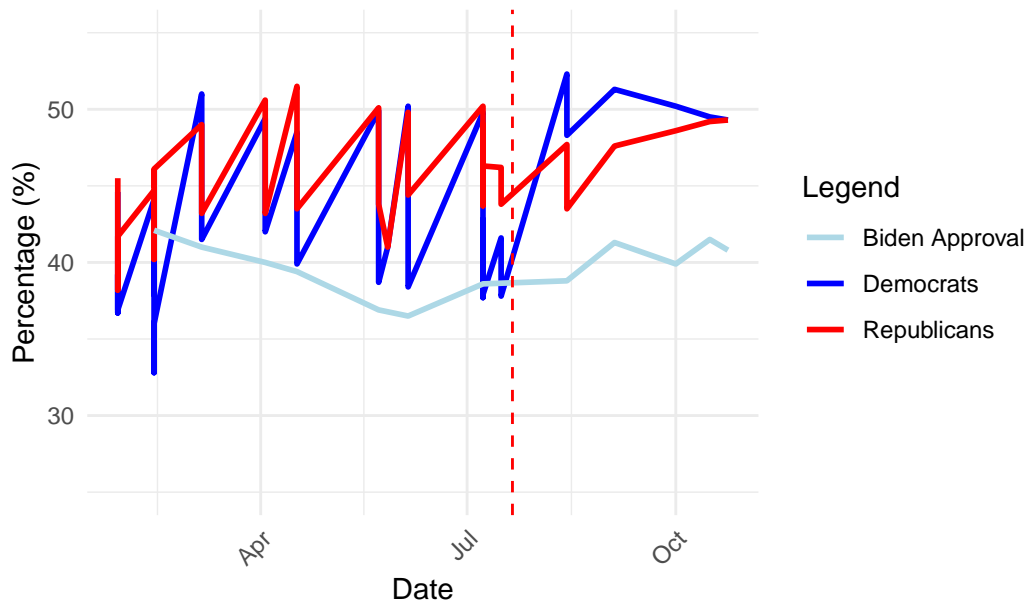


Figure 8

Lastly, we examine Democratic and Republican support alongside Biden’s approval over time, marking July 21st when Joe Biden concluded his campaign, as reported by Rabson (2024). During this period, Democratic support was initially low but increased following this change. By analyzing specific demographic factors and previous voting behavior, we aim to predict the 2024 election outcome.

Key predictor variables:

- **Education:** This measures education levels, with higher education often correlating with more Democratic support.
- **Race/Ethnicity:** This examines how different racial and ethnic groups tend to align politically, with minority groups often leaning Democratic and White/Caucasian voters leaning Republican.
- **Gender:** This captures how gender influences party preference, with women generally leaning more Democratic and men more Republican.
- **Priors (Past Voting Behavior):** This looks at who people voted for in the past, helping us understand ongoing political support or disapproval.
- **Biden’s Approval Rating:** This assesses how approval or disapproval of Joe Biden impacts Democratic support, providing insight into shifting voter sentiments.

3 Pollster Methodology Overview (Appendix A)

Append a detailed review of a chosen pollster’s methodology, survey techniques, strengths, and weaknesses.

4 Model

4.1 Model Development

The goal of our model is to forecast the likely vote outcome in 8 swing states for the 2024 US presidential election. This model will allow us to estimate the expected vote share for each candidate based on state-level characteristics and prior voting patterns, which are especially important in determining close elections.

4.2 Model Set-up

We define the model as a Bayesian linear model where the outcome, (y_i) , represents the predicted vote share in state (i) . The predictors in the model capture key demographic and historical voting patterns across swing states, and we aim to model this outcome with an

intercept, (α_i) , and a state-specific predictor coefficient, (β_{ji}) , that represents the effect of a given variable (x_{ji}) on vote share.

4.2.1 Model Components and Assumptions

- **Outcome Variable:** The predicted vote share (as a percentage) for each candidate in each state.
- **Predictors:** Variables include demographic factors (e.g., education, income, race distribution) and historical voting patterns (e.g., previous election vote shares, approval ratings).
- **Intercept:** Represents the baseline vote share in the absence of additional predictors.
- **Coefficients:** These capture the effect size of each predictor on vote share, allowing for state-specific adjustments.
- **Residual Standard Deviation:** Accounts for unexplained variation in vote share predictions, assumed to follow an exponential distribution.

4.2.1.1 Assumptions

1. **Linearity:** The relationship between each predictor and vote share is assumed to be approximately linear.
2. **Independence:** Vote share predictions for each state are independent, conditional on the predictors.
3. **Normality of Errors:** Errors in vote share predictions are assumed to be normally distributed.
4. **Prior Distributions:** We use weakly informative priors to regularize predictions without imposing strong assumptions.

4.2.2 Implementation

We run this model in R using the `rstanarm` package, which provides Bayesian estimation with sensible defaults. The default priors in `rstanarm` are chosen to be weakly informative to prevent overfitting while still regularizing our model coefficients.

4.3 Model Justification and Results

By using a Bayesian linear model, we leverage historical and demographic data to provide probabilistic estimates of vote share, giving us a distribution of possible outcomes for each state. This approach allows for uncertainty quantification around predictions, which is particularly important in close elections.

Table 4: Forecasting model results based on poll data

[!h]

Table 5: Prediction Summary for Swing States with Declared Winner

State	Democrat Points	Republican Points	Winner
Michigan	0.49	0.50	Republican
Georgia	0.49	0.50	Republican
Nevada	0.49	0.50	Republican
North Carolina	0.49	0.49	Republican
New Hampshire	0.49	0.49	Republican
Wisconsin	0.49	0.50	Republican
Pennsylvania	0.49	0.50	Republican
Arizona	0.49	0.48	Democrat
National	0.48	0.49	Republican

Our results are summarized in Table Table 6, which shows the estimated vote shares, posterior means, and credible intervals for each candidate in each swing state. These findings indicate the influence of demographic and historical factors on vote share predictions. Furthermore, the model’s predictive accuracy is evaluated through posterior predictive checks, indicating how well the model captures actual voting patterns. The implications of these findings are significant for forecasting election outcomes, particularly in identifying which states are likely to be more competitive.

5 Results

Our results are summarized in Table 6.

Person B: Visualize the model’s results and include any performance metrics (e.g., RMSE, test/train split).

6 Discussion

6.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Table 6

[!h]

Table 7: Prediction Summary for Swing States with Declared Winner

State	Democrat Points	Republican Points	Winner
Michigan	0.49	0.50	Republican
Georgia	0.49	0.50	Republican
Nevada	0.49	0.50	Republican
North Carolina	0.49	0.49	Republican
New Hampshire	0.49	0.49	Republican
Wisconsin	0.49	0.50	Republican
Pennsylvania	0.49	0.50	Republican
Arizona	0.49	0.48	Democrat
National	0.48	0.49	Republican

6.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

6.3 Third discussion point

Discuss what the model reveals about the election forecast and its potential impact on understanding voting behavior.

Person C: Discuss limitations of the model and areas for further improvement.

6.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

.1 Appendix A: Pollster Methodology Overview

Person A: Provide a detailed methodology review for the chosen pollster, including survey design, population sampling, non-response handling, etc.

.2 Appendix B: Idealized Survey Design for \$100K Budget

Person C: Develop and describe an ideal survey design for forecasting the election with a \$100K budget, and include survey link.

Sampling approach: The sampling approach we will use is stratified random sampling where we divide the population into strata based on... (age, gender, education level, etc.) Recruit Respondents: We will recruit respondents using an online survey on Google Forms so that we can minimize the cost and maximize the range of respondents we can reach. We will spend a portion of our budget (specify here) to advertise these surveys and also send out emails, with an additional monetary incentive (specify here) to encourage more participation. Data validation: IP address tracking to prevent duplicate responses. Poll Aggregation: Incorporate Bayesian Inference.

Survey Link: (Short questionnaire asking for demographic questions to be added here) Copy of Survey:

.3 Additional Data & Model Details

Include any technical details on data cleaning, model diagnostics, and posterior checks.

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...

Examining how the model fits, and is affected
by, the data

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...

Checking the convergence of the MCMC algo-
rithm

References

- FiveThirtyEight. 2024. “Dataset: US Presidential General Election Polls.” https://projects.fivethirtyeight.com/polls/data/president_polls.csv.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rabson, Mia. 2024. “Kamala Harris Would Be Democrats’ ‘Best Choice’ If Biden Doesn’t Run, Expert Says.” <https://globalnews.ca/news/10683018/kamala-harris-democratic-nomination-us-election-2/>.