Kamala Harris Rises as the 2024 Election Front Runner: A Poll-of-Polls Forecast*

How Poll Aggregation Reveals US Presidental Race Trends

Yunkyung Ko

October 22, 2024

In the anticipation of the 2024 US presidential election, this paper intends to analyze and predict the level of support that Kamala Harris will gain. The support polls for Harris remained relatively constant. However, there were some variations detected when accounting for pollster and geographical differences. Our objective is to propose a reliable electoral prediction considering those variations help global stakeholders take reasonable preemptive measures.

Table of Contents

---

1 Introduction

The US presidential election is an event that receives a lot of international attention due to its far-reaching implications beyond the country's borders. The effects of US elections are not only related to economy and international relations around the world, but also link to social and environmental issues such as climate change (Bijune and Ha 2024). In the anticipation of the 2024 US electoral competition, this paper is aimed at predicting possible outcomes of the election by analyzing the level of support that Kamala Harris will gain.

We forecast the support of Kamala Harris based on the polling results at the national and state levels, and apply a linear regression and a Bayesian approach. The main parameter of interest is the proportion of vote or support that Harris received in surveys, which is traced over time. By considering the effect of changes in poll-making organizations and geographical distinctions, our objective is to correct for variation across different voter bases with the pooling the polls approach (Jackman 2024).

Our initial linear model examining the support for Harris over time suggests that the rate of support remained relatively stable around 47%, with no significant increase or decrease. However large variability was detected, as seen in the spread of points around the fitted line. Consequently, pollster-specific effects and state-level random effects were added to the model but resulted in even higher variations. Some pollsters or states consistently reported higher or lower support for Harris compared to others, proposing the significance of considering distinct pollster and state environment when interpreting electoral polling results.

These prognostics provide much more than merely forecasting the election in question. Hinting the trajectory the US might take in matters of foreign policy, economics, and global politics, predictions enable stakeholders worldwide to take preemptive measures for the resulted changes

of a newly elected government (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 US presidential election(CSIS 2024).

The paper is structured as follows. Initially, Section 2 and Section 3 explores the data and methodology used, including filtering and modeling techniques applied to the polling data. Following that, Section 4 presents the results from the linear and Bayesian models, while the next section Section 5 discusses the broader implications of these findings. Finally, the paper concludes with remarks on future directions for research and applications of these models Section 5.

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 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. Wickham et al. (2019) was used for data manipulation and visualization and Goodrich et al. (2024), Arel-Bundock (2022) was respectively used for Bayesian modeling and generating model summaries. For visualizing results, Wickham (2016) was used and Zhu (2024) helped format tables for presentation. These packages provided a framework for efficient data processing, modeling, and reporting.

(Data last updated on [Day] [Month] 2024.) - [!!! MODIFY AT FINAL !!!]

2.2 Measurement

[!!! MODIFY OBSV. NUMBERS AFTER DATA UPDATES !!!] - ERASE AT FINAL

The original dataset sourced from FiveThirtyEight (FiveThirtyEight 2024) aggregates a wide range of poll results (15971 observations, based on dataset available at Oct 21, 22:00). 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.

Each poll represents a predictor of an actual event, namely voter opinion at a particular moment. Nevertheless, like all survey results, the raw data is susceptible to many potential limitations including the following: sampling error, variation in polling methods, distortion because of inappropriate survey responses such as missing data or response from respondents who misunderstood the questions (Alexander 2023).

While applying several filters to the original dataset such as restricting to those with a numeric grade of 2.7 or higher or pollsters with more than 30 polls improves data reliability, certain limitations still exist. Selection bias and sampling error remains as a concern, since polls always represent only part of the population. Differences in the way different organizations conducted their polling might introduce more inconsistencies. Finally, by focusing our attention on post-declaration polls only, we exclude earlier trends that could add more insight into how Harris' support has evolved over time.

2.3 Outcome variables

2.3.1 pct

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 40% 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 'pct' Variable (Raw Data)

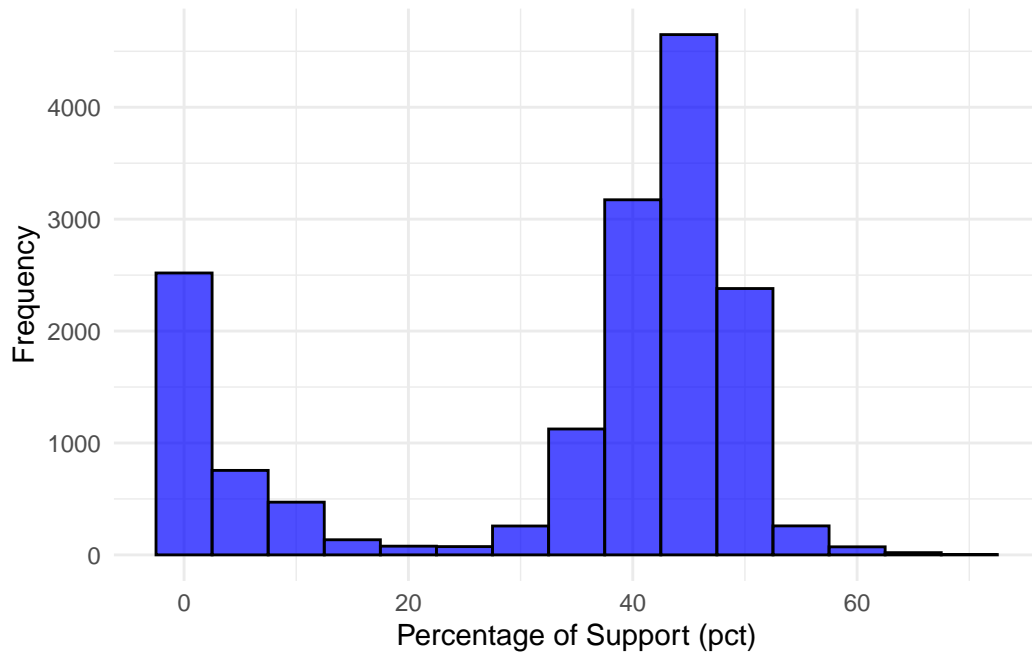| mean	median	min	max	sd	n
33.68017	42	0	70	18.03627	15971

Figure 1: Distribution of the 'pct' Variable (Raw Data)

Table 2: Summary Statistics for the 'pct' Variable (Filtered Data)

mean	median	min	max	sd	n
47.28909	48	38	65.3	3.854899	165
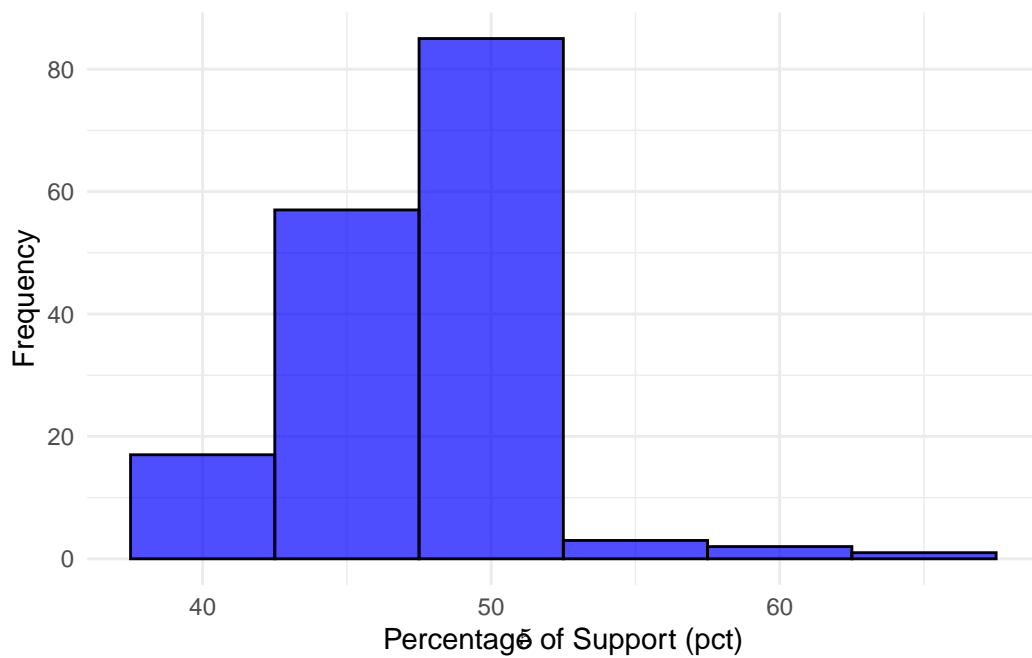
Figure 2: Distribution of the 'pct' Variable (Filtered Data)

2.4 Predictor variables

Min	Max
1/1/23	9/9/24

[!!! MODIFY DATE RANGE AFTER DATA UPDATES !!!] - ERASE AT FINAL

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 ranges from 23 July, 2024 to 14 October, 2024 (Table 4). Figure 4 shows that overall, polling is conducted regularly but intensifies around specific dates.

Table 4: Summary Statistics for the 'end_date' Variable (Filtered Data)

Min	Max
2024-07-23	2024-10-14

2.4.2 pollster, 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 222 distinct pollsters and 54 distinct states. After filtering for high-quality polls and assigning 'other' for states with fewer than 60 polls, the analysis data contains 3 distinct poll-making organizations and 19 geographical distinctions as shown in Table 6.

Table 5: Number of Distinct Values for the 'pollster', 'state' Variable (Raw Data)

Pollster	State
222	54

Table 6: Number of Distinct Values for the 'pollster', 'state' Variable (Filtered Data)

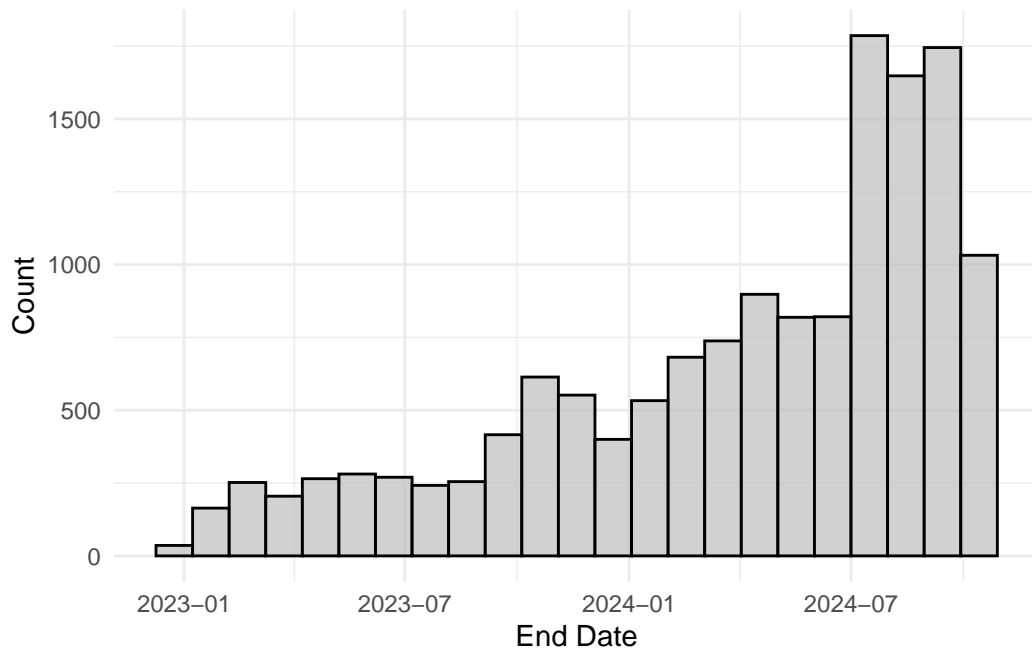| Pollster	State
3	19

Figure 3: Distribution of the 'end_date' Variable (Raw Data)
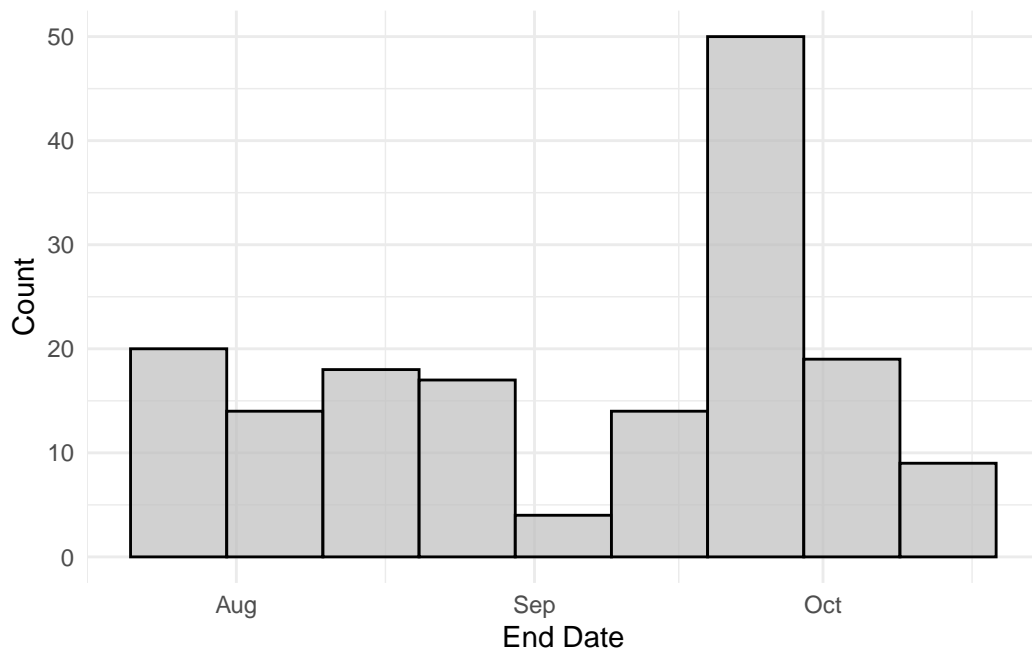

Figure 4: Distribution of the 'end_date' Variable (Filtered Data)

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 B.
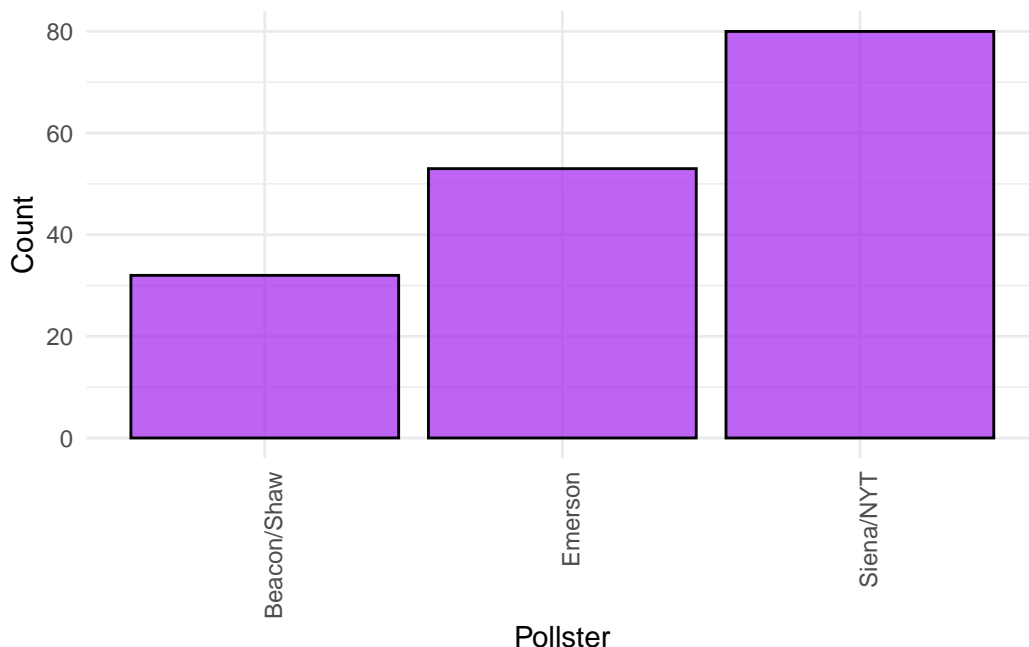

Figure 5: Distribution of the 'pollster' Variable (Filtered Data for Harris)

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).

2.4.3 pollscore

This variable was another factor we had put into consideration to check whether the validity of the polling questions affects the polling results. The 'pollscore' variable represents the score or reliability of the pollster in question. The numeric values are the error and bias that can be attributed to the pollster, which means negative numbers are better. Table 7 and Figure 7 suggests that while the majority of the polls are moderately to highly qualitative in the original dataset, a fraction of the polls with low-quality or no scores could add noise or uncertainty to the analysis.
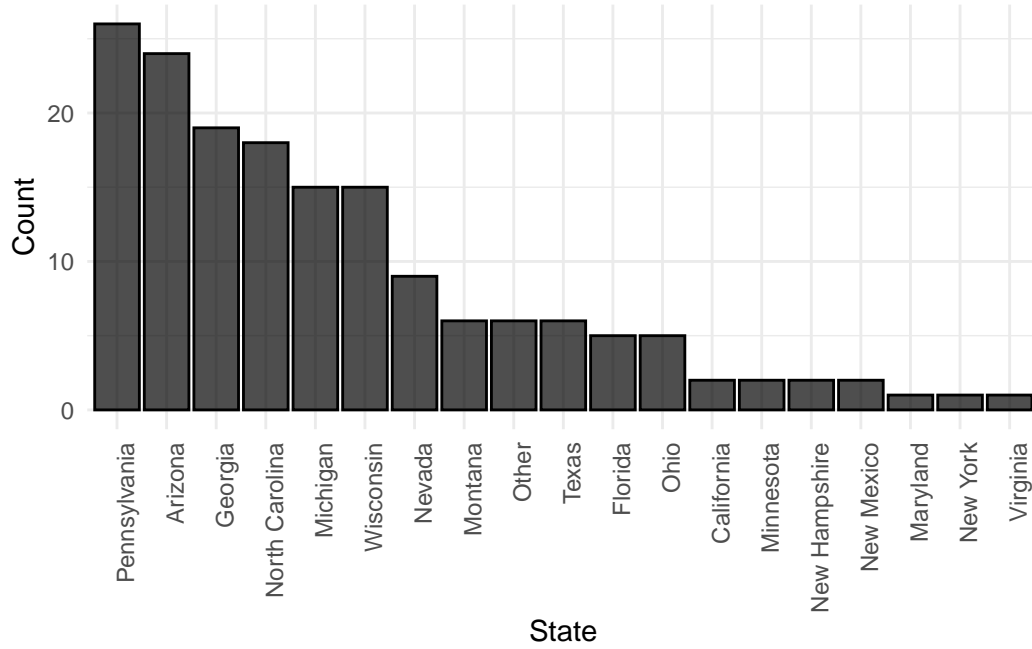
Figure 6: Distribution of the 'state' Variable (Filtered Data for Harris)

Table 7: Summary statistics for the 'pollscore' Variable (Raw Data)

mean	median	min	max	sd	n
-0.3804771	-0.3	-1.5	1.7	0.7021844	15971
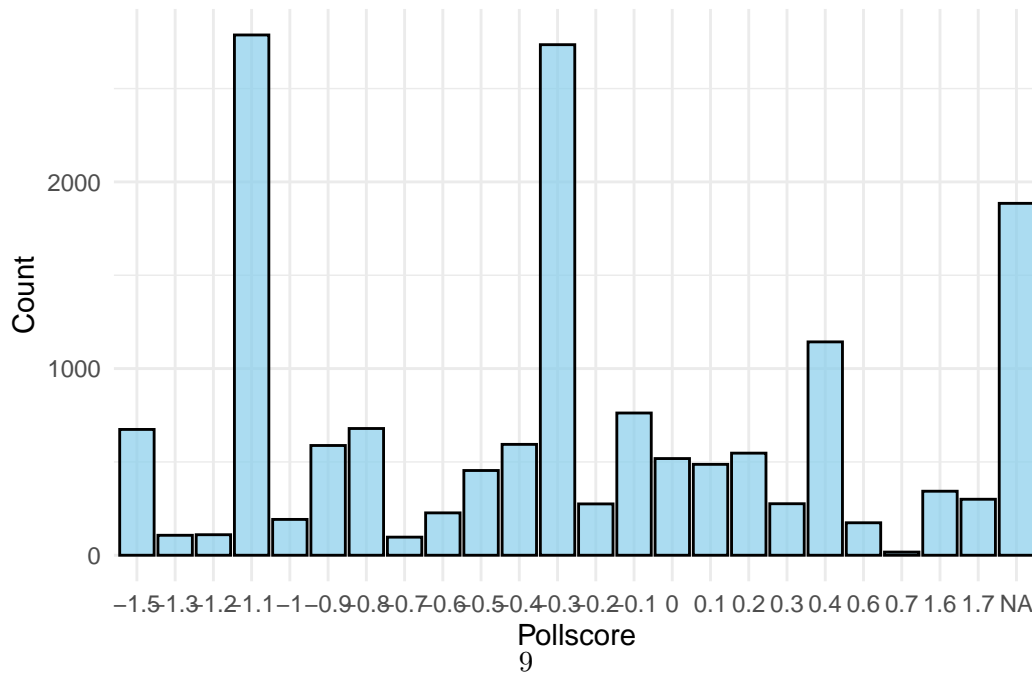
Figure 7: Distribution of the 'pollscore' Variable (Raw Data)

9

After the filtering to polling data of high-quality polling organizations, we can find that the overall value and standard deviation of pollscores went down in Table 8. This implies that the polling data narrowed down to the responses from more reliable survey questions. Figure 8 also indicates that the data cleaning process effectively excluded less reliable sources, which
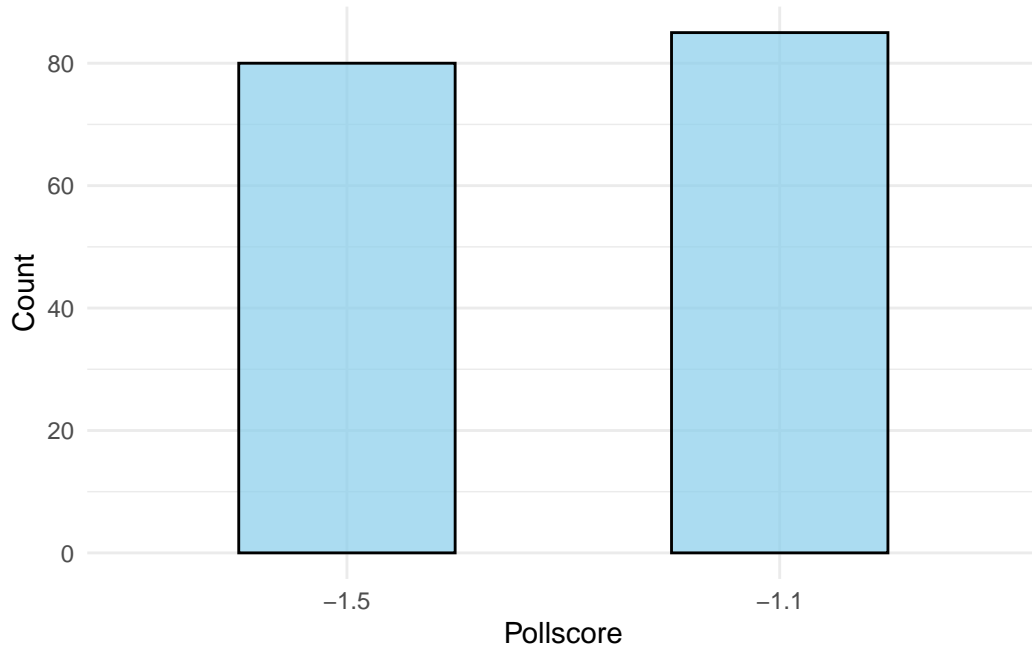
Figure 8: Distribution of the 'pollscore' Variable (Filtered Data)

3 Model

The goal of our modelling strategy is to estimate Kamala Harris's support percentage in the 2024 US election polls, accounting for potential variations over time, as well as across different pollsters and states. The model balances complexity with interpretability, incorporating both linear and Bayesian frameworks to capture patterns in the data. Background details and diagnostics are included in Appendix C.

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), which makes Bayesian modeling available through the use of Stan's strong inference engine. To validate the models, RMSE and WAIC have been made use of to check the goodness of fit; 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.

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 exploring interactions between pollsters and states.

4 Results

Our results are summarized in Table 9 and Table 10. [!!! ADD MORE WORDS !!!]

5 Discussion

[!!! NEED MODIFICATION !!!]

5.1 Stability in Support Levels for Harris

Kamala Harris' polling numbers show stability, a common trend for well-known candidates whose electorate solidifies in its opinion over time. Articles from The New York Times (Goldmacher and Weisman 2024) note that "once voters form opinions about candidates, those views rarely fluctuate significantly unless a major event occurs, such as a scandal or a high-profile

Table 9: Linear models of support percentages for Harris based on date and pollster

	Linear by Date	Linear by Date, Pollster
(Intercept)	189.49	−48.99
	(295.49)	(288.71)
end_date	−0.01	0.00
	(0.01)	(0.01)
pollsterSiena/NYT		−3.11
		(0.74)
Num.Obs.	133	133
R2	0.004	0.135
R2 Adj.	−0.019	0.096
Log.Lik.	−378.353	−369.442
ELPD	−381.6	−373.4
ELPD s.e.	13.9	12.8
LOOIC	763.1	746.8
LOOIC s.e.	27.8	25.7
WAIC	763.1	746.7
RMSE	4.14	3.87

Table 10: Bayesian models of support percentages for Harris based on pollster and state

	Bayesian with Pollster	Bayesian with Pollster, State
(Intercept)	−0.10	−0.04
	(0.13)	(0.12)
Sigma[pollster × (Intercept),(Intercept)]	0.05	0.02
	(0.06)	(0.03)
Sigma[state × (Intercept),(Intercept)]		0.06
		(0.02)
Num.Obs.	133	133
ICC	1.0	1.0
Log.Lik.	−798.443	−573.758
ELPD	−805.1	−601.8
ELPD s.e.	68.2	19.8
LOOIC	1610.2	1203.7
LOOIC s.e.	136.5	39.5
WAIC	1610.3	1198.3
RMSE	0.04	0.02

policy shift. In the case of Harris, her polling numbers have been consistent, only changing slightly since she announced in 2024 (Figure 4). This suggests that her supporters are more or less consistent and unlikely to change much as the election draws closer.

5.2 Pollster-Specific Variability

Various polling entities, including Siena/NYT and Quinnipiac, use distinct methodologies, resulting in potential systematic discrepancies within their outcomes. According to Michael Wines, the inherent biases of each pollster result in different interpretations of the elections prediction (Wines 2024). For instance, Siena/NYT's live-interview method often skews slightly older and more conservative in its responses compared to online polling methods used by other organizations (Institute 2024). This variability makes clear that when interpreting for polling results, the source of the poll must be included at minimum.

5.3 State-Level Differences

Certain states, particularly battleground states like Florida, Pennsylvania, and Georgia, show significantly different polling results compared to national trends due to regional issues and voter priorities (Piper 2024). These reflect state-specific issues, such as local economies, immigration, and access to healthcare. Surveys carried out within these states often show wider variation in candidate support, which makes state-level analysis necessary for more accurate predictions. Incorporating state-specific factors into electoral models allows a better representation of the diverse political landscapes seen across the United States, hence improving the precision of prediction in these key areas.

5.4 Weaknesses and next steps

Future studies could focus on the incorporation of time-varying effects to capture dynamic shifts in public opinion, especially at times of large campaign events. Interaction terms between pollster and state can be included in the model, since some polling organizations may be more effective or influential in specific regions. This can give further explanation to how pollsters and regional dynamics affect overall support for a candidate. Moreover, extending the model to incorporate voter demographics, such as age, gender, and education, could show which segments of the population are driving changes in support.

While our model provides a foundation to understand Harris' polling support, it will take further refinements are necessary to enhance its accuracy and applicability. The combination of hierarchical modeling with Bayesian methods is effective at accounting for the heterogeneity across pollsters and states, yet there remain considerable directions of inquiry to explore at the intersection of polling methodology, voter behavior, and regional electoral dynamics.

A Appendix

B Additional data details

B.1 Pollster Methodology Overview and Evaluation

The New York Times/Siena College polling partnership, the polling organization that accounted for the majority of polls in our analysis (Figure 5), conducts polls tailored for specific elections, such as state or national races ("New York Times/Siena Poll Methodology - June 2020" 2020). Their sample size typically includes 600 to 1000 likely voters per poll, with oversampling in battleground states to capture regional nuances ("New York Times/Siena Poll Methodology - June 2020" 2020). The methodology uses random-digit dialing (RDD) for landlines and mobile phones to ensure representative sample coverage across demographics. In addition, online surveys are administered to complement phone-based responses, ensuring broader accessibility ("New York Times/Siena Poll Methodology - June 2020" 2020). The stratified random sampling approach is employed, where the population is divided into strata (based on demographic variables like race, education, and geography), and a random sample is drawn from each stratum (Alexander 2023). This allows for precision in reflecting the political leanings and key demographic shifts in specific regions.

The organization intends to enhance transparency in how public opinion is assessed, ensuring that questions are carefully designed to represent contemporary political discussions, and that the terminology is polished through a process of iterative testing to achieve clarity. They devote extensive resources to cognitive testing to ensure question wording reflects what the public thinks (Institute 2024). Their polling methodology stands out in that its strategic focus on using representative samples reflect political leanings and demographics of a region for more contextual and precise polling. Siena/NYT has its reputation for accurately predicting key battleground state outcomes during previous elections, such as Florida in 2016(News 2024).

The limitations of Siena/NYT's methodology are the challenge of polling itself. Since polling is a "snapshot in time", the results can fluctuate based on recent political events or campaign dynamics. Additionally, while the effort to represent a broad demographic is laudable, there are still issues with nonresponse bias in polling-particularly among the hard-to-reach voter or voters suspicious of polling organizations themselves (Center 2023).

B.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. This could alleviate bias and make the sample more representative of the population (Center 2023). The data collection process will encompass both telephone

surveys and internet polling, thereby effectively engaging a diverse range of voters (Center 2023). Survey respondents would include older populations via conventional methods and younger, technologically intellectual individuals through digital platforms. This multifaceted approach enhances the reliability and inclusiveness of the gathered data.

In addition to robust sampling, the methodology incorporates weighting methods that account for groups that are underrepresented, ensuring that the outcomes are not skewed by sampling errors (Center 2023). 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 (Center 2023). The inclusion of external factors, such as major political events, debates, or sudden economic shifts, would help the model remain responsive to rapid changes in voter sentiment.

B.3 Idealized Survey

The proposed survey questionnarie design is provided in the following link:

C Model details

C.1 Posterior predictive check

In the first posterior predictive check (Figure 9a), 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 9b), 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).

C.2 Diagnostics

Figure 10a 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
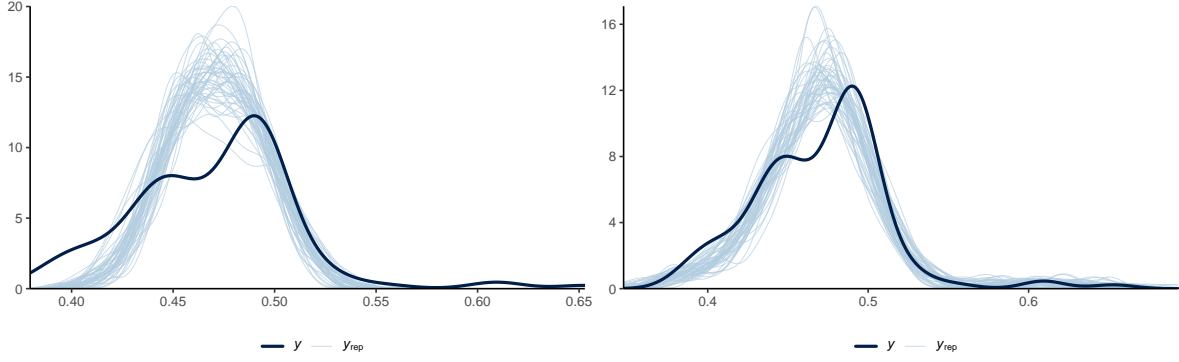
(a) Posterior prediction check   (b) Posterior prediction check

Figure 9: Examining how the Bayesian model fits, and is affected by, the data

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 10b 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.
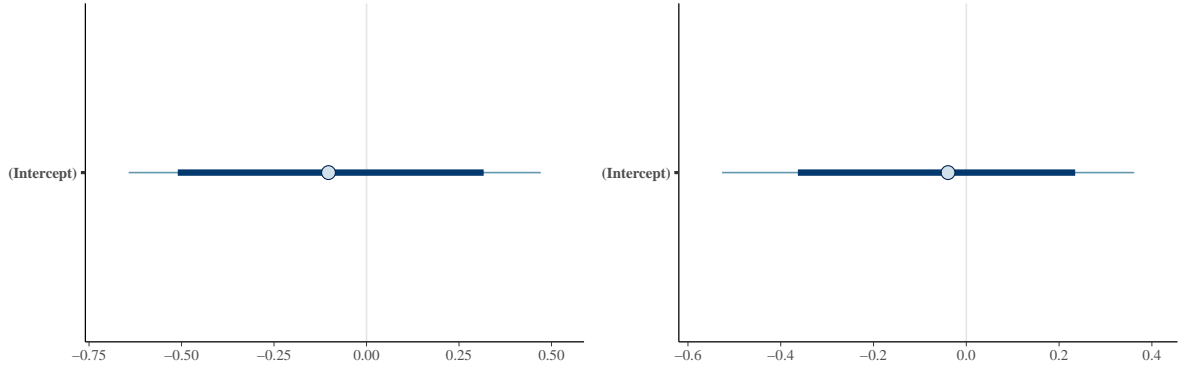
(a) Trace plot   (b) Rhat plot

Figure 10: 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.

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.

Center, Pew Research. 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/.

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

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.

Goldmacher, Shane, and Jonathan Weisman. 2024. "Harris Gains in Polling Against Trump, According to National Poll." *The New York Times.* https://www.nytimes.com/2024/10/08/us/politics/harris-trump-poll-national.html?campaign_id=60&emc=edit_na_20241008&instance_id=0&nl=breaking-news&regi_id=193003212&segment_id=179873&user_id=2d04b9e4643940584b7476aee2111e62.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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

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

"New York Times/Siena Poll Methodology - June 2020." 2020. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://int.nyt.com/data/documenttools/nyt-siena-poll-methodology-june-2020/f6f533b4d07f4cbe/full.pdf.

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

Piper, Jessica. 2024. "Trump Maintains Narrow Lead in Key Battleground States, NYT Poll Finds." *Politico.* https://www.politico.com/news/2024/05/13/trump-narrow-lead-battleground-00157541.

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

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

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.

Wines, Michael. 2024. "Polling Problems: How Margins of Error Could Mislead in the 2024 Election." *The New York Times.* https://www.nytimes.com/2024/10/14/us/elections/poll-problems-margin-of-error.html.

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