A Poll-of-Polls Forecast for the 2024 U.S. Presidential Election: Analyzing Support Trends for Harris by Pollster and State*

Kamala Harris Leads with Stable Support Around 48%

Yunkyung Ko

November 4, 2024

In the anticipation of the 2024 U.S. presidential election, this paper analyzes trends in Harris's support, focusing on pollster and state-specific factors. Harris's support has remained steady around 48%, though key pollsters like Siena/NYT and battleground states such as Michigan show notable deviations. Our goal is to propose a reliable electoral prediction by examining correlations among date, pollster, and state variables and developing predictions through linear and Bayesian models.

Table of Contents

---

*Code and data are available at: https://github.com/koyunkyung/us_election_2024.

1

1 Introduction

The 2024 U.S. presidential election receives a lot of international attention due to its far-reaching implications, affecting global economy, foreign policy, and social issues like climate change (Bijune and Ha 2024). This paper aims to predict possible outcomes of the election by analyzing Kamala Harris's support in the polling data with a linear regression and a Bayesian approach.

Our estimand is Harris's support rate in polls, tracked over time and adjusted for pollster and state specific variations. Using a "pooling the polls" approach, we aim to correct for variations across different voter bases (Jackman 2024). Initial linear models show Harris's support rate remains stable at around 48% over time, though with large variability as seen in the spread of points around the fitted line. Some pollsters or states such as Siena/NYT and Michigan consistently reported higher or lower support for Harris compared to others, proposing the importance of accounting for specific polling contexts.

Beyond election forecasting, this analysis signals global stakeholders on potential shifts in the U.S. policy and strategy (CSIS 2024). As such, this study not only contributes to the domestic political discourse but also provides a valuable tool for global actors seeking to navigate the uncertainty surrounding the 2024 U.S. presidential election (CSIS 2024). The paper is organized as follows: Section 2 and Section 3 detail the data and methodology used, including filtering and modeling techniques applied to the polling data. Section 4 presents findings from the linear and Bayesian models, while Section 5 interprets broader implications. Finally, Section A covers pollster methodology and ideal survey practices.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) to analyze US presidential polling data from FiveThirtyEight (FiveThirtyEight 2024), focusing on support for Kamala Harris. The dataset, last updated on 30 October 2024, includes a wide range of poll results from various national and state-level polls, with key variables such as pollster, sample size, percentage of support for Harris, and end date of the poll. Following the guidance of Alexander (2023), we compiled the results of each opinion poll over a period of time and compared them taking into account the methodological peculiarities of polling by pollsters and geographical scope of the conducted polls.

To ensure data quality, we filtered the dataset to include only polls that measured Kamala Harris' support, with a numeric grade of pollster 2.7 or higher for reliability. We also limited the analysis to polls conducted after July 21, 2024, when Harris officially declared her candidacy, and excluded pollsters with fewer than 30 polls to focus on those with sufficient data for robust results.

In performing the analysis, we utilized several R packages. `tidyverse` (Wickham et al. 2019) was used for data manipulation and visualization and `rstanarm` (Goodrich et al. 2024), `modelsummary` (Arel-Bundock 2022) was respectively used for Bayesian modeling and generating model summaries. For visualizing results, `ggplot2` (Wickham 2016) was used and `kableExtra` (Zhu 2024) helped format tables for presentation. These packages provided a framework for efficient data processing, modeling, and reporting.

2.2 Measurement

The original dataset sourced from FiveThirtyEight (FiveThirtyEight 2024) aggregates a wide range of poll results, showing 16817 observations based on dataset available at October 30. The polls conducted by various polling organizations capture voter preferences by taking a representative sample of the electorate and asking for the voters' candidate of choice. Surveys were conducted at the state and national levels, providing the wide perspective on public feelings across the country.

In converting an individual's complex set of opinions into "plausibly measurable" (Alexander 2023) numbers, each poll attempts to capture and simplify voters' choice in a straightforward way. Polls capture a momentary snapshot of opinion, but this process is inevitably influenced by factors such as question phrasing, the timing of the poll, and contextual events. A person may have mixed or conflicting feelings about candidates, and the process of capturing these preferences in a single response imposes much error (Alexander 2023).

The outcome variable, which is intended to reflect overall support levels, is constructed from respondents' stated candidate preferences. As part of the data refinement process to enhance

the reliability of predictions, we restricted our analysis to polls with a numeric quality grade of 2.7 or higher and to pollsters with over 30 polls in the dataset. We set a minimum threshold of 2.7 for the numeric grade to focus on the top 30% of the dataset, ensuring a sufficient sample size of 5,612 observations. Additionally, we included only pollsters with at least 30 polls, reasoning that a higher number of polls from a given pollster would enhance the reliability and validity of the aggregated data.

While efforts to filter and refine the dataset mitigate some limitations, they cannot entirely eliminate the challenges inherent in measuring multifaceted concept like voter preference. Limitations including sampling error, response bias, and inconsistencies in polling methods remain. Some responses may also be distorted due to misunderstandings or incomplete answers, as respondents might interpret questions differently depending on context (Alexander 2023). This measurement process is essentially an estimation, translating complex human sentiments into quantifiable data, subject to various limitations.

2.3 Outcome variable

2.3.1 The Proportion (%) of Support that a Candidate Received in the Poll

The main variable of interest that we aim to forecast is the 'pct' variable, which represents the proportion of vote or support that a candidate received in the poll. Table 1 and Figure 1 respectively shows the summary statistics and distribution of the 'pct' variable in the original dataset (FiveThirtyEight 2024). Table 2 and Figure 2 shows the summary statistics and distribution of the same variable, but in the filtered dataset that only comprises of the supporting votes for Harris from relatively high-quality polling organizations. Comparing the summary statistics for the raw data (Table 1) and filtered data (Table 2), higher numbers were derived from data filtered only by Harris supporters. Also, Figure 2 illustrates that a significant number of polls indicate support levels ranging from 45% to 50%, which suggests a stable yet not substantial endorsement. This proposes that Harris possesses a reliable foundational support, although her capacity to obtain a majority remains ambiguous.

Table 1: Summary statistics for the proportion (%) of support that candidates received in the poll

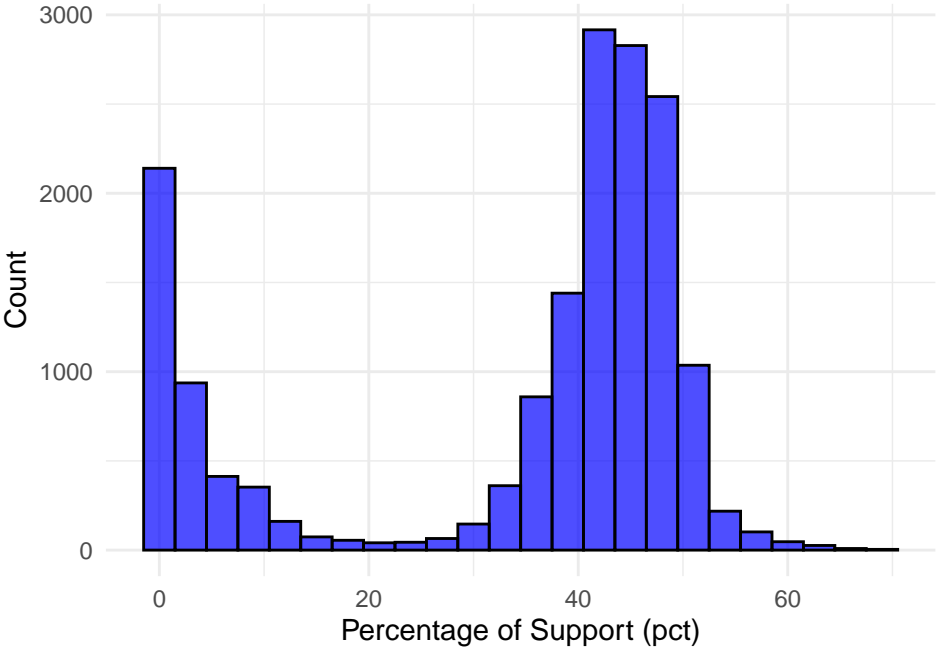| mean	median	min	max	sd	n
33.78	42	0	70	18.19	16817

Figure 1: Distribution of the proportion (%) of support that candidates received in the poll

Table 2: Summary statistics for the proportion (%) of support that Harris received in high-quality polls

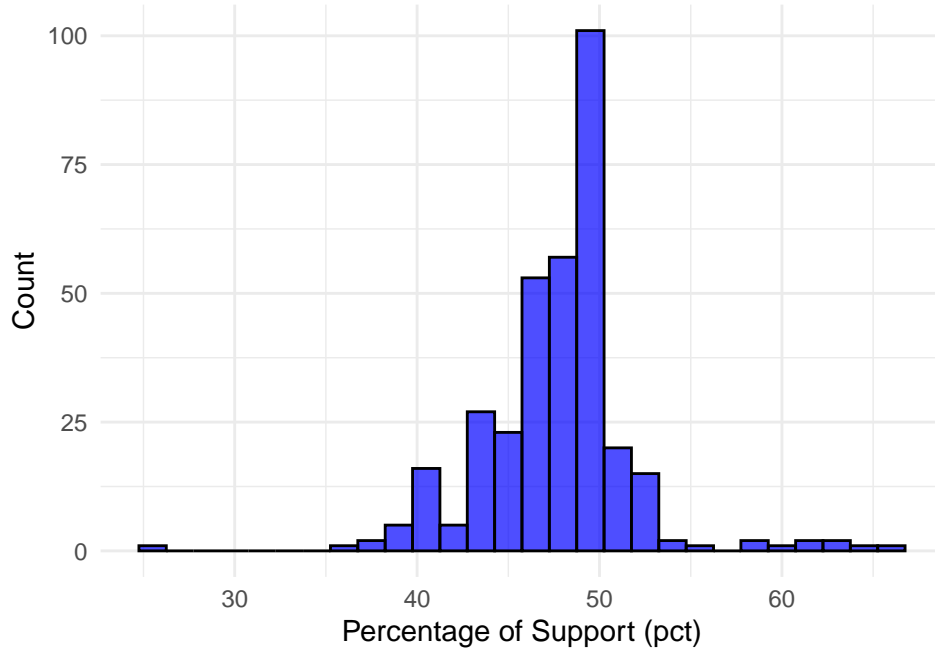| mean	median	min	max	sd	n
47.66	48	25	65.3	4.1	338

Figure 2: Distribution of the proportion (%) of support that Harris received in high-quality polls

2.4 Predictor variables

2.4.1 The Date the Poll was Concluded

The 'end_date' variable representing the time the poll was concluded was put into account to keep track of how support for a candidate changes in time. The reported end dates in the original dataset (FiveThirtyEight 2024) ranges from 7 April, 2021 to 29 October, 2024 (Table 3). Figure 3 shows that the polling data is more concentrated on survey results conducted in the recent period.

Table 3: Summary statistics for the date that the poll was concluded

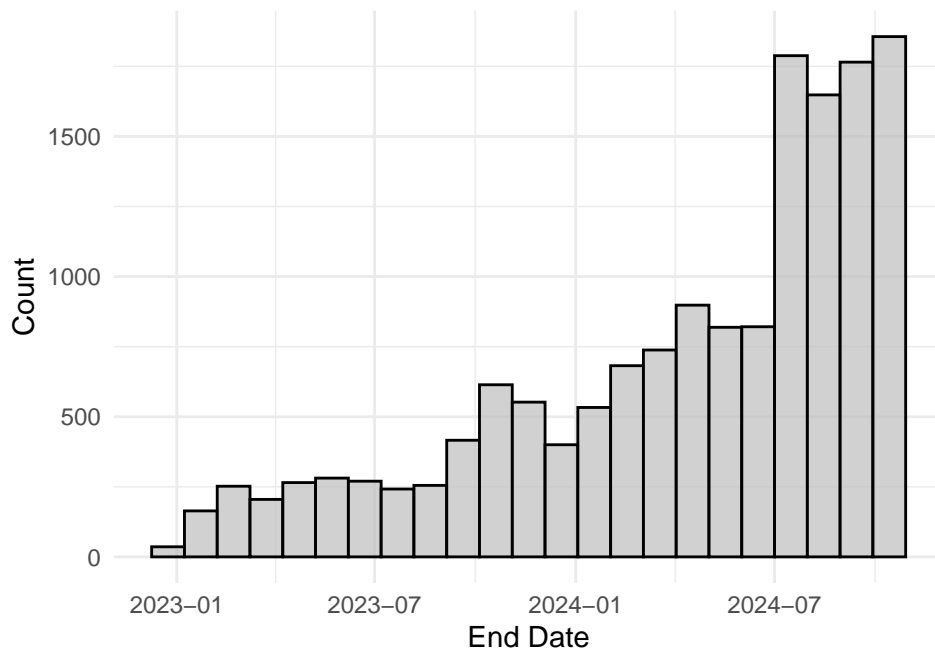| Min	Max
2021-04-07	2024-10-29

Figure 3: Distribution of the date that the poll was concluded

When filtering the data, not only pollster quality and candidate type but also the date variable was considered. The filtered data contains only the polling data after the declaration of Harris. So, the date variable for filtered data in Table 4 ranges from 23 July, 2024 to 29 October, 2024. Figure 4 shows that overall, polling is conducted regularly but intensifies around specific dates.

Table 4: Summary statistics for the date that high-quality polls for Harris was concluded

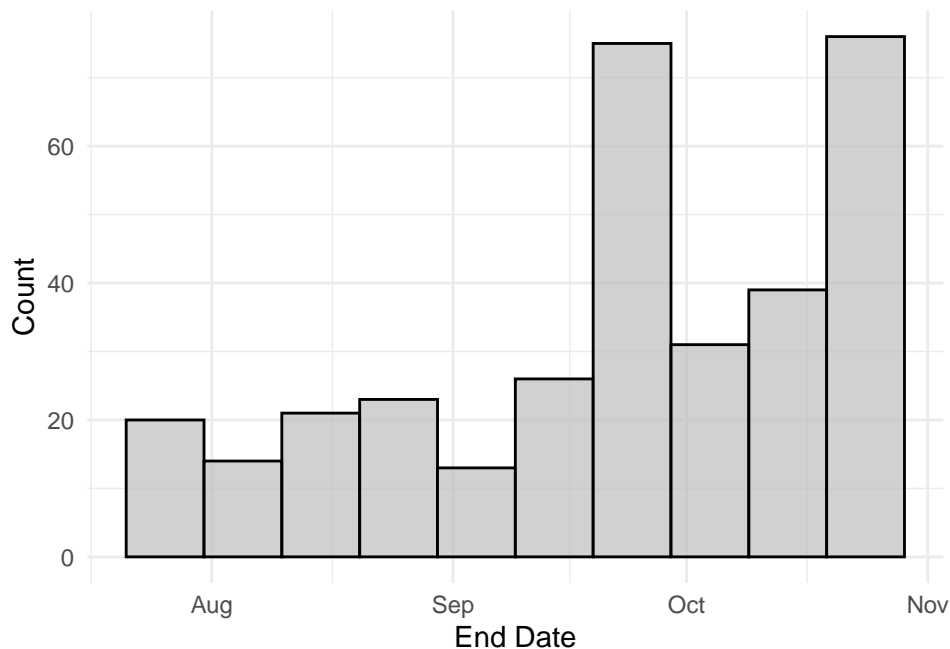| Min	Max
2024-07-23	2024-10-29

Figure 4: Distribution of the date that high-quality polls for Harris was concluded

2.4.2 Pollster and State

The 'pollster' and 'state' variable were selected to consider the effect of changes in poll-making organizations and geographical distinctions. The two variables respectively represent the polling organization that conducted the poll and the US state where the poll was conducted or focused.

Table 5 shows that the original dataset (FiveThirtyEight 2024) contains 232 distinct pollsters and 56 distinct states. After filtering for high-quality polls and assigning 'other' for states with fewer than 60 polls, the analysis data contains 8 distinct poll-making organizations and 20 geographical distinctions as shown in Table 6.

Table 5: Number of distinct polling organizations and US states where the poll was conducted

Pollster	State
232	56

Table 6: Number of distinct high-quality polling organizations and US states where more than 60 polls for Harris were conducted

Pollster	State
8	20

The distribution of polling counts for different pollsters in Figure 5 suggests that the analysis data is dominated by a few pollsters, particularly Siena/NYT. Depending on their polling methodology, the general results may have potential biases. A detailed analysis of the polling methodology and possible errors of the organization will be covered in Section A.
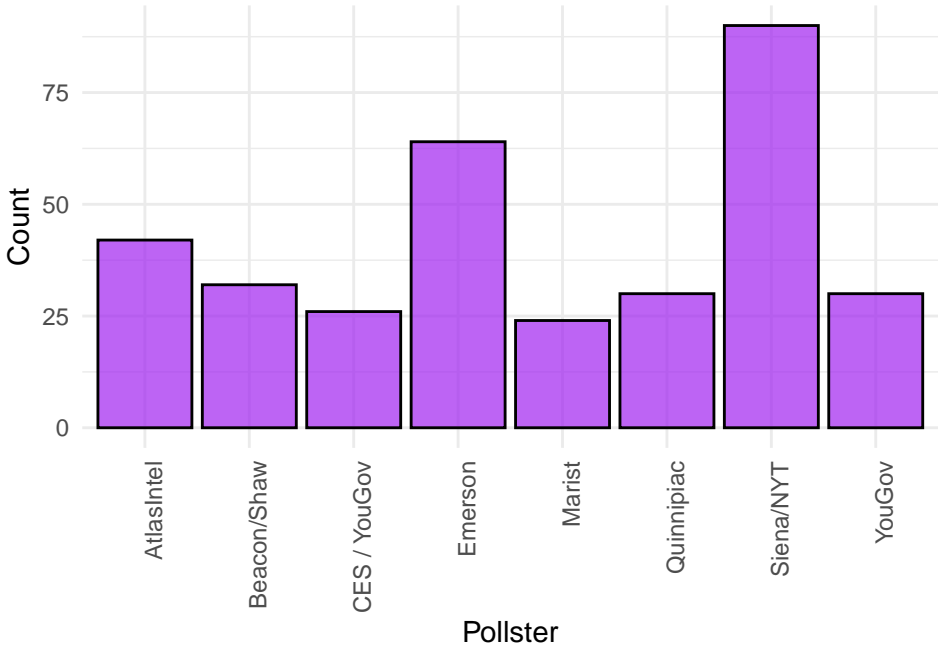

Figure 5: Distribution of polling organizations where high-quality polls for Harris were conducted

Figure 6 displays the distribution of polls across different states in the analysis data. Pennsylvania, Arizona, and Georgia are the top 3 states with high number of polls while states like Montana, New Mexico, and Maryland have much fewer polls. Note that a significant number of national or unspecified state-level polls are aggregated in this analysis data regarding the high count in 'Other' category. The concentration of polls in certain states further suggests a strategic focus on areas likely to impact the election outcome (11Alive 2024).
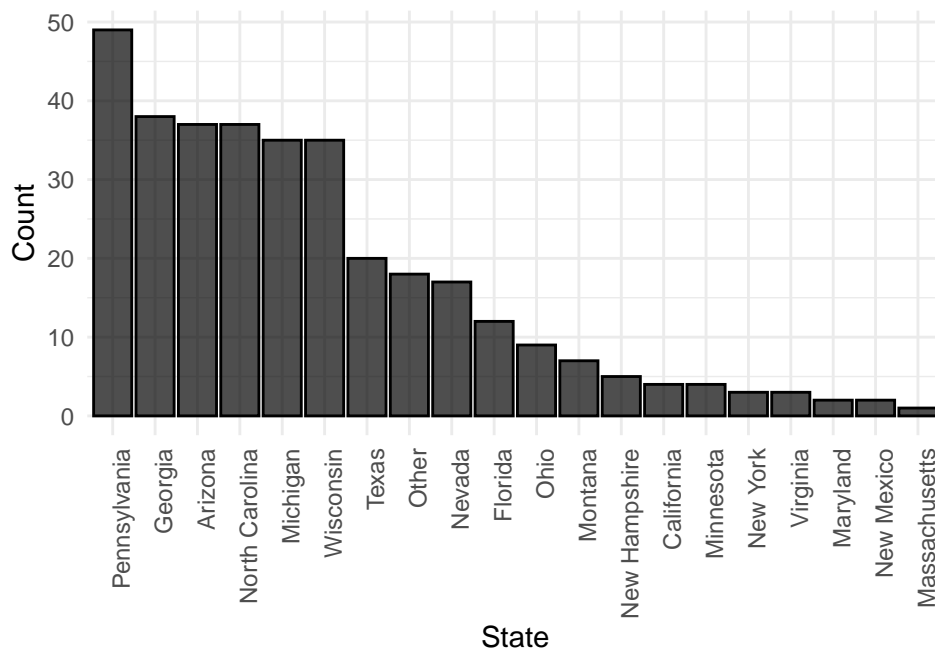
Figure 6: Distribution of US states where the more than 60 polls for Harris were conducted

2.5 Correlation between predictor variables

By examining the relationships between predictor variables in the analysis dataset, we aim to identify potential biases in interpretation and model instability. This is expected to ensure a transparent understanding of the model's robustness and interpretability.

2.5.1 End Date and State

Figure 7 shows notable polling concentrations on certain states such as Pennsylvania, Ohio, and Nevada in week 2024-37 (September 9, 2024) or 2024-38 (September 15, 2024). This suggests that these states are key battelgrounds or areas of strategic focus during the election period. Moreover, polling activities are not consistent across all weeks and there are clear peaks in polling activity in week 2024-38 (September 15, 2024), which may correspond to significant political events, debates, or media focuses.
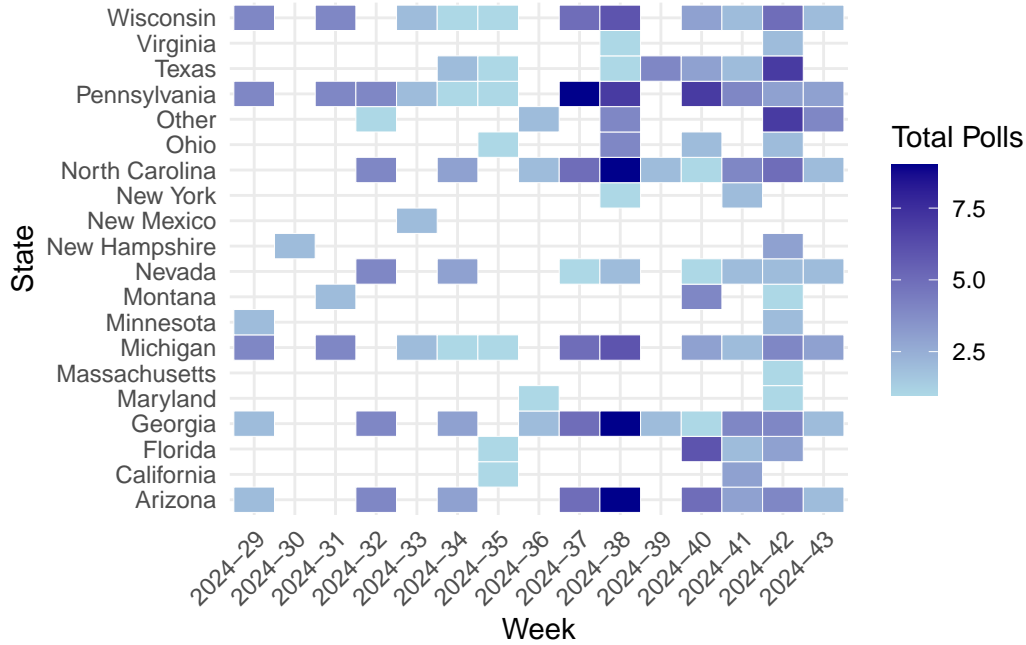
Figure 7: Polling concentration over time by state (*Note:* Strategic focus of polling is directed in battleground states such as Pennsylvania and Ohio in September 15, 2024)

2.5.2 End Date and Pollster

Figure 8 shows that Siena/NYT, which made up the largest proportion of polls in our analysis dataset, has concentrated polling efforts in week 2024-37 (September 9, 2024) and week 2024-38 (September 15, 2024). Notably active periods, possibly around key election events, might introduce temporal bias and overemphasize methodologies of Siena/NYT in that specific time period. Emerson shows overall consistency of polling activities across several weeks indicating steady involvement, while other polling organizations show more sporadic or minimal involvement.
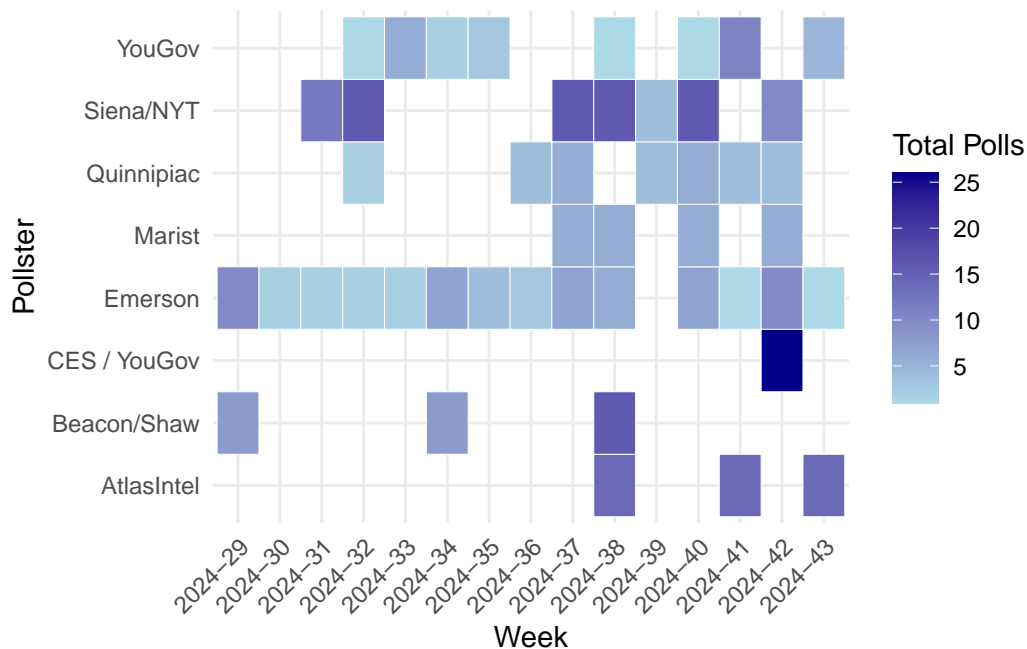
Figure 8: Polling concentration over time by pollster (*Note:* Concentrated polling efforts of Siena/NYT are shown in the first half of September, 2024.)

2.5.3 Pollster and State

Figure 9 shows that Siena/NYT has the strongest state-level presence in key battleground states like Michigan and Arizona. Others like AtlasIntel and Beacon/Shaw has minimal polling coverage in targeted states such as Nevada and Georgia. This variation in coverage could influence the overall analysis in that some states might receive disproportionately more attention from certain pollsters.
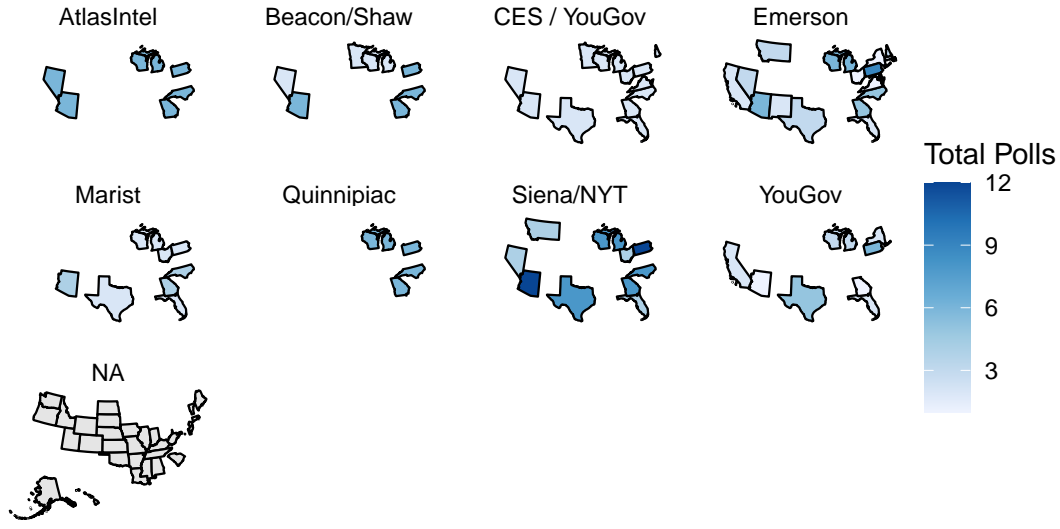
Figure 9: Polling activity by pollster and state (*Note:* Siena/NYT has strongest state-level polling coverage in key battleground states)

3 Model

Our modelling strategy estimates Kamala Harris's support percentage in the 2024 US election polls, accounting for potential variations over time including pollster and state differences. The model balances complexity with interpretability, using both linear and Bayesian frameworks to capture patterns in the data. We turned from the initial linear model to a Bayesian model for robust predictions considering the variables' uncertainties. The Bayesian approach, which incorporates both prior beliefs and observed data to make predictions, allowed us to capture the variances among pollsters and states more accurately and responsively. Further background details and diagnostics are included in Section A.3.

3.1 Model set-up

Define y_i as the percentage of support that Kamala Harris receives in poll i. We begin with two simple linear models and progress to more complex Bayesian models that account for hierarchical structures.

The following models outline our approach:

3.1.1 Linear Model by Date

$$y_i = \beta_0 + \beta_1 \cdot \text{end\_date}_i + \epsilon_i \tag{1}$$
$$\epsilon_i \sim \text{Normal}(0, \sigma^2) \tag{2}$$

where: - y_i is the percentage of support for Harris in poll i, - β_0 is the intercept, - β_1 represents the effect of the poll's end date, - ϵ_i is the error term.

3.1.2 Linear Model by Date and Pollster

$$y_i = \beta_0 + \beta_1 \cdot \text{end\_date}_i + \gamma_{p[i]} + \epsilon_i \tag{3}$$
$$\epsilon_i \sim \text{Normal}(0, \sigma^2) \tag{4}$$

where: - $\gamma_{p[i]}$ is a fixed effect for pollster p conducting poll i (e.g., Siena/NYT).

3.1.3 Bayesian Model with Random Intercept for Pollster

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{5}$$
$$\mu_i = \beta_0 + \beta_1 \cdot \text{end\_date}_i + \gamma_{p[i]} \tag{6}$$
$$\gamma_p \sim \text{Normal}(0, \sigma_\gamma) \tag{7}$$

where: - γ_p is a random effect for pollster p.

3.1.4 Bayesian Model with Random Intercept for Pollster and State

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{8}$$
$$\mu_i = \beta_0 + \beta_1 \cdot \text{end\_date}_i + \gamma_{p[i]} + \delta_{s[i]} \tag{9}$$
$$\gamma_p \sim \text{Normal}(0, \sigma_\gamma) \tag{10}$$
$$\delta_s \sim \text{Normal}(0, \sigma_\delta) \tag{11}$$

where: - $\delta_{s[i]}$ is a random effect for state s.

The Bayesian models are fit using `rstanarm` in R. The priors used are weakly informative: - $\beta_0 \sim \text{Normal}(0, 10)$ - $\sigma \sim \text{Exponential}(1)$

3.1.5 Model justification

Different pollsters and states induce variations in polling results, as pollsters may have distinct methodologies and states represent diverse voter bases. Incorporating random effects for both pollsters and states allows us to improve the robustness of the model.

These models are run through the rstanarm package (Goodrich et al. 2024) in R (R Core Team 2023). RMSE and WAIC were used to check for the goodness of fit of the model; Bayesian models with reduced RMSE and WAIC outperform linear models. We use weakly informative priors; for example, $\beta_0 \sim \text{Normal}(0, 10)$ and $\sigma \sim \text{Exponential}(1)$. This reflects our initial uncertainty but prevents overfitting. The priors were chosen conservatively to ensure that the model remains consistent.

Model diagnostics, including posterior predictive checks and convergence diagnostics, were carried out to ensure the reliability of the results. The Bayesian models converged successfully, as indicated by $\hat{R} = 1$ for all parameters. Further details about model diagnostics are included in Section A.3.

The main assumption in these models is that the pollster and state effects can be treated as random. This assumes that the effects are normally distributed across pollsters and states, which may not always be accurate. Additionally, the model assumes that polling data is representative of the actual electorate, an assumption that can be violated if polls are biased or have non-random sampling issues. Despite these limitations, the hierarchical structure allows us to capture important variability, making the model suitable for predicting Harris's support. Future improvements could involve incorporating time-varying effects or interaction terms between pollsters and states.

4 Results

4.1 Results from examining the analysis data

Figure 10 shows an initial increase in support for Harris since her declaration to join the electoral race, peaking around mid-September. However, this initial momentum stabilized with minor decrease in support as the election gets closer. Moreover, the scattered data points around the summary line show high variance between individual polls.
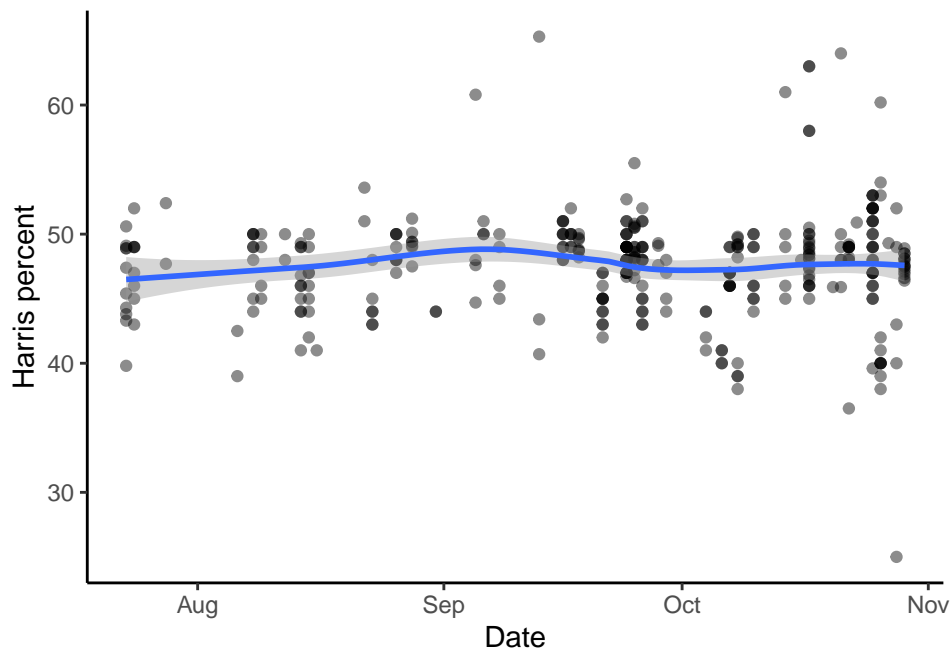
Figure 10: Polling votes for Harris over time (*Note:* The date starts from after Harris officially announced her campaign for 2024 U.S. presidential election on July 21, 2024.)

When accounting for the pollster difference, Figure 11 and Figure 12 show variations in Harris' support levels that are obscured in the aggregated view of Figure 10. Siena/NYT (pink) shows relatively stable polling results for Harris, with fewer extreme values and a slight downward trend in October. On the other hand, Emerson (green) shows more fluctuation throughout the period with a wider spread of polling values.
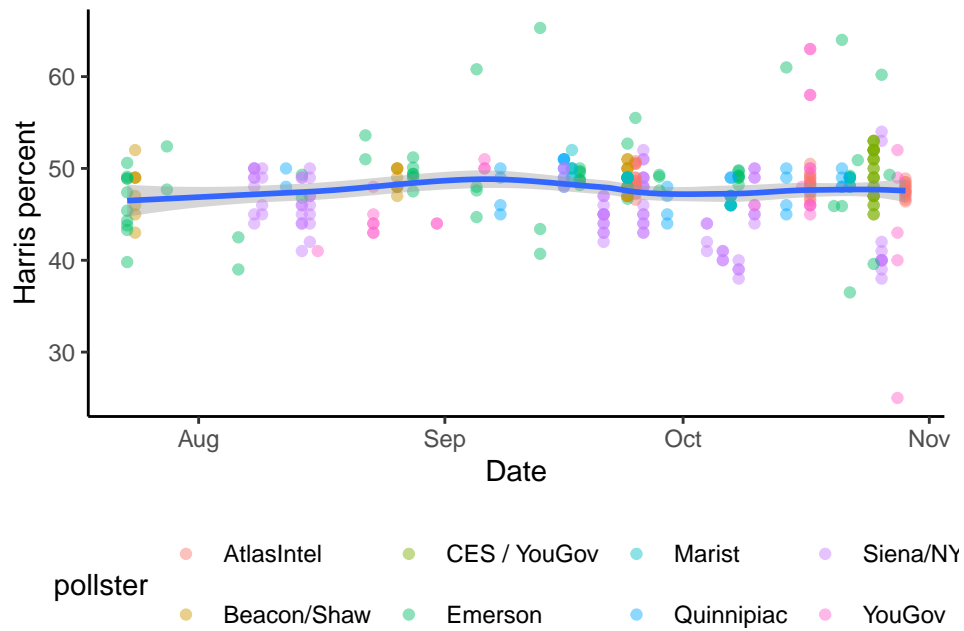
Figure 11: Polling votes for Harris over time by pollster (*Note:* The date starts from after Harris officially announced her campaign for 2024 U.S. presidential election on July 21, 2024. Only filtered pollsters with high-quality polling questions are shown in the following.)
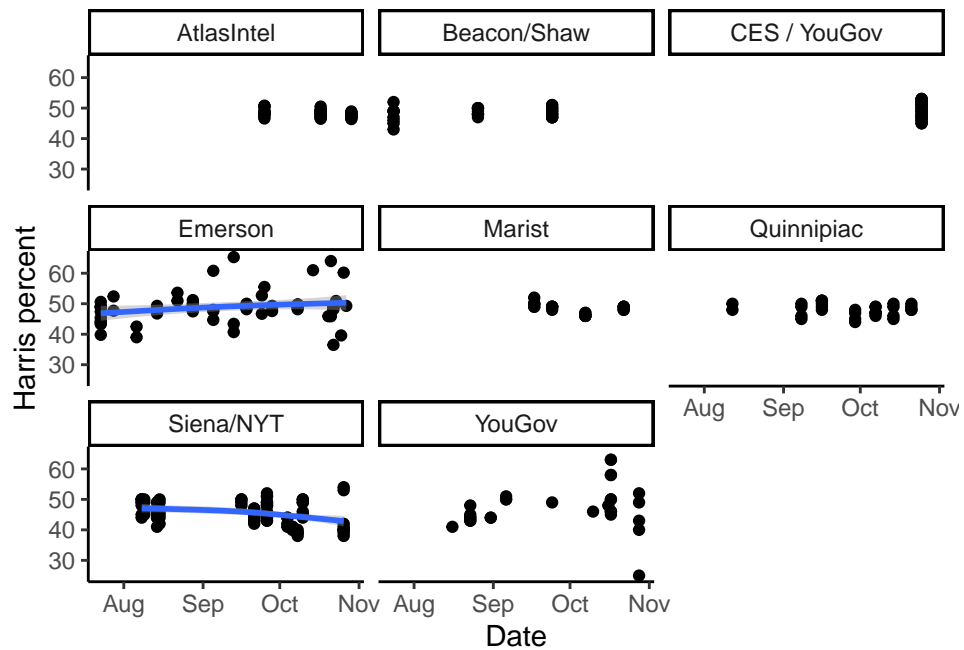
Figure 12: Polling votes for Harris over time by pollster (facets) (*Note:* The date starts from after Harris officially announced her campaign for 2024 U.S. presidential election on July 21, 2024. Only filtered pollsters with high-quality polling questions are shown in the following.)

When accounting for the state difference, Figure 13 and Figure 14 show variations in Harris' support levels. Critical battleground states like Pennsylvania and Michigan show distinct trends from the national analysis in Figure 10. Pennsylvania shows relatively stable trends slightly above the national average while Michigan shows fluctuations over time.
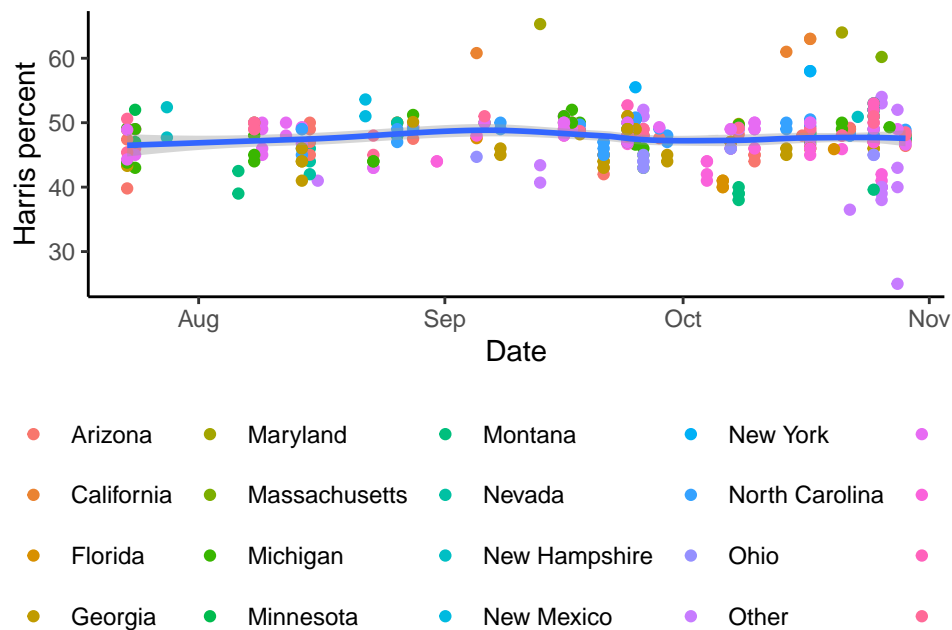

Figure 13: Polling votes for Harris over time by state (*Note:* The date starts from after Harris officially announced her campaign for 2024 U.S. presidential election on July 21, 2024. Only filtered pollsters with high-quality polling questions are shown in the following.)

Figure 14: Polling votes for Harris over time by state (facets) (*Note:* The date starts from after Harris officially announced her campaign for 2024 U.S. presidential election on July 21, 2024. Only filtered pollsters with high-quality polling questions are shown in the following.)
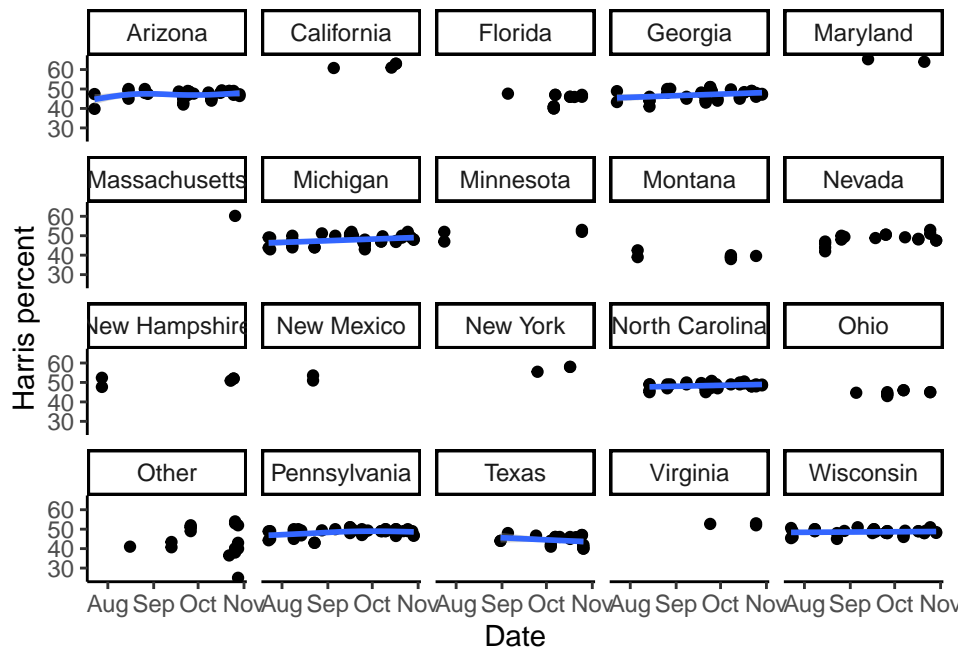
4.2 Results from the prediction model

Prediction results derived from our model frameworks are summarized in Table 7 and Table 8. In the linear models (Table 7), higher R^2 and lower WAIC, RMSE values in the second column show that the inclusion of pollster effects improves accuracy and reliability of the prediction model. In the Bayesian models (Table 8), higher Log-Likelihood and lower ELPD, WAIC values in the fourth column shows that including both pollster and state effects enhances model performance.

Figure 15 and Figure 16 show the trends in the percentage of supportive polls for Harris predicted by linear models, respectively representing a time-series analysis and consideration of pollster differences. Moreover, Figure 17 and Figure 18 show the supportive trends predicted by Bayesian models accounting for pollster and state differences. Each model, from the simple linear regression to the more complex Bayesian models, seem to reflect a relatively consistent support pattern over time. This alignment of underlying support trends across model shows the credibility of the observed overall support trend.
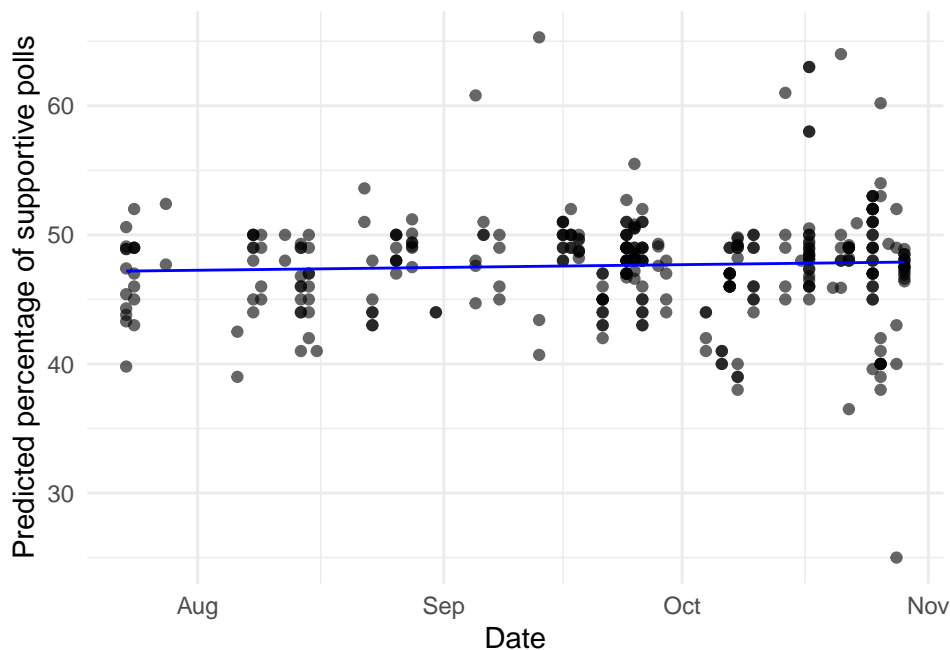

Figure 15: Predicted percentage of supportive polls for Harris in the linear model by date

Table 7: Linear models of support percentages for Harris based on date and pollster

	Linear by Date	Linear by Date, Pollster
(Intercept)	−94.75	48.22
	(159.49)	(175.27)
end_date	0.01	0.00
	(0.01)	(0.01)
pollsterBeacon/Shaw		0.11
		(0.96)
pollsterCES / YouGov		1.14
		(0.99)
pollsterEmerson		0.59
		(0.83)
pollsterMarist		0.20
		(0.99)
pollsterQuinnipiac		−0.23
		(0.94)
pollsterSiena/NYT		−2.82
		(0.77)
pollsterYouGov		−0.79
		(0.93)
Num.Obs.	338	338
R2	0.003	0.131
R2 Adj.	−0.006	0.072
Log.Lik.	−955.998	−936.341
ELPD	−959.7	−946.3
ELPD s.e.	25.8	27.7
LOOIC	1919.4	1892.6
LOOIC s.e.	51.6	55.3
WAIC	1919.5	1892.6
RMSE	4.08	3.84

Table 8: Bayesian models of support percentages for Harris based on pollster and state

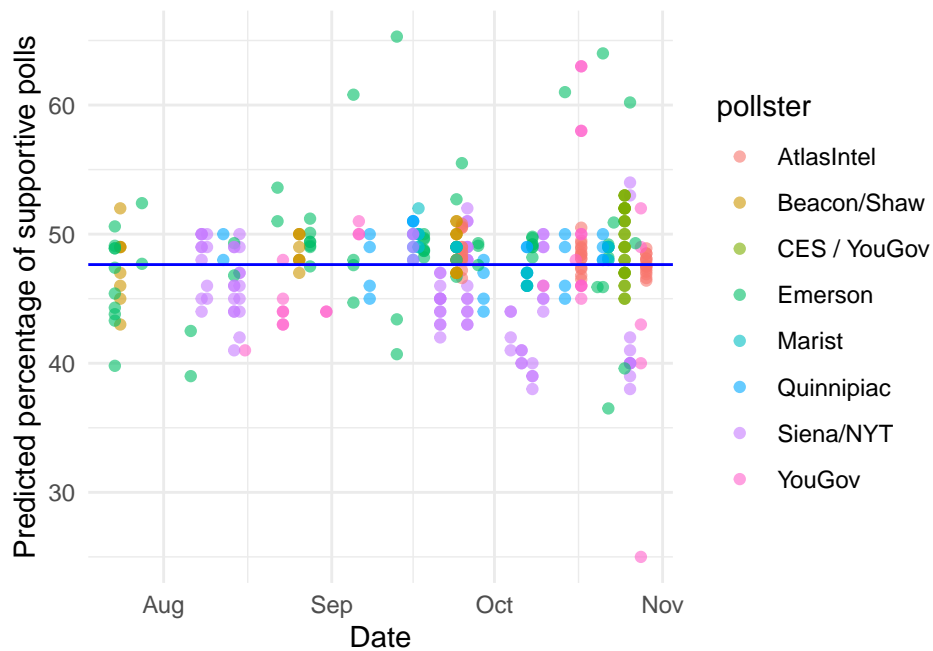|	Bayesian with Pollster	Bayesian with Pollster, State
(Intercept)	−0.08	0.01
	(0.02)	(0.06)
Sigma[pollster × (Intercept),(Intercept)]	0.00	0.00
	(0.00)	(0.00)
Sigma[state × (Intercept),(Intercept)]		0.07
		(0.02)
Num.Obs.	338	338
ICC	0.5	1.0
Log.Lik.	−2154.335	−1563.746
ELPD	−2179.4	−1599.1
ELPD s.e.	127.3	44.8
LOOIC	4358.9	3198.2
LOOIC s.e.	254.7	89.7
WAIC	4358.6	3195.8
RMSE	0.04	0.03

Figure 16: Predicted percentage of supportive polls for Harris in the linear model by date and
pollster (*Note:* The blue summary line calculates the mean of the predictions across
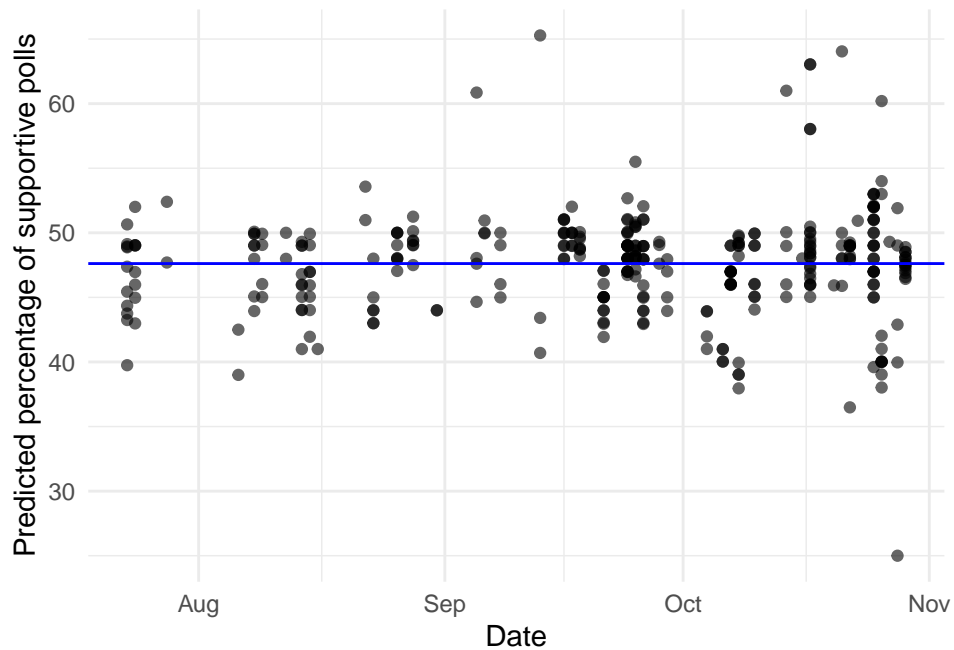all dates and pollsters.)

Figure 17: Predicted percentage of supportive polls for Harris in the Bayesian model with random intercept for `pollster` variable (*Note:* In order to avoid the overfitting problem, overall mean of prediction was separately calculated in creating the summary line.)

Figure 18: Predicted percentage of supportive polls for Harris in the Bayesian model with random intercept for `pollster` and `state` variable (*Note:* In order to avoid the overfitting problem, overall mean of prediction was separately calculated in creating the summary line.)
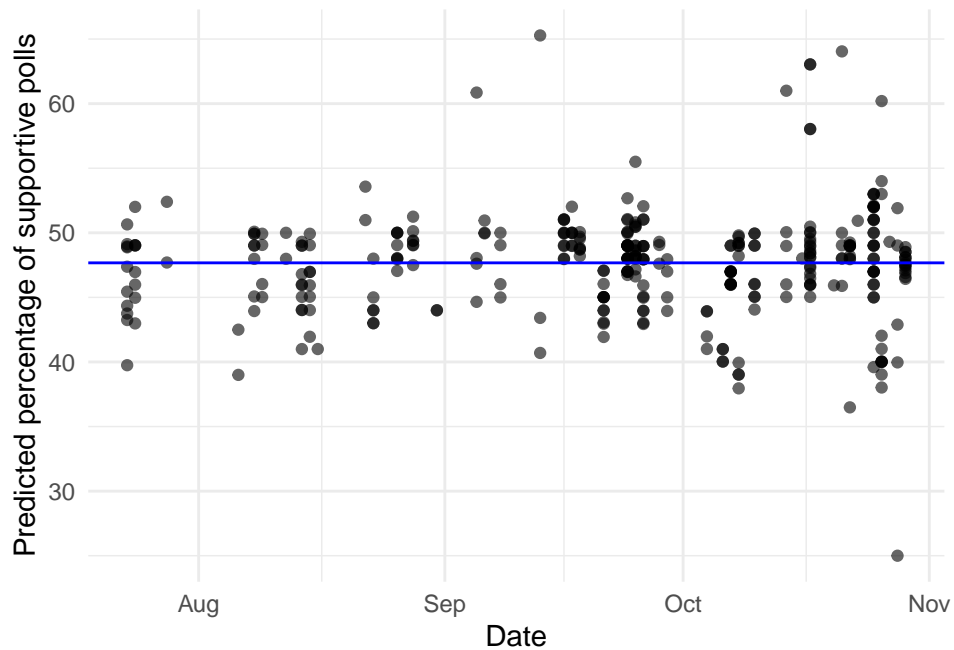
5 Discussion

5.1 Why Harris could beat her polls

Table 1 and Table 2 shows that filtering data based on more reliable polling criteria makes the average rate of support for Harris increase. Given the current polling landscape where Harris and Trump have nearly equivalent levels of support (FactCheck.org 2024), higher support rates for Harris in the filtered analysis serves as substantial evidence that Harris could win the election.

Furthermore, our data analysis that incorporates state-specific factors (Figure 14) reveals that Harris holds relatively stable support in key battleground states, including Pennsylvania, Michigan, and Wisconsin. As it will be discussed in more detail in Section 5.3, this suggests that Harris' support base is both resilient and potentially insulated from the unpredictable swings (Kaiser Family Foundation 2024). Regarding the winner-take-all nature of the electoral college, Harris' stability in these key states not only strengthens her position but also shows high probability in her victory, as even slight leads in battleground states can result in a decisive electoral advantage (Kaiser Family Foundation 2024). The combination of broad polling support and specific regional strength positions Harris as a leading presidential candidate in the 2024 election.

5.2 Pollsters herding around false consensus

Figure 11 and Figure 12 shows notable differences in the supportive trends for Harris when considering the variations of polling organizations. In Table 7, the inclusion of pollster-specific variables had statistical significance. This proposes that pollster-specific factors, like methodology and sample composition, produce a range of outcomes even for the same candidate's support levels (The New York Times Editorial Board 2024). As pollsters often fear publishing outlier results that might undermine credibility, they may adjust the results with a perceived consensus (Silver 2024). This is likely to reduce the accuracy of polling aggregates, misrepresenting the true differences in public sentiment.

To be more specific, this adjustment based on previous consensus data would limit the visibility of demographic shifts that might favor one candidate unexpectedly (Public Opinion Research 2022). For instance, the polling discrepancies in the 2016 and 2020 U.S election were partly attributed to underestimated certain voter segments supportive of Trump (Silver 2024). Assessing the potential limitations of aggregated data, recognizing that it may reflect methodological biases more than independent public opinion should be on the way for a reliable prediction of electoral support.

5.3 The electoral college and the power of battleground states

In the U.S. electoral college, the Constitution assigns each state a set number of electoral based on its Congressional representation and a candidate must secure more than 270 votes to win the presidency (WHYY News 2024). This often results in candidates focusing their campaigns primarily on battleground states, where voter preferences are more fluid and results are less predictable (Ash Center for Democratic Governance and Innovation 2024). Figure 14 show that key battleground states' polling trends differ from that of the national average (Figure 10).

The stability of polling trends in Pennsylvania suggests a consistent voter base despite of the battleground status, which often means that even slight shifts in support could be pivotal (Ash Center for Democratic Governance and Innovation 2024). Harris' focus on healthcare reform, especially her commitment to strengthening the Affordable Care Act (ACA), is expected to have particular resonance in Pennsylvania, where a large population of low-to-middle income residents benefit from ACA subsidies (FactCheck.org 2024). On the other hand, Michigan, where there are economically diverse population, national economic policy discussions could swing voter sentiment more sharply and result in relatively large fluctuations (Kaiser Family Foundation 2024). Surveys carried out within battleground states show wider variation in candidate support, making the incorporation of state-specific factors into electoral models allow a better representation of the diverse political landscapes seen across the United States.

5.4 Weaknesses and next steps

While our model provides a foundation to understand Harris' polling support, further refinements are necessary to enhance its accuracy and applicability. For instance, future studies could focus on the incorporation of voter demographics, such as age, gender, and education. Showing which segments of the population are driving changes in support, this can give further explanation to how certain policies or campaign events of candidates affect overall support for a candidate (Brown et al. 2024). In addition, interaction terms between pollster and state can be included in the model, considering that some states might receive disproportionately more attention from certain pollsters as shown in Figure 9. This can give further explanation to how pollsters and regional dynamics affect overall support for a candidate (Pew Research Center 2020).

This study has its constraints in that sampling error and systematic biases inevitably occur in the process of measuring social phenomenon with data. In particular, due to the nature of polling surveys, non-responses or misunderstanding of the survey questions result in missing data or outliers (Pew Research Center 2020). Considering the possibility of prediction errors, we should take caution in interpreting the analysis and prediction results. By conducting more in-depth research of the domain knowledge and constructing appropriate variables to put into account, it would help derive accurate conclusions rather than just acknowledging the numbers given itself (GeeksforGeeks 2023).

A Appendix

A.1 Pollster methodology overview and evaluation

The New York Times/Siena College polling partnership, the polling organization that accounts for the majority of polls in our analysis (Figure 5), employs a methodology tailored to capture voter preferences for specific elections, whether at the state or national level (The New York Times and Siena College Research Institute 2020). Sample sizes typically range from 600 to 1000 likely voters, with an emphasis on oversampling in battleground states to better reflect regional political dynamics (The New York Times and Siena College Research Institute 2020).

To enhance representativeness, the organization uses a mixed-mode approach that combines random-digit dialing (RDD) for landlines and mobile phones with online surveys, increasing accessibility and demographic coverage (The New York Times and Siena College Research Institute 2020). The polling process relies on stratified random sampling, a technique where the population is divided into mutually exclusive and collectively exhaustive subgroups, or strata (Alexander 2023), based on key demographic variables such as race, education, and geographic location. Within each stratum, random sampling ensures proportional representation. This approach is especially effective in capturing variation within diverse populations (Lohr 2019), as it controls for demographic factors that could otherwise introduce bias. However, stratified sampling can also introduce complexities, such as difficulty in defining strata boundaries, which can affect overall precision if strata are not well-chosen (Lohr 2019).

In addition, the Siena/NYT intends to enhance transparency and reliability through cognitive testing of survey questions. Cognitive testing allows the pollster to evaluate whether respondents interpret questions as intended, which is a necessary process in minimizing response bias and improving question clarity (Presser and Blair 1994). This iterative approach ensures that question wording resonates with respondents' current understanding of political issues, thus reduces measurement error(Siena College Research Institute 2024).

A distinctive strength of this organization's approach is its commitment to accurately reflecting the demographics and political leanings of specific regions. This kind of approach has been validated in high-profile cases, such as the accurate prediction of battleground state outcomes in the 2016 Florida election (Siena College News 2024). By using robust sampling methods and iterative question refinement, their polls are well-positioned to reflect regional voter behavior with a higher degree of accuracy.

However, challenges remain in achieving a fully representative sample. Polling captures a "snapshot in time"(Alexander 2023) and is vulnerable to shifts caused by recent political events, which can introduce volatility into results. Additionally, non-response bias remains an issue, as certain groups that are distrustful of polling organizations or those difficult to reach by phone are often underrepresented (Pew Research Center 2023). Simulation studies in the

literature show that non-response bias can skew results, especially in polarized contexts, unless corrective weighting or follow-up efforts are applied (Presser and Blair 1994).

By continuously refining their approach by using corrective weighting (The New York Times and Siena College Research Institute 2020), Siena/NYT intends to remain a credible source of insights into the evolving dynamics of U.S. elections. To be more specific about this weighting method, it assigns different weights to responses based on demographic factors such as age, race, gender, or geographic location (The New York Times and Siena College Research Institute 2020). For instance, if a poll has fewer responses from young adults compared to their actual proportion in the population, each response from a young adult might be given a higher weight to compensate. This adjustment helps balance out the representation of various groups, making the poll more reflective of the overall population (Siena College News 2024).

A.2 Idealized methodolgy

The proposed methodology for forecasting the 2024 U.S. presidential election with a budget of $100,000 would be designed as follows. First, a stratified random sampling method will be employed that allows for the capture of the demographic elements such as age, gender, race, and education level (Pew Research Center 2023). Stratified sampling involves dividing the population into mutually exclusive subgroups, or strata, based on demographic factors, then randomly sampling from each stratum (Alexander 2023). This approach allows us to control for variation within subgroups, which improves representation and reduces sampling error compared to simple random sampling. However, as noted by Lohr (2019), defining appropriate strata boundaries can be challenging and may impact the precision of estimates if not carefully managed (Lohr 2019). By focusing on 10 battleground states including Florida, Pennsylvania, Michigan, and Arizona, we can capture the diverse political landscapes critical to the election outcome. Each state will receive a $7,000 budget, with $5,000 allocated to conducting telephone and online surveys and $2,000 for training local staff for outreach and data entry. This budget allocation aims to address potential sources of bias, ensuring the sample is reflective of the population and enhancing data reliability.

To further strengthen our methodology, we will invest $15,000 in questionnaire design and testing. High-quality questionnaire design is essential for minimizing response bias and ensuring that questions are clear and unbiased (Alexander 2023). With $5,000 on initial drafting and expert review, $7,000 on conducting pre-tests and pilot studies (Drive Research 2023) and $3,000 on final review and translation of the survey questions, we expect to improve the validity and reliability of the collected data. In particular, pre-tests, rooted in cognitive testing, is expected to help assess whether respondents interpret questions as intended, a critical step for reducing measurement error and enhancing question validity (Presser and Blair 1994).

Then, $10,000 will be dedicated to data weighting and cleaning. Post-stratification weighting will adjust for demographic imbalances in the sample, helping to correct for potential biases from non-response and underrepresented groups (Valliant 2019). By assigning $5,000 for these

weighting procedures, we aim to ensure that the sample aligns as accurately as possible with the U.S. voting population. Additionally, $5,000 will be allocated to data cleaning, which includes consistency checks and removal of duplicates or erroneous entries, further improving data quality and reducing bias (Pew Research Center 2023).

Our final prediction model will employ Bayesian hierarchical modeling, which allows us to account for uncertainty and variability across states, pollsters, and external factors. Bayesian models are particularly effective in election forecasting, as they enable dynamic updates based on real-world changes, such as shifts in public opinion following political events. By allocating $3,000 for model setup and calibration and $2,000 for cross-validation and out-of-sample testing, we ensure the model's robustness in capturing complex electoral dynamics (Pew Research Center 2023). This approach not only provides flexibility in handling diverse data sources but also improves predictive accuracy through structured handling of variability and uncertainty.

The final model would use Bayesian hierarchical modeling, which allows for more flexible modeling of uncertainty and variation across states, pollsters, and other external factors. These models, along with out-of-sample testing and cross-validation, enable accurate prediction sensitive to the dynamics of real-world changes, including political events (StataCorp 2024). By investing $3,000 on model setup and calibration and $2,000 on cross-validation and testing, we expect to ensure the model remains robust and provides reliable predictions under various electoral scenarios.

In summary, this methodology balances rigorous sampling, robust question design, and advanced statistical modeling, all underpinned by established survey and sampling principles. These practices aim to deliver a reliable and comprehensive forecast for the 2024 U.S. presidential election.
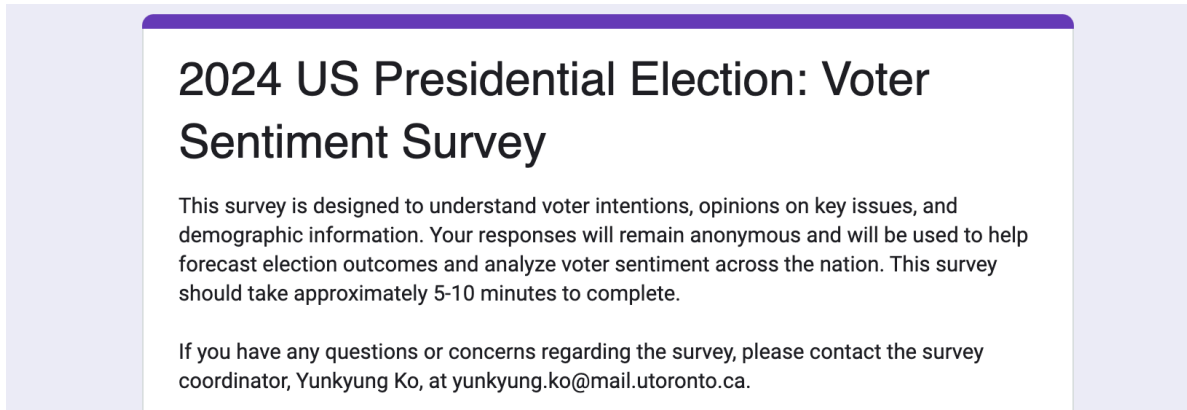
A.2.1 Idealized survey

The proposed survey questionnarie design is in the following link:
https://forms.gle/vehDvpDNDGZ8WCn47

Following the guidance of Alexander (2023), we aimed to have survey questions that are conversational and flow from one to the next, grouped within topics (Elson 2018). We refined the wording of questions to be based on what the respondent would be comfortable with, simultaneously checking for whether the survey questions could appropriately relate to measuring the estimand (Alexander 2023).
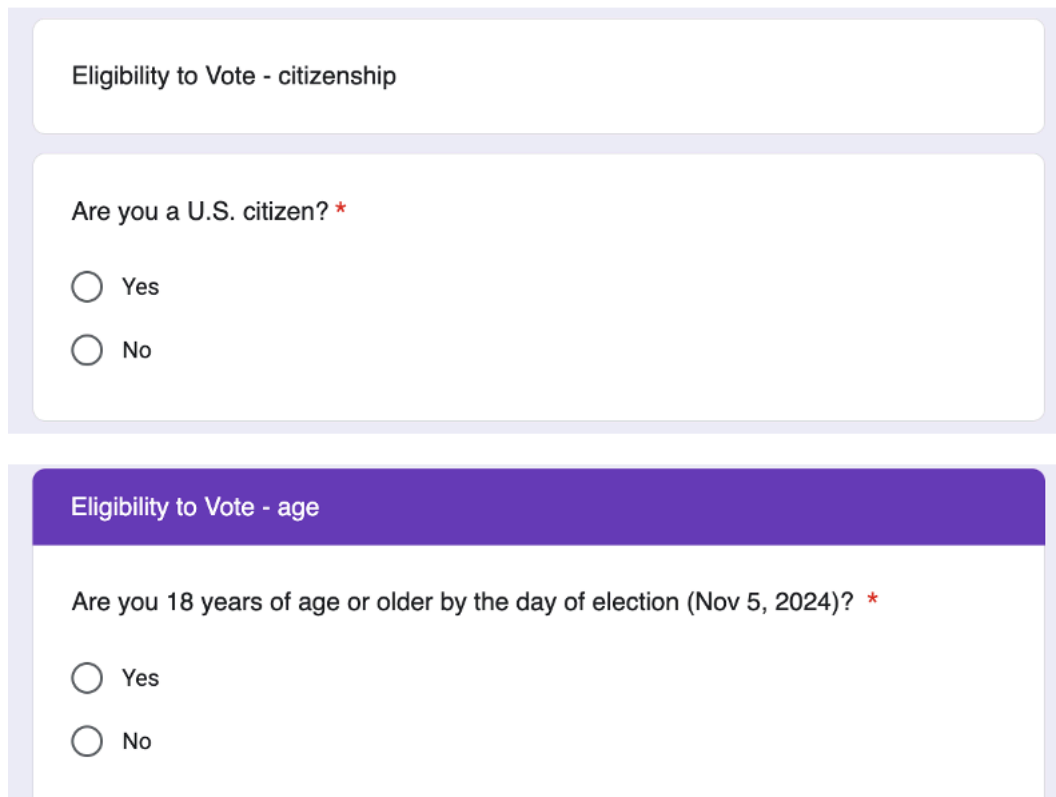
A.2.2 Survey demo

Figure 19: Survey Intro

Figure 20: Questions asking for vote eligibility - citizenship, age

Eligibility to Vote - registered or not

Are you registered to vote in the United States? *

○ Yes

○ No

○ I'm not sure

If registered, in which U.S. state or territory are you registered to vote?

○ California

○ Texas

○ Florida

○ New York

○ Pennsylvania

○ Illinois

○ Ohio

○ Georgia

○ North Carolina

○ Michigan

○ New Jersey

○ Virginia

○ Washington

○ Arizona

○ Massachusetts

○ 기타:

Figure 21: Questions asking for vote eligibility - voter status and registration state

Demographics

What is your age? *

○ 18-24

○ 25-34

○ 35-44

○ 45-54

○ 55-64

○ 65-74

○ 75-84

○ 85+

○ 기타: _____

What is your gender *

○ Male

○ Female

○ Prefer not to say

○ 기타: _____

Which of the following best describes your race/ethnicity? *

○ White

○ African American

○ Hispanic or Latino

○ Asian

○ Prefer not to say

○ 기타: _____

What is your highest level of education? *

○ High school or less

○ College

○ Bachelor's degree

○ Graduate degree

○ Prefer not to say

○ 기타: _____

Figure 22: Demographics Questions

Voting Intentions

If the 2024 U.S. presidential election were held today, for whom would you vote? *

○ Kamala Harris

○ Donald Trump

○ Other

○ Undecided

Candidate Favorability
Please rate your favorability to the following candidates on the scale below.

How do you view Kamala Harris? *

	1	2	3	4	5	
Very Unfavorable	○	○	○	○	○	Very Favorable

How do you view Donald Trump? *

	1	2	3	4	5	
Very Unfavorable	○	○	○	○	○	Very Favorable

Figure 23: Voting Intentions and Candidate Favorability

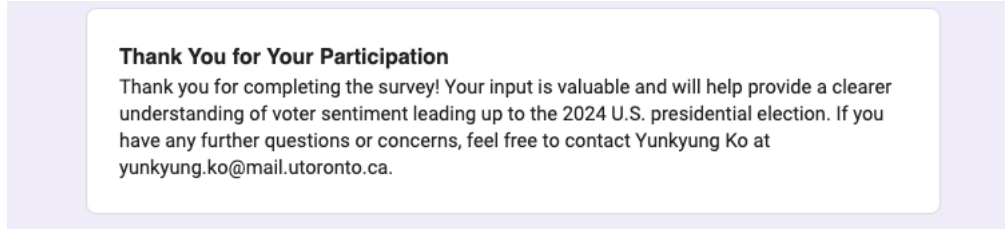Figure 24: Issues of Interest and Likelihood to Vote

Figure 25: Survey Final Thanks

A.3 Model details

A.3.1 Posterior predictive check

In the first posterior predictive check (Figure 26a), we compare the observed data with replicated data generated from the posterior distribution. This shows that the model is able to replicate the overall distribution of the observed data, with the replicated curves (light blue) closely following the true data (dark blue line). This indicates that the model fits the data well in terms of capturing the main pattern or trend (Stan Development Team 2023).

In the second plot (Figure 26b), the replicated data which had both the pollster and state variable as random intercepts, shows relatively closer approximation to the true data distribution. The narrowing of uncertainty in the posterior relative to the prior indicates the impact of the data on refining the model's predictions. This reassures that the model fits the data reasonably well and that the prior information has been appropriately updated by the observed data (Stan Development Team 2023).
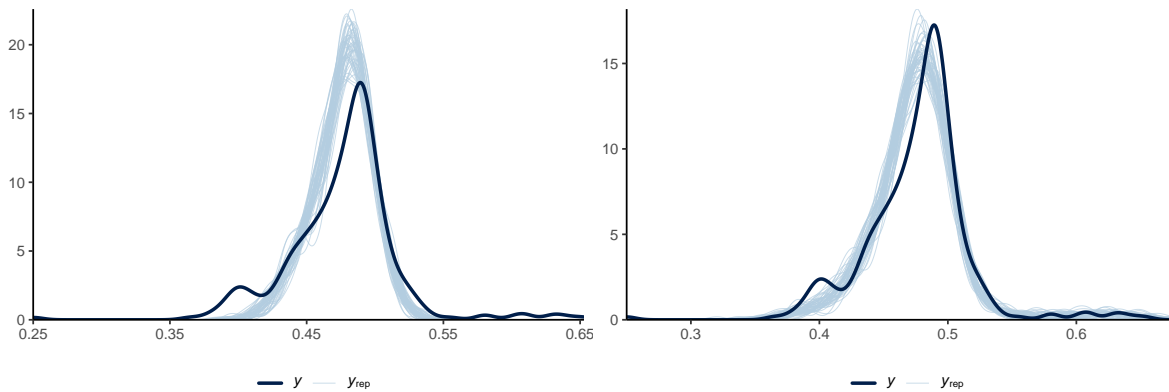
(a) Posterior prediction check   (b) Posterior prediction check

Figure 26: Examining how the Bayesian model fits, and is affected by, the data

A.3.2 Diagnostics

Figure 27a is a trace plot. The sampled values for posterior distribution of intercept parameter across iterations of the MCMC algorithm shows good convergence (Gabry, Češnovar, and contributors 2021b). The lines for the parameter appear to be stable and fluctuating around a central value without any clear trends or patterns. This suggests that the MCMC algorithm has likely converged, and the posterior samples are representative of the target distribution.

Figure 27b is a Rhat plot. The Rhat value is approximately 1.0 for the intercept, which shows that the variance within and between multiple chains have converged. An Rhat value close to 1 indicates that the chains have mixed well and are drawing from the same distribution while values significantly greater than 1 would indicate that further iterations are needed (Gabry, Češnovar, and contributors 2021a).This suggests that the Bayesian models for both "pollster" and "state" have likely converged, and the results derived from these models are reliable.
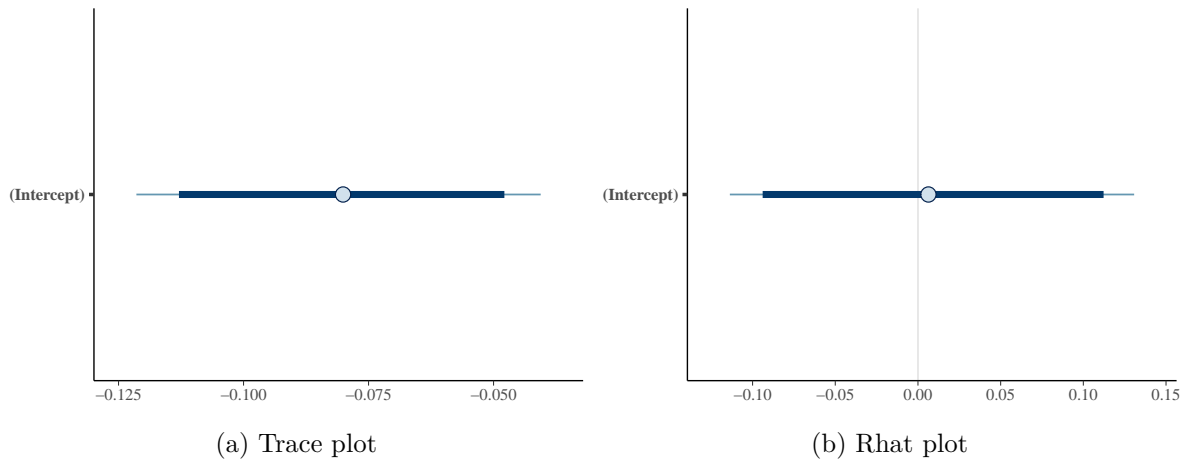

(a) Trace plot (b) Rhat plot

Figure 27: Checking the convergence of the MCMC algorithm

References

11Alive. 2024. "What Does a Battleground State Mean? What Are the Swing States?" 2024. https://www.11alive.com/article/news/politics/elections/what-does-a-battleground-state-mean-what-are-the-swing-states/85-89793c86-c6b3-4394-a1f9-e4cf6136f35e.

Alexander, Rohan. 2023. *Telling Stories with Data.* Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Ash Center for Democratic Governance and Innovation. 2024. "The Electoral College and Our Broken Presidential Election System." 2024. https://ash.harvard.edu/articles/the-electoral-college-and-our-broken-presidential-election-system/.

Bijune, Aiste, and Lan Ha. 2024. "US 2024 Election: Implications for the Global Economy." *Euromonitor International.* https://www.euromonitor.com/article/us-2024-election-implications-for-the-global-economy.

Brown, Nadia, Shana Kushner Gadarian, Antoine J Banks, and Kathleen Searles. 2024. "Candidate Characteristics, Identity, and Perceptions of Electability." *Political Behavior.* https://doi.org/10.1007/s11109-023-09909-3.

CSIS. 2024. "The Global Impact of the 2024 u.s. Presidential Election." https://features.csis.org/2024-us-election-global-impact/.

Drive Research. 2023. "Cognitive Testing Surveys [Examples + Benefits]." https://www.driveresearch.com/market-research-company-blog/cognitive-testing/.

Elson, Malte. 2018. "Question Wording and Item Formulation." PsyArXiv Preprints. https://doi.org/10.31234/osf.io/e4ktc.

FactCheck.org. 2024. "Trump Vs. Harris on u.s. Manufacturing." 2024. https://www.factcheck.org/2024/09/trump-vs-harris-on-u-s-manufacturing/.

FiveThirtyEight. 2024. "Dataset: US Presidential General Election Polls." https://projects.fivethirtyeight.com/polls/data/president_polls.csv.

Gabry, Jonah, Šimon Češnovar, and contributors. 2021a. *MCMC Diagnostics in Bayesplot.* https://mc-stan.org/bayesplot/reference/MCMC-diagnostics.html.

———. 2021b. *MCMC Trace Plots in Bayesplot.* https://mc-stan.org/bayesplot/reference/MCMC-traces.html.

GeeksforGeeks. 2023. "Role of Domain Knowledge in Data Science." https://www.geeksforgeeks.org/role-of-domain-knowledge-in-data-science/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Jackman, Simon. 2024. "Pooling the Polls over an Election Campaign." https://cdn-uploads.piazza.com/paste/ket0p3a9re9628/8a55272a479e1aaeabbd01b3030aa1d38c4a7fc323a9c3feee3bd5fec2fdca44/Pooling_the_polls_over_an_election_campaign.pdf.

Kaiser Family Foundation. 2024. "KFF: Health Policy Analysis, Polling, and News." 2024. https://www.kff.org/.

Lohr, Sharon L. 2019. *Sampling: Design and Analysis.* 2nd ed. Boca Raton, FL: Chapman;

Hall/CRC. https://www.routledge.com/Sampling-Design-and-Analysis/Lohr/p/book/9781138059559.

Pew Research Center. 2020. "Understanding How 2020's Election Polls Performed and What It Might Mean for Other Kinds of Survey Work." https://www.pewresearch.org/short-reads/2020/11/13/understanding-how-2020s-election-polls-performed-and-what-it-might-mean-for-other-kinds-of-survey-work/.

———. 2023. "How Public Polling Has Changed in the 21st Century." *Pew Research Center*. https://www.pewresearch.org/methods/2023/04/19/how-public-polling-has-changed-in-the-21st-century/.

Presser, Stanley, and Johnny Blair. 1994. "Pretesting Survey Instruments: An Overview of Cognitive Methods." *Sociological Methods & Research* 24 (1): 109–30. https://doi.org/10.1177/0049124194024001005.

Public Opinion Research, American Association for. 2022. "Herding in Polling: A Report by the American Association for Public Opinion Research." https://aapor.org/wp-content/uploads/2022/12/Herding-508.pdf.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Siena College News. 2024. "A Perfect Partnership in Polling." https://www.siena.edu/news/story/a-perfect-partnership-in-polling/.

Siena College Research Institute. 2024. "New York Times/Siena College National Poll - October 2024." https://scri.siena.edu/2024/10/08/new-york-times-siena-college-national-poll-3/.

Silver, Nate. 2024. "Trust a Pollster More When It Publishes 'Outliers'." 2024. https://www.natesilver.net/p/trust-a-pollster-more-when-it-publishes.

Stan Development Team. 2023. *Posterior Predictive Checks*. https://mc-stan.org/docs/stan-users-guide/posterior-predictive-checks.html.

StataCorp. 2024. "Bayesian Predictions in Stata." https://www.stata.com/features/overview/bayesian-predictions/.

The New York Times and Siena College Research Institute. 2020. "New York Times/Siena Poll Methodology - June 2020." https://int.nyt.com/data/documenttools/nyt-siena-poll-methodology-june-2020/f6f533b4d07f4cbe/full.pdf.

The New York Times Editorial Board. 2024. "Election Polls and the Risks of Misinterpretation in a Trump Vs. Harris Race." *The New York Times*. https://www.nytimes.com/2024/10/23/opinion/election-polls-results-trump-harris.html?unlocked_article_code=1.UU4.pFkQ.F2hD-woxmiEj&smid=url-share.

Valliant, Richard. 2019. "Post-Stratification or Non-Response Adjustment?" *Survey Practice* 12 (1). https://doi.org/10.29115/SP-2019-0001.

WHYY News. 2024. "What Is the Electoral College and How Does the u.s. Use It?" 2024. https://whyy.org/articles/electoral-college-united-states-presidential-election/.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of*

Open Source Software 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.