

Predicting the 2024 US Presidential Election: A Polling-Based Forecast*

Trump will get support about 46.87% by pollster and have probability of 31%
lead the election over Harris

Tianrui Fu Yiyue Deng Jianing Li

November 4, 2024

This paper uses advanced statistical techniques, including generalized linear regression and Bayesian models, to forecast the 2024 U.S. Presidential Election. Our analysis of polling data predicts an average support rate of 46.87% for Donald Trump and 48.50% for Kamala Harris across multiple polling sources. Among the 29 polling organizations examined, only 9 anticipate a higher support rate for Trump over Harris; however, the Bayesian model still assigns a 31% probability of Trump leading, underscoring the competitiveness of the race. These findings emphasize the importance of credible pollsters and robust statistical methods in electoral forecasting, highlighting the volatility of voter sentiment and the value of continual methodological refinement.

Table of contents

1	Introduction	3
2	Data	4
2.1	Overview	4
2.2	Cleaning Data	4
2.3	Measurement	5
2.4	Outcome variables	5
2.5	Predictor variables	6
2.5.1	Pollster	6
2.5.2	Poll Score	7
2.5.3	Transparency Score	8

*Code and data are available at: <https://github.com/FrankFU323/U.S.-presidential-election.git>

2.5.4	Numeric Grade	9
2.5.5	End date	10
2.5.6	Relationship between pollster, pollscore and end date	11
2.5.7	Relationship between transparency score and numeric grade	12
3	Model	13
3.1	Model Selection	13
3.2	Model set-up	13
3.2.1	Model 1 – GLM	14
3.2.2	Model 2 – Bayesian model for Trump	14
3.3	Model justification	15
4	Results	15
4.1	Result of model for analysis data	16
4.2	Result for model prediction	17
5	Discussion	19
5.1	Summary of findings	19
5.2	Evaluating Predictive Models in Electoral Forecasting	20
5.3	Quantifying Pollster Influence on Electoral Outcomes	21
5.4	Weaknesses and next steps	21
A	Appendix	23
A.1	Methodology Overview and Evaluation of YouGov Polling	23
A.2	Idealized Survey Methodology	25
A.3	Clean data	28
A.4	Results of model	28
A.4.1	Table for summary of model results	28
A.4.2	Bayesian model for Harris data set	30
A.4.3	Predictions for both Trump and Harris	31
B	Model details	33
	References	34

1 Introduction

With the 2024 U.S. Presidential Election on the horizon, the nuanced interpretation of voter sentiment is paramount for anticipating, understanding, and informing the strategies of campaign teams and political analysts about electoral outcomes. This study offers a comprehensive analysis of polling data to forecast voter support for potential President, Donald Trump, focusing on how various influences—ranging from media portrayals to strategic campaign adjustments and significant socio-political developments—mold public perception and shape electoral dynamics. By utilizing a robust dataset aggregated from diverse polling sources, this research endeavours to illuminate the predictors of voter behaviour. We delve into both the quantitative metrics of polling results and the qualitative aspects, such as the credibility of polling agencies and the specific timing of polls. It is through this dual approach that our work seeks to contribute to the academic conversation by addressing the often-overlooked subtleties that affect the interpretation and reliability of polling data.

At the heart of our inquiry is the modelling of Donald Trump’s voter support, which incorporates a broad spectrum of variables including the trustworthiness of pollsters, geographic and demographic influences, and the temporal context relative to Election Day. By applying sophisticated statistical techniques such as generalized linear regression and Bayesian analysis, our study seeks to distill clear, actionable insights from the complex interplay of these factors. This methodology will not only highlight the dynamics of voter sentiment but also clarify how each determinant, from polling methodology to regional variations, impacts public opinion.

Our analysis reveals a pattern of gradually increasing support for Trump as the election nears, accentuated by the methodological integrity of the polls. This trend underscores the critical role of employing high-quality polling data to predict electoral results with greater accuracy and to strategize more effectively. Access to such data allows campaign strategies to be not only data-driven but also precisely targeted, maximizing impact where it is most effective. This approach enables campaigns to rapidly adapt to changing voter sentiments and demographic shifts, thus enhancing their responsiveness. Thus, We believe that these insights gained from rigorous data analysis are essential for refining campaign strategies and ensuring a nuanced understanding of the electoral landscape. Moreover, these insights are essential for improving the precision of electoral forecasts, thereby providing campaigns with a clearer understanding of where to allocate resources to achieve the greatest effect and sway electoral outcomes. Precise forecasts enable campaigns to allocate resources more effectively, potentially swaying the election outcomes. Moreover, the implications of this study extend to practical applications in political strategy. By elucidating the primary influences on polling statistics, the research provides campaign strategists and political consultants with the foundations to tailor their initiatives more adeptly, aiming to sway voter opinions in pivotal states and demographics.

The paper will begin with an in-depth review of our data sources and the methodologies employed, ensuring clarity and methodological transparency. The rest of the paper will be structured as follows: Data 2 provides a detailed examination of our findings, progressing from the initial section, which analyzes the data compilation and preprocessing methods. Model 3

discusses our modelling approach and justification, Results 4 shows the result of predictions, and Discussion 5 delves into the statistical analysis and its implications, with each section building upon the last to paint a comprehensive picture of the current electoral landscape.

2 Data

2.1 Overview

We conducted our analysis using the statistical programming language R Core Team (2023), employing a suite of packages designed to facilitate efficient data manipulation and modeling. The analysis framework was based on the template from Alexander (2023) complemented by the primary packages used which include tidyverse Wickham et al. (2019) for data manipulation and visualization, dplyr Wickham et al. (2023) for data transformation, rstanarm Goodrich et al. (2022) for Bayesian inference, arrow Richardson et al. (2024) for data integration, modelr Wickham (2023) for modeling within the tidyverse framework, modelsummary Arel-Bundock (2022) for generating model summaries, ggplot2 Wickham (2016) for creating advanced graphics, kableExtra Zhu (2024) for enhancing table outputs, knitr Xie (2023) for dynamic report generation and here Müller (2020) to help read the parquet. Our analysis is based on the latest polling outcomes available from the website FiveThirtyEight (2024), which includes a dataset of 52 variables and 17765 observations, covering aspects such as pollster, poll score, and more.

Table 1: The example of chosen variable dataset

pollster	pollscore	numeric_grade	transparency_score	end_date	pct
Quinnipiac	-0.5	2.8	9.0	2024-09-29	50.0
YouGov	-1.1	3.0	9.0	2024-09-24	44.0
Siena/NYT	-1.5	3.0	9.0	2024-10-26	51.0
SurveyUSA	-1.3	2.8	8.0	2024-09-18	42.0
CES / YouGov	-1.1	3.0	10.0	2024-10-25	51.0
The Washington Post	-1.2	3.0	9.0	2024-10-30	47.0
YouGov	-1.1	3.0	9.0	2024-10-17	38.0
Siena/NYT	-1.5	3.0	9.0	2024-10-23	47.0
Ipsos	-0.9	2.8	9.5	2024-10-22	47.0
AtlasIntel	-0.8	2.7	6.0	2024-11-02	50.4

2.2 Cleaning Data

We have cleaned the data set and the detailed procedure see from the Appendix A.3.

2.3 Measurement

In political polling, measurement plays a critical role by translating complex and often conflicting voter sentiments into quantifiable data that can inform electoral predictions. The fundamental challenge in measurement lies in transforming a mix of opinions—where individual voters may hold conflicting views or ambivalence about candidates—into a single, structured outcome variable that reflects electoral support accurately. This process begins with careful design of polling methodologies aimed at capturing voter preferences through survey questions that effectively reduce complex sentiments into measurable data points.

This dataset, sourced from FiveThirtyEight (2024), maintains high standards in polling data collection, ensuring broad voter representation and thorough documentation of variables like pollster identity, polling dates, and trust levels. In our dataset of 803 observations focused on Donald Trump’s polling results, we examine 38 key variables relevant to understanding his electoral support, including the numeric grade, transparency score, poll score, pollster, end date, and percentage of support. Each variable is measured rigorously to capture both the nuances and shifts in public opinion accurately.

The results are represented as proportions of support for each candidate per poll, adjusting over time to account for factors such as changing voter sentiment in response to candidate speeches or approaching Election Day. This transformation from raw opinions to structured data enables us to track and model trends with greater predictive reliability.

In addition to measuring the outcome variable (electoral support), we also ensure accurate measurement of predictor variables, such as pollster transparency scores and numeric grades, which may affect a poll’s reliability. By applying these measurement principles, we systematically convert public opinion into a structured dataset that provides a robust basis for predictive modeling. Ultimately, this rigorous approach allows us to build a Bayesian model to evaluate how various pollsters may influence final election outcomes, thereby bridging individual opinions and broader electoral predictions with a high degree of credibility.

2.4 Outcome variables

The `pct` is the percentage of the vote or support that the candidate received in the poll and keep integer. The Figure 1 shows the distribution of this variable, labeled “`pct`,” with most values clustered around 47%. The distribution appears approximately normal but has a slight skew due to a few higher values. Overlaying the Figure 1 is a density curve that illustrates the general shape of the data. Accompanying the Figure 1 is Table 2, a summary statistics table for the “`pct`” variable, which indicates a mean of 46.98%, a median of 47%, and a standard deviation of 3.8. The range spans from 31% to 70%, showing some spread in the data, with most observations relatively close to the mean, suggesting that while there is some variability, support percentages are largely concentrated around the central values.

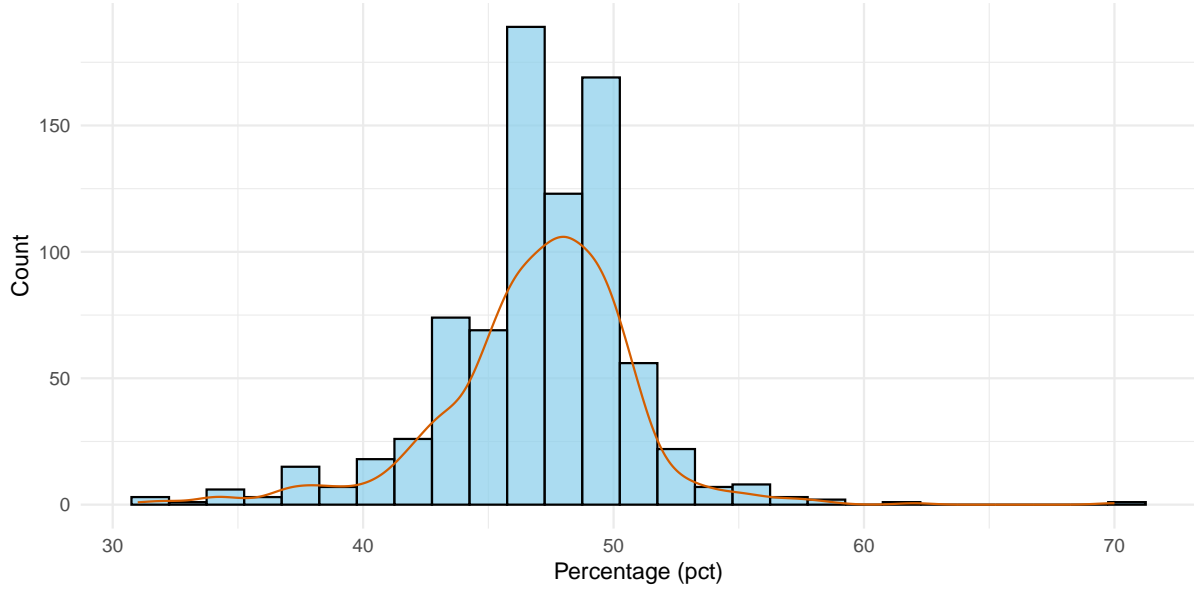


Figure 1: Distribution of percentage

Table 2: Summary Statistics for 'pct' Variable

Mean	Median	Lower	Upper	SD
46.98	47	31	70	3.8

2.5 Predictor variables

2.5.1 Pollster

The pollster is the name of the polling organization that conducted the poll included Marquette Law School, CNN/SSRS and etc. In Figure 2, the predictor variable is the count of polls conducted by each pollster. The bar chart displays the number of polls released by different polling organizations, ordered from highest to lowest frequency. The bars range in color from light blue to dark blue, indicating the frequency of polls conducted by each organization. Siena/NYT and YouGov are the most active pollsters, with 136 and 121 polls conducted, respectively. Emerson and AtlasIntel also have high polling frequencies, with 73 and 82 polls. In contrast, several pollsters, like Christopher Newport University and McCartney Institute/YouGov, conducted only one poll, indicating their minimal activity in comparison.

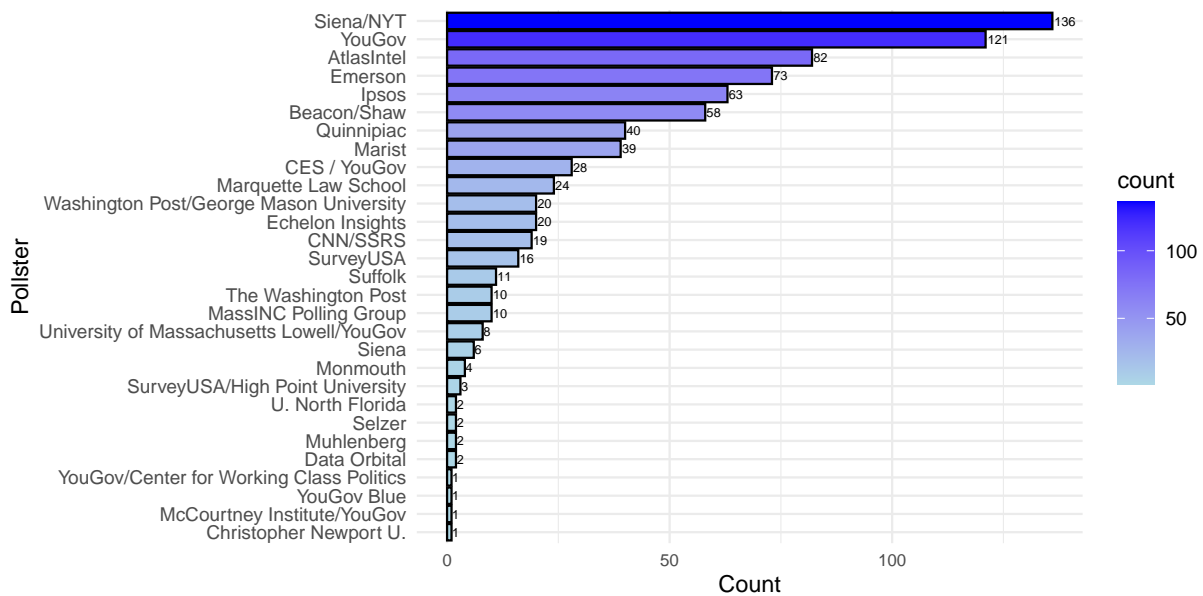


Figure 2: Frequency of each pollster

2.5.2 Poll Score

The poll score is the numeric value representing the score or reliability of the pollster in question from -1.5 to -0.5 and negative numbers of poll score are better. Figure 3 is a vertical bar chart displaying the distribution of poll scores, ranging from -0.5 to -1.5. The most frequent poll score is -1.1, indicating a high concentration around this value. Most pollsters' poll scores are below -0.8, which suggests that overall scores are relatively low, reflecting a higher level of trust in the polls. Further analysis shows that the lower the poll score, the higher the trust level. Therefore, the distribution is notably skewed, with most data concentrated in the lower range (around -1.1 and below). Higher negative scores (closer to -0.5) are relatively rare, indicating that most polling agencies demonstrated high credibility in this survey, earning broad approval from the public or voters.

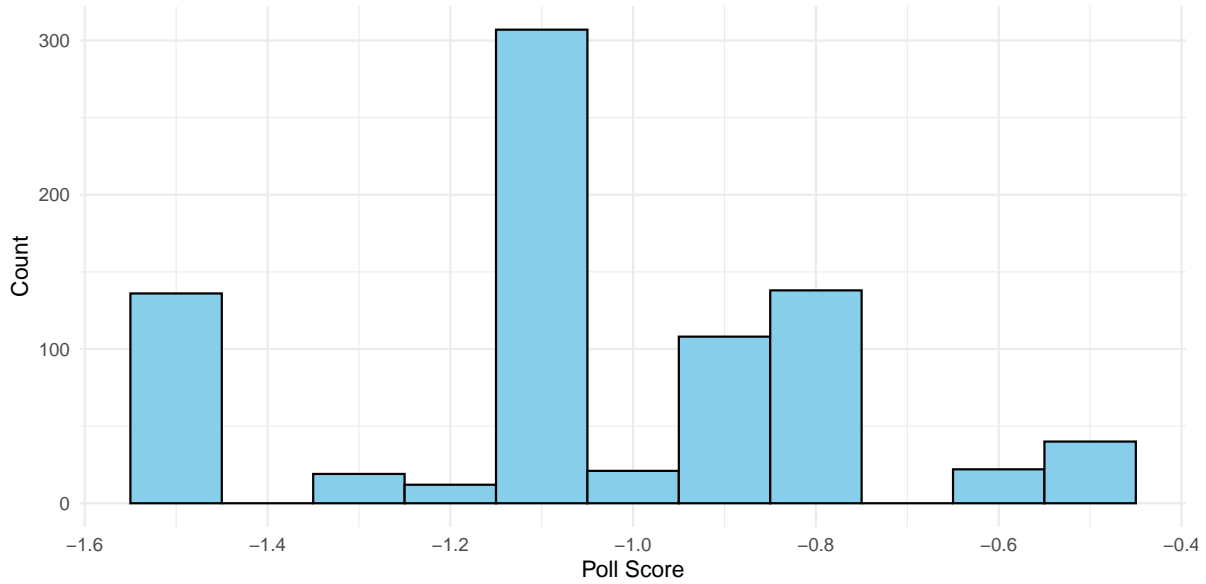


Figure 3: Distribution of Poll Score

2.5.3 Transparency Score

The `transparency_score` is the grade for how transparent a pollster is, calculated based on how much information it discloses about its polls and weighted by recency from 4.5 to 10.0. Figure 4 illustrates the distribution of “Transparency Scores” related to the 2024 election, possibly representing public or media ratings of candidate transparency, ranging from 5 to 10. The majority of scores are clustered around 8, showing that most respondents rate transparency fairly high. Scores of 6 and 7 are less frequent, while extreme scores of 5 and 10 are rare. This indicates that most voters or media sources perceive candidates as moderately to highly transparent, with scores skewed toward the higher end of the scale.

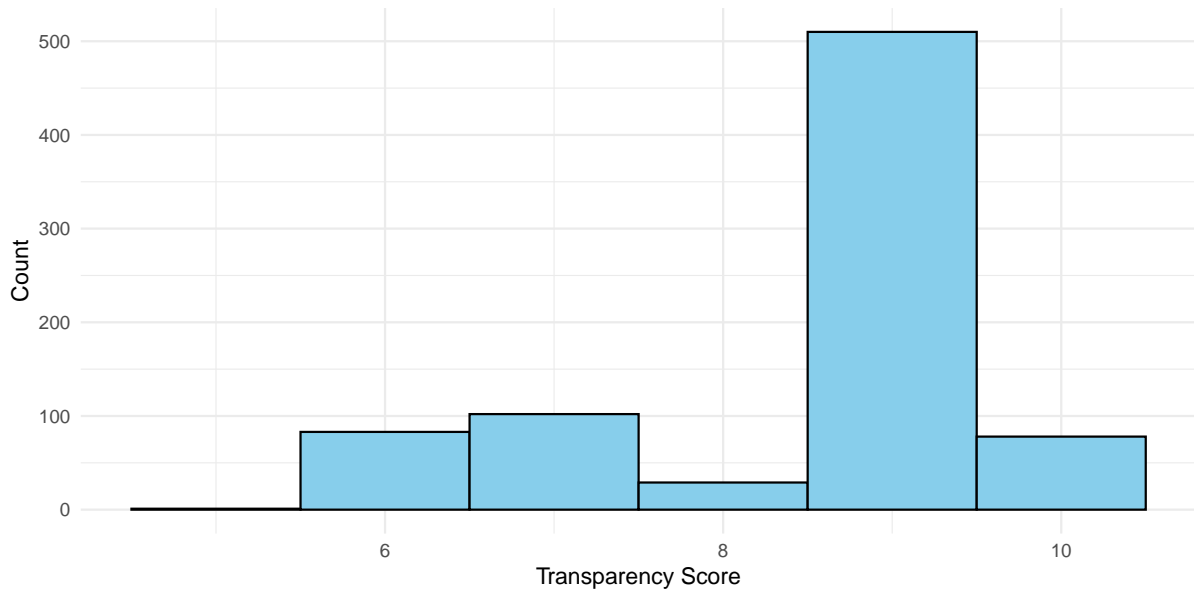


Figure 4: Distribution of Transparency Score

2.5.4 Numeric Grade

The `numeric_grade` is the numeric rating given to the pollster to indicate their quality or reliability from 2.7 to 3.0. Figure 5 displays the distribution of a “Numeric Grade” associated with the 2024 election, which could represent candidate approval ratings, public opinion scores, or voter satisfaction. The grades are primarily concentrated between 2.7 and 3.0, with a noticeable peak close to 3.0, indicating a higher frequency of favorable ratings. This distribution suggests strong public support or positive sentiment, with the majority of respondents giving relatively high scores on this variable.

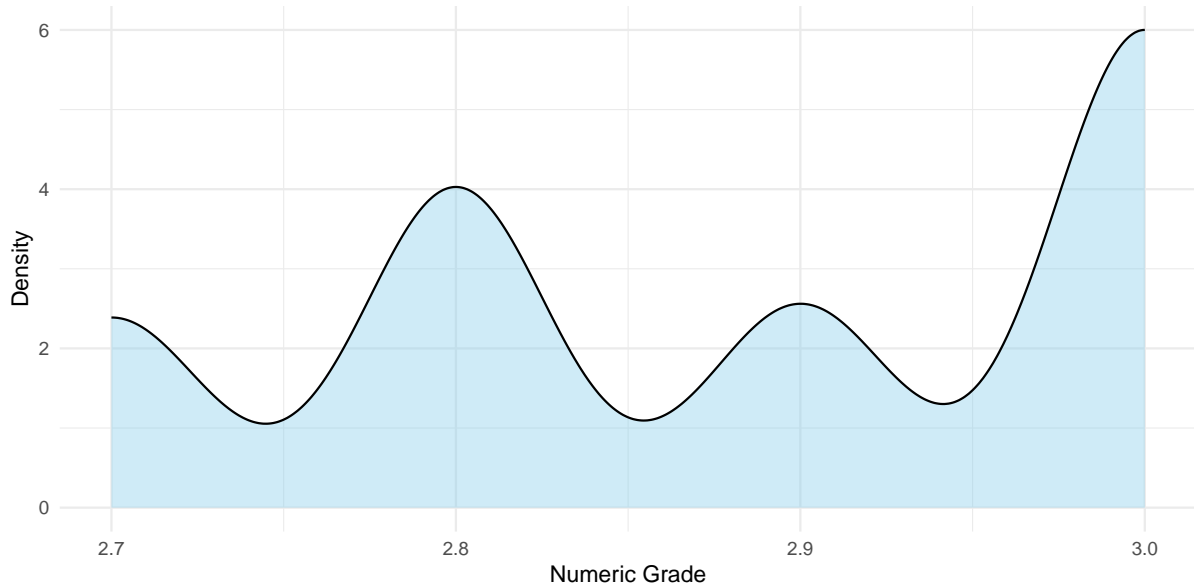


Figure 5: Density Plot of Numeric Grade

2.5.5 End date

The `end_date` is the date of polling ends. Figure 6 shows the distribution of a particular variable over time related to the 2024 U.S. election. The x-axis represents dates from August to November, and the y-axis shows the count. This variable might represent the frequency of certain events, such as news coverage, voter registrations, or polling responses. The data is relatively sparse from August to October, with small peaks in some areas. In November, however, there is a significant increase in counts, particularly in early November, possibly linked to the heightened election activity or voter engagement as Election Day approaches.

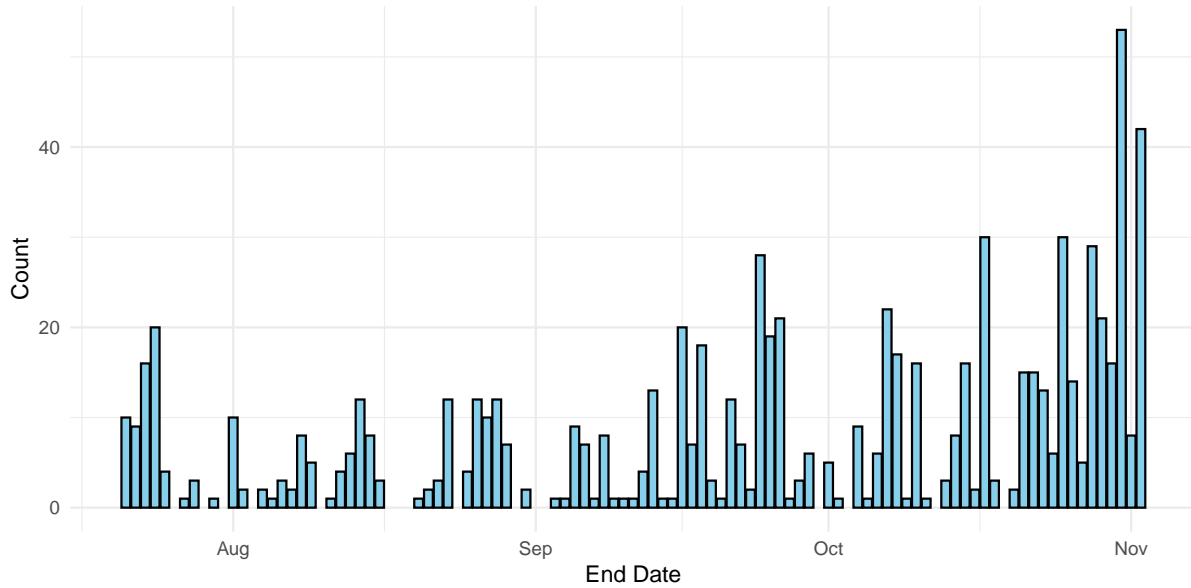


Figure 6: Frequency of Records by End Date

2.5.6 Relationship between pollster, pollscore and end date

In Figure 7, the predictor variable is the average poll score for each pollster over time, from August to October. This heat map visualizes the poll scores of each organization on different dates, with color intensity indicating the score levels—darker colors represent lower scores. The pollsters with the most consistently low average poll scores (darker colors) include Selzer, Siena/NYT, and Marquette Law School. These pollsters frequently show results lower than others over time, suggesting they may consistently lean toward one direction in their results. In contrast, pollsters like YouGov Center for Working Class Politics and YouGov Blue display lighter colors, indicating relatively higher scores across their polls.

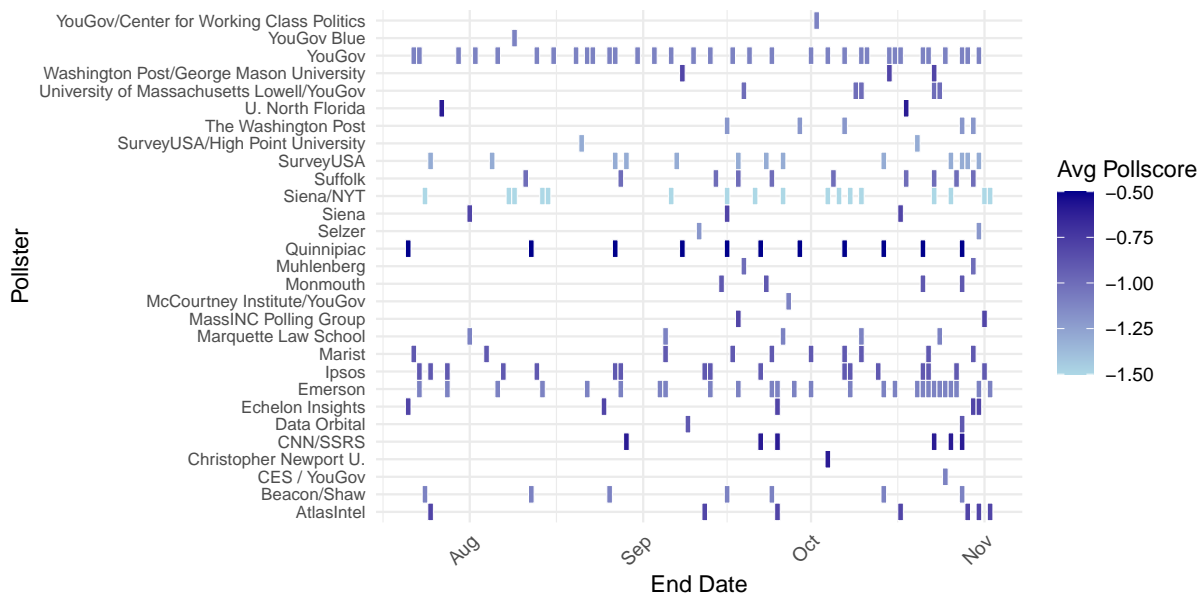


Figure 7: Average Pollscore by Pollster and End Date

2.5.7 Relationship between transparency score and numeric grade

In Figure 8, there is a clear positive relationship between transparency score and numeric grade. Pollsters with higher transparency scores, such as those scoring around 10, tend to achieve higher numeric grades, close to 3.0. Conversely, pollsters with lower transparency scores tend to have lower numeric grades, closer to 2.7. This pattern suggests that greater transparency is associated with better overall pollster ratings.

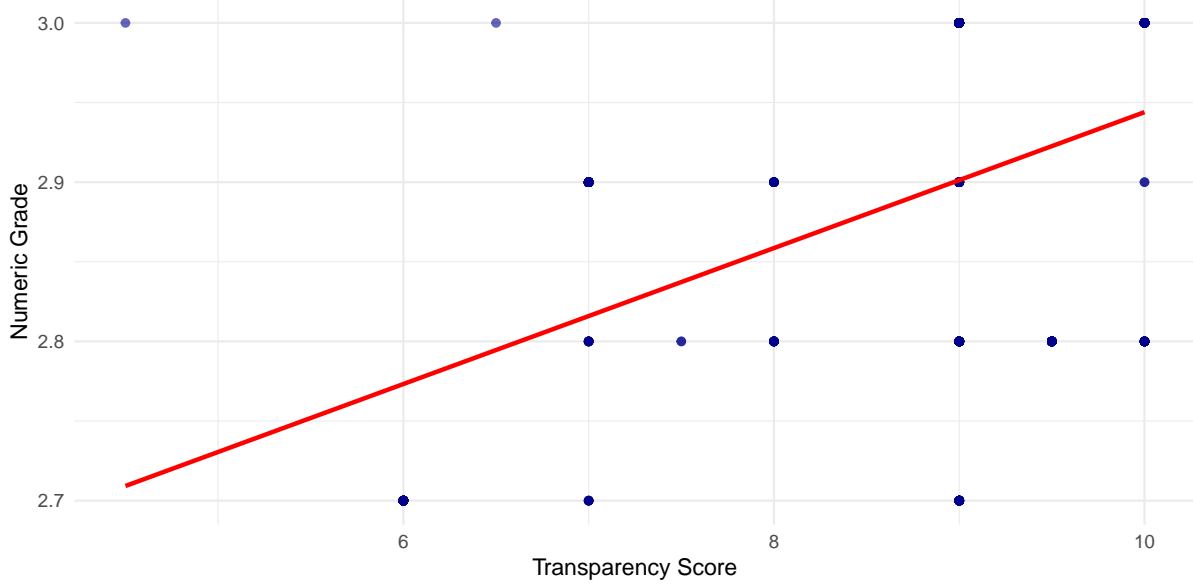


Figure 8: Scatter Plot of transparency score and numeric grade

3 Model

3.1 Model Selection

Our model aims to predict the changes in Donald Trump’s probability of winning the 2024 U.S. election over time while examining the influence of key factors, such as pollster reliability scores, transparency, and specific pollster effects, on polling results. For model selection, we initially considered a generalized linear regression to provide a straightforward analysis of the relationships between these factors. However, given the limitations of generalized linear regression in handling uncertainty and dynamic data, we ultimately chose a Bayesian model. The Bayesian approach enables the integration of prior information and dynamically updates predictions, offering greater stability and accuracy under high uncertainty. Below is a brief overview of our model.

Background details and diagnostics are included in [Appendix B](#).

3.2 Model set-up

In this Bayesian framework, we assume a normal distribution of poll results around a mean affected by key predictors: numeric grade, transparency score, pollscore, pollster and end date. Define y_i as the percentage of Donald Trump.

3.2.1 Model 1 – GLM

$$y_i = \alpha + \beta_1 \cdot \text{numeric_grade}_i + \beta_2 \cdot \text{transparency_score}_i + \beta_3 \cdot \text{pollscore}_i \quad (1)$$

$$+ \sum_{j=1}^N \gamma_j \cdot \text{pollster}_j + \delta \cdot \text{end_date}_i + \epsilon_i \quad (2)$$

where:

- α is the intercept, representing the average poll level.
- β_1 , β_2 , and β_3 are the regression coefficients for numeric grade, transparency score, and pollscore, respectively.
- γ_j indicates the fixed effect of each pollster.
- δ is the regression coefficient for the end date.
- ϵ_i is the error term.

3.2.2 Model 2 – Bayesian model for Trump

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (3)$$

$$\mu_i = \alpha + \beta_1 \cdot \text{numeric_grade}_i + \beta_2 \cdot \text{transparency_score}_i + \beta_3 \cdot \text{pollscore}_i \quad (4)$$

$$+ \sum_{j=1}^N \gamma_j \cdot \text{pollster}_j + \delta_i \cdot \text{end_date}_i \quad (5)$$

$$\alpha \sim \text{Normal}(50, 10) \quad (6)$$

$$\beta_1, \beta_2, \beta_3 \sim \text{Normal}(0, 5) \quad (7)$$

$$\gamma_j \sim \text{Normal}(0, 5) \quad (8)$$

$$\sigma \sim \text{Exponential}(1) \quad (9)$$

where:

- α represents an average poll result.
- β_1 , β_2 , and β_3 capture the unique effect of their respective predictors on the poll percentage.
- γ_j denotes a specific pollster effect.
- δ_i accounts for time trends.

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

3.3 Model justification

The choice of a Bayesian model to analyze Donald Trump’s voting support rate stems from the need to effectively integrate uncertainty and leverage prior knowledge in estimating outcomes. Given that the dataset includes key predictors related to polling organizations—such as numeric grade, transparency score, pollster, poll score, end date, and pct—this approach allows for a comprehensive understanding of how these variables influence voting results.

The numeric grade and transparency score are crucial in assessing the reliability of polling organizations. By integrating these factors, the Bayesian model can provide a nuanced analysis of how the quality and transparency of polling organizations impact Trump’s support rate. This is particularly important in the context of public opinion polling, as perceptions of reliability can significantly affect the interpretation of results.

Moreover, the Bayesian framework enables the incorporation of prior distributions for model parameters, reflecting beliefs about their possible values before observing the data. This is especially beneficial in a domain where historical data and expert opinion can provide valuable insights.

Additionally, by including the pollster variable, the model accounts for the fixed effects of different polling organizations, allowing for a more targeted analysis that acknowledges the unique characteristics of each organization. The `end_date` variable helps to consider time trends, ensuring that the analysis remains relevant to the evolving political landscape, especially as elections approach.

However, there are areas for improvement within the model. First, refining the selection of variables related to voting support could enhance the model’s predictive power. For instance, incorporating socio-economic factors, demographic data, or shifts in voter behavior may provide a more comprehensive perspective. Additionally, the model’s predictive performance could be assessed through cross-validation or other model evaluation techniques to ensure the robustness and reliability of the results.

By employing this Bayesian model, our aim is to provide robust estimates of Trump’s support rate, considering both the statistical characteristics of the data and the inherent uncertainty related to polling, while exploring avenues for further model improvement.

4 Results

The results part is combined with the result for models of analysis data and the result of predictions by Bayesian model on November 5, 2024. In order to get to know whether the votes of Trump will win the election, we combined the data of Harris who have the most competitive candidate other than Trump to do predictions. It would help us be faster and more accurate to do the predict judgment.

4.1 Result of model for analysis data

In Figure 9, we applied a generalized linear regression model to predict the percentage of polls Trump according each pollster. Specifically, the support rate data points range from 40% to 55%, with most clustered between 45% and 50%. The orange trend line shows a subtle upward trend, indicating a gradual increase in Trump’s average support rate over time. For example, in early August, the support rate is around 45%, while by mid-October, the average support rate shown by the trend line is close to 50%. This suggests that Trump’s support rate has risen by about 5% over these three months. For the model summary of Generalized Linear Regression and Bayesian Model, the table can see at Appendix A.4.1.

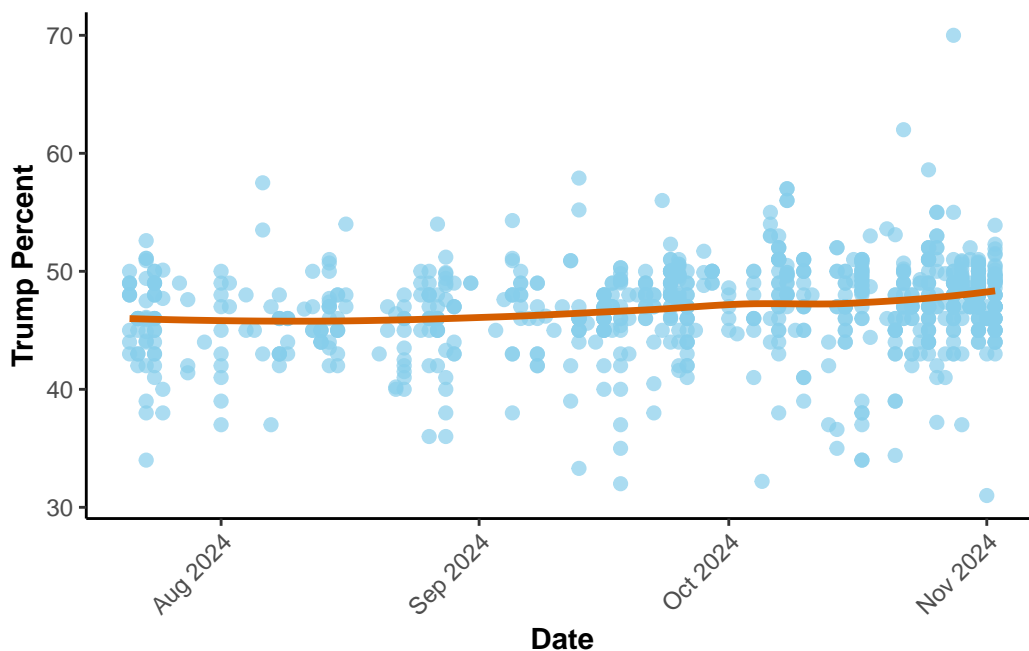


Figure 9: Donald Trump Support Over Time by Generalized Linear Regression

In Figure 10, we applied a bayesian model to predict the percentage of polls Trump according each pollster. Each data point is color-coded by pollster, with support rates ranging from 35% to 60%. The shaded gray area around the trend line represents the confidence interval, indicating the uncertainty in support rate variations. The overall trend line (blue) shows a slight upward trend, increasing from about 45% in August to nearly 50% in October. In particular, data points in early August are more dispersed, with some pollsters reporting low support rates (around 40%) and others reporting higher rates (over 50%). By November, most poll results have converged, with support rates centered between 45% and 50%. Notably, certain pollsters, like YouGov and Marist, provide either consistently higher or lower support rate estimates compared to others, indicating some bias or variation among polling organizations.

We also applied a Bayesian model to predict the percentage of polls Harris according each pollster. The figure can see it in Appendix A.4.2.

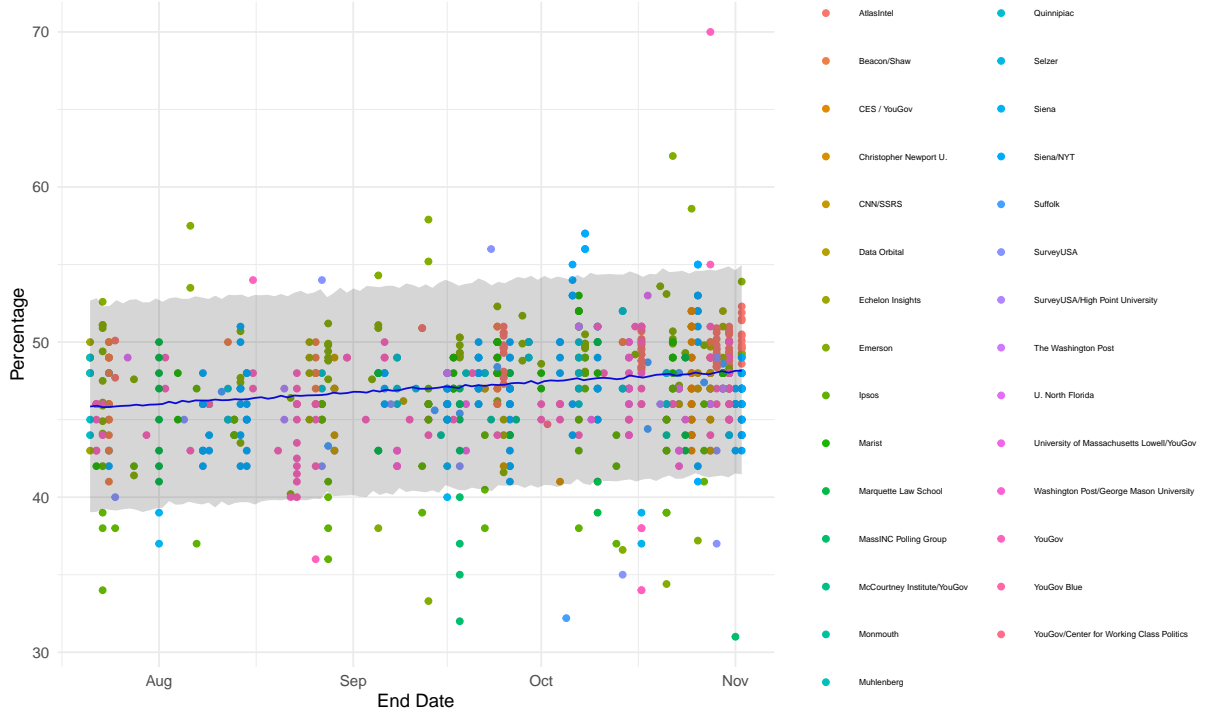


Figure 10: Poll Percentage over Time with Bayesian Fit for Trump

4.2 Result for model prediction

Figure 11 compares the predicted mean support rates for Trump and Harris across different polling organizations in 5th November of 2024. The vertical axis shows the predicted support rate (in percentage), and the horizontal axis lists the various pollsters. The red line represents Harris's predicted support rate, while the blue line represents Trump's.

For example, in predictions from Christopher Newport U. and MassINC Polling Group, Harris has a noticeably higher average support rate than Trump, exceeding 50%, while Trump's support rate is below 50%. In Siena's predictions, Trump's support rate is slightly lower than Harris's, reaching around 41%, while Harris's support is close to 53%. In Christopher Newport U.'s, MassINC Polling Group's, Siena's and Washington Post/George Mason University's predictions, Trump's support rate is all lower than Harris's and have a big gap.

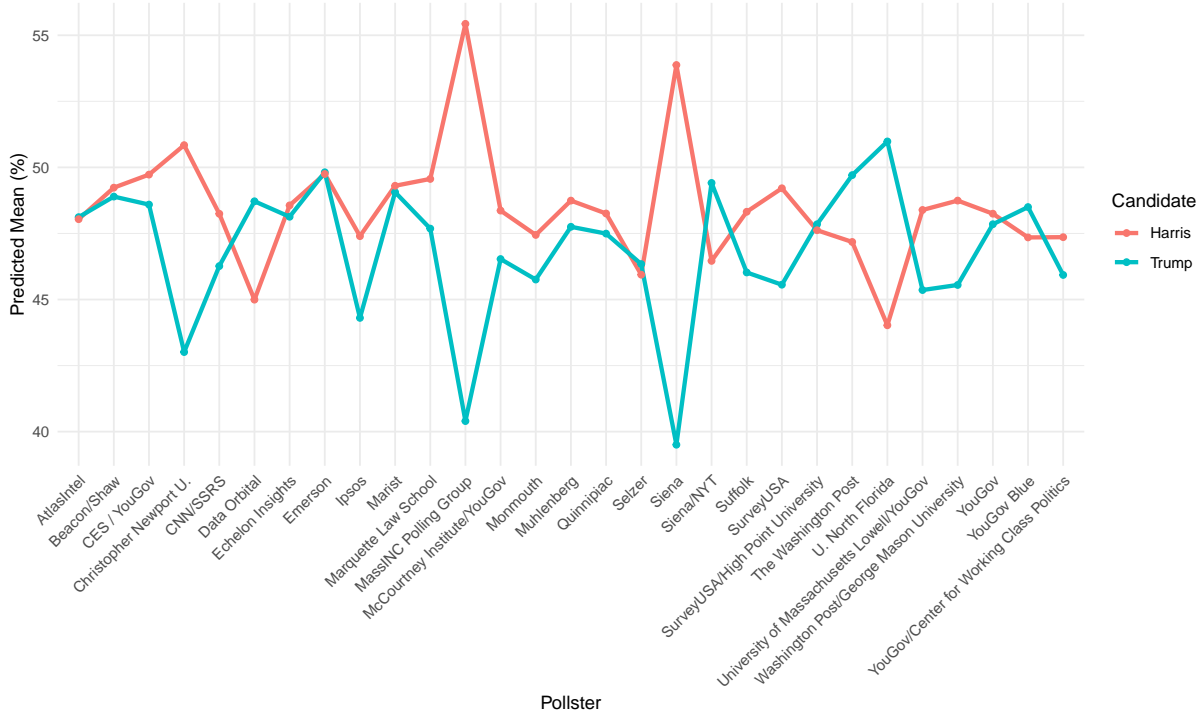


Figure 11: Predicted Mean Percentage by Pollster for Trump and Harris

In Table 3, we merged and compared the Bayesian model predictions of the vote shares for Trump and Harris on November 5, 2024, by pollster. The green background color represents the same pollster, indicating that the side with the larger predicted vote share is expected to win based on the data collected and summarized by that pollster. Conversely, red indicates the losing side. In Table 3, a total of 29 pollsters participated in the polling, with 9 pollsters predicting a Trump victory and 18 pollsters predicting a Harris victory. In Table 4, we calculated the mean support for Trump and Harris, estimating Trump’s probability of leading probability to be 31% and the mean support be 46.87%. For the further more predictions about the Bayesian model, we can see at Appendix A.4.3.

Table 3: Predictions for both Trump and Harris by pollster

Predictions Summary			
	pollster	pred_mean.Trump	pred_mean.Harris
2.5%	Emerson	49.81549	49.74995
2.5%1	AtlasIntel	48.11692	48.04041
2.5%2	Siena/NYT	49.41257	46.46172
2.5%3	MassINC Polling Group	40.40029	55.43243
2.5%4	Ipsos	44.30453	47.39779
2.5%5	YouGov	47.84888	48.24803
2.5%6	SurveyUSA	45.56023	49.20986

2.5%7	Echelon Insights	48.13166	48.56242
2.5%8	Selzer	46.33986	45.94006
2.5%9	Suffolk	46.02405	48.32055
2.5%10	Marist	49.06094	49.30425
2.5%11	Muhlenberg	47.75231	48.74332
2.5%12	The Washington Post	49.70756	47.17822
2.5%13	Data Orbital	48.71777	44.99492
2.5%14	Monmouth	45.75619	47.44451
2.5%15	Quinnipiac	47.49468	48.25494
2.5%16	Beacon/Shaw	48.89540	49.23308
2.5%17	CNN/SSRS	46.26501	48.24482
2.5%18	CES / YouGov	48.59251	49.72693
2.5%19	Marquette Law School	47.68378	49.56380
2.5%20	University of Massachusetts Lowell/YouGov	45.35861	48.38879
2.5%21	Washington Post/George Mason University	45.55097	48.74419
2.5%22	SurveyUSA/High Point University	47.84759	47.62245
2.5%23	U. North Florida	50.98170	44.02510
2.5%24	Siena	39.50460	53.87453
2.5%25	Christopher Newport U.	43.01220	50.84202
2.5%26	YouGov/Center for Working Class Politics	45.92983	47.35936
2.5%27	McCourtney Institute/YouGov	46.53130	48.36614
2.5%28	YouGov Blue	48.49834	47.34990

Table 4: Summary of Predictions by mean and lead probability

Metric	Value
Trump Mean Support	46.87
Harris Mean Support	48.50
Trump Lead Probability	0.31

5 Discussion

5.1 Summary of findings

The predictive models deployed in our study delineate a subtle yet pivotal dynamic: economic indicators and significant public events exert substantial influence on voter behavior as elections approach. The Bayesian analysis reveals a close support rate between Donald Trump (46.87%) and Kamala Harris (48.50%), highlighting an electorate acutely responsive to shifts in economic sentiment and public perception. This responsiveness is intensified by the fact

that nearly 31% of pollsters show varying support levels for Trump, underscoring a voter sentiment that is not merely divided but remarkably fluid. This data suggests that electoral outcomes can pivot rapidly following major economic updates or notable public engagements by candidates, potentially reshaping the electoral narrative almost instantaneously.

Furthermore, this analysis accentuates the imperative for campaign strategies to remain highly adaptable to these evolving conditions. The efficacy of Bayesian models, which we will talk about in the next section, in integrating and updating predictions with incoming economic data and changes in public sentiment is particularly invaluable. This capability enables campaigns to refine their strategic approaches almost in real-time, customizing their communications to exploit shifts in voter sentiment precipitated by economic improvements or declines. Such strategic agility is crucial; it equips campaign teams to seize opportunities to influence undecided or swing voters at decisive moments, effectively harnessing economic and social currents to secure electoral leverage.

5.2 Evaluating Predictive Models in Electoral Forecasting

In the task of forecasting the U.S. presidential election, the strategic selection of predictive models is paramount due to the complex interplay of electoral dynamics, such as shifts in voter behavior driven by economic changes, social movements, and regional voting patterns. These dynamics are characterized by non-linear relationships and interactions that are not only highly contextual but also vary significantly over time. Generalized Linear Models (GLMs) are valued for their straightforwardness and computational efficiency. However, they often simplify the intricate relationships within electoral data, which can result in overlooking subtle yet crucial nuances that are vital for accurate predictions. In contrast, Bayesian models excel in environments where prior knowledge and expert insights are available. These models integrate these elements through prior distributions that inform and refine the model's estimations, making them particularly advantageous for electoral forecasting. This integration allows for a nuanced understanding of voter intentions, providing the ability to adaptively update as new data becomes available—a critical feature in the rapidly changing landscape of political campaigns.

Additionally, Bayesian models produce probabilistic outputs—posterior distributions of parameters—that offer not only point estimates but also the probabilities of these estimates. This aspect enhances the interpretability and reliability of the forecasted electoral outcomes, making Bayesian models especially adept at handling the uncertainties inherent in polling data, such as non-response biases and sampling errors. Our project's application of Bayesian methodologies not only evaluates but also quantifies the effects of various predictive factors within a sophisticated probabilistic framework, ensuring a robust approach to capturing the complexities of national election predictions.

5.3 Quantifying Pollster Influence on Electoral Outcomes

The influence of pollster methodology on electoral forecasts is particularly critical for candidates like Donald Trump, whose support base may be heterogeneously distributed and exhibits high polarization. Our analysis of aggregated polling data from various sources demonstrates significant variations in predictions due to differing pollster methodologies. This variability highlights the crucial need for a rigorous assessment of how pollster-specific biases and methodological differences can skew the overall election forecasts. To systematically address this issue, an in-depth evaluation of each pollster’s historical accuracy and methodological rigor is required. Pollsters with a track record of high accuracy, especially in contexts demographically and politically akin to the current electoral environment, should be weighted more heavily in the models.

Furthermore, adopting a Bayesian hierarchical approach can refine these forecasts by pooling data from a variety of sources and adjusting for known biases. This model enhances the predictions by modulating the influence of each pollster based on their historical performance and methodological transparency, thereby improving the reliability and accuracy of electoral forecasts. Such a comprehensive approach not only mitigates the impact of pollster-specific biases but also provides a more detailed and contextually informed perspective on potential electoral outcomes.

5.4 Weaknesses and next steps

While our predictive models provide a comprehensive analysis of electoral dynamics, they are not without limitations. One notable challenge is the potential for overfitting in complex Bayesian models, which may adapt too closely to specific datasets and fail to generalize across different electoral cycles. Additionally, despite the sophisticated integration of economic indicators and polling data, the unpredictable nature of voter behavior—especially in response to sudden political or economic events—presents a significant challenge to the accuracy of our forecasts. The reliance on historical data to inform model priors can also lead to biases if past trends do not accurately reflect future outcomes.

To address these challenges, future research should focus on enhancing the generalizability of our models. This could involve the development of new statistical techniques that reduce the risk of overfitting and improve the models’ ability to adapt to new and varying datasets. Incorporating real-time data analytics into our Bayesian framework could also enhance the responsiveness of our models to sudden changes in the political landscape. Furthermore, expanding our data sources to include more granular demographic and psychographic information could refine our understanding of voter behavior and improve the accuracy of our predictions.

Additionally, conducting longitudinal studies to track changes in voter sentiment over multiple election cycles could provide deeper insights into the long-term efficacy of our predictive models and help identify persistent biases in pollster methodologies. Finally, increasing collaboration

with political scientists and behavior analysts could enrich the theoretical foundations of our models, ensuring that they not only predict electoral outcomes but also contribute to a broader understanding of electoral behavior.

A Appendix

A.1 Methodology Overview and Evaluation of YouGov Polling

Introduction

YouGov is an online polling company using a mix of non-probability sampling and a specialized model called multilevel regression with post-stratification (MRP) to estimate voter intentions. This appendix reviews YouGov’s methodology, covering sample recruitment, data collection, non-response management, questionnaire design, and MRP modeling. An assessment of the strengths and limitations of YouGov’s approach is also provided.

Population, Frame, and Sample Composition

The target population for YouGov’s election polling consists of American adults, particularly registered voters. YouGov uses an online panel framework, recruiting participants who voluntarily sign up and provide detailed demographic information. The sampling frame is further refined by aligning samples with demographic data from the TargetSmart voter file, which ensures that the sample accurately reflects national voter demographics. YouGov adjusts sample sizes during election cycles to improve representativeness and model robustness, starting with approximately 100,000 responses and increasing by an additional 20,000 during peak months, such as September and October.

Sample Recruitment and Representativeness

Panelists are recruited online through digital ads and partnerships with other websites, allowing any American adult with internet access to participate. Demographic data provided by participants enables YouGov to weight samples to reflect the population more closely. Respondents earn points redeemable for small rewards, which promotes engagement and data quality. However, reliance on online recruitment can underrepresent groups with limited internet access, such as some rural and low-income populations.

Sampling Approach and Methodological Trade-offs

YouGov’s sampling is non-probabilistic, meaning not everyone has an equal chance of selection. To mitigate this, YouGov applies MRP, a hierarchical modeling technique that segments respondents into subgroups based on characteristics like age, gender, race, education, and region. Subgroups are weighted to match their proportions in the broader population. While MRP efficiently models population-wide estimates, it may have trade-offs, especially in accurately representing smaller states and hard-to-reach demographic groups.

Non-response Handling and Quality Control Measures

YouGov employs quality control checks to address non-response bias. The platform incorporates response speed and consistency checks to filter out low-quality responses, and panelists failing these criteria are removed from future studies. These quality measures enhance the

MRP model’s accuracy by ensuring reliable data, critical for generating robust estimates of voter intentions.

Questionnaire Design

YouGov’s questionnaires are concise and randomized to minimize bias, with neutral wording and varied question types (text, images, audio) to enhance respondent understanding. For sensitive topics, options like “prefer not to say” are included to protect privacy. The online format, however, excludes certain demographics, and shorter survey lengths may limit the depth of data collected.

MRP Model for Vote Estimation

YouGov’s MRP model estimates voter support through three main steps: predicting voter turnout likelihood, determining likely voter candidate support, and aggregating to estimate overall support. The model draws on responses from the TargetSmart voter file, enabling estimates at both national and regional levels. MRP also allows YouGov to track voter opinion shifts, reflecting changing public sentiment over time.

Strengths and Limitations of the YouGov Approach

YouGov’s methodology has strengths, including the MRP model’s ability to generate precise regional and national predictions, and voter file use that enhances sample validity. Continuous panelist engagement allows YouGov to monitor shifts in voter support. However, potential limitations stem from the non-probability sampling approach, which may affect representativeness, especially in smaller states or underrepresented groups. Additionally, reliance on internet-based methods may miss some populations lacking consistent internet access.

Conclusion

YouGov employs a structured methodology combining MRP modeling, quality control measures, and a large online panel, producing valuable insights into U.S. voter intentions. While challenges in achieving full representativeness and internet reach persist, YouGov’s methodology has demonstrated success in past elections and serves as a reliable tool for future polling and tracking electoral trends.

A.2 Idealized Survey Methodology

The budget for this survey is \$100,000, aimed at predicting the outcome of the 2024 United States Presidential Election. This methodology includes stratified sampling, multi-platform recruitment, data validation, and multi-wave data aggregation to ensure representative and reliable data.

Sampling Approach: Stratified Random Sampling

To obtain a representative sample, we use stratified random sampling with a sample size of 5,000 respondents. Sampling is stratified by age, gender, education, and geographic region to ensure broad coverage across voter demographics. The age groups include 18-29, 30-44, 45-64, and 65 and above. Gender is categorized as male, female, and other, while education is classified as high school or below, bachelor's degree, and master's degree or above. The geographic region includes all U.S. states. Stratified sampling ensures that the sample accurately reflects voter characteristics, providing a solid foundation for subsequent analysis.

Recruitment: Multi-Platform and Interactive Engagement

Recruitment is conducted through a multi-platform strategy to ensure wide coverage among voters. Targeted ads are deployed on Google, Facebook, and Twitter to attract respondents with specific demographic characteristics. The ad content is concise and highlights anonymity and the research purpose to encourage participation. In addition, we use Random Digit Dialing (RDD) to contact older adults and residents in remote areas who may be less accessible online, ensuring their participation in the survey. To increase the completion rate, a small reward system is implemented, with one out of every 100 participants receiving a \$5 to \$10 incentive.

Survey Platform: Google Forms and Phone Outreach

Google Forms is used as the primary data collection platform, enabling respondents to complete the survey on a computer or mobile device with ease. We have set Google Forms to restrict submissions to one per Google account to prevent duplicate responses. For those who may not have easy online access, telephone outreach will be conducted, especially targeting older adults and rural voters, to ensure the diversity and inclusivity of the sample.

Data Validation: Post-Stratification Weighting

To ensure data quality, post-stratification weighting will be applied during the analysis phase based on demographic characteristics. This adjustment aligns the sample structure with the national voter distribution, reducing bias and increasing the accuracy of our findings. Multi-layered data validation measures enhance the reliability and representativeness of the results.

Poll Aggregation: Multi-Wave Polling and Adjustments

To track changes in voter sentiment over time, we will conduct multiple waves of data collection and aggregation. The survey will be administered in multiple rounds throughout the election cycle, with each round spaced 3-4 weeks apart. Each wave of data will be collected and

analyzed independently, then aggregated using weighted averages to smooth out single-instance fluctuations and help identify long-term trends, providing a reliable basis for predictions.

Budget Allocation: Phased Spending and Testing

To maximize budget efficiency, we will use a phased spending strategy. An initial 30% of the advertising budget will be allocated to testing across platforms to determine effectiveness, with the remaining 70% directed to the most successful channels. The budget breakdown is as follows: social media advertising (\$40,000) for broad and targeted outreach, phone outreach (\$20,000) for RDD calls to less accessible populations, small rewards (\$15,000) to encourage completion, and data cleaning and analysis (\$25,000) for validation, weighting, and aggregation across multiple waves.

Survey Content and Link

Survey Title: 2024 United States Presidential Election Survey

Survey Link: <https://forms.gle/EzyHp3zuX8Cu6Ep8A>

Introduction:

Thank you for taking part in our survey on the 2024 United States Presidential Election. Your input will help us understand voters' views on candidates and key policy issues. All responses are confidential and will be used solely for statistical purposes. As a token of appreciation, participants will have a chance to win a small reward, with one in every 100 respondents receiving \$5 to \$10.

For Questions or Concerns, Please Contact:

Tianrui Fu

Email: tianrui.fu@mail.utoronto.ca

Yiyue Deng

Email: yiyue.deng@mail.utoronto.ca

Jianing Li

Email: lijianing.li@mail.utoronto.ca

Survey Questions:

1. What is your age?

- ☐ 18-29
- ☐ 30-44
- ☐ 45-64
- ☐ 65 and above

2. What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other
- ☐ Prefer not to say

3. What is your highest level of education?

- ☐ High school or below
- ☐ Bachelor's degree
- ☐ Master's degree or above

4. Which state do you currently reside in?

(Dropdown menu with state names)

5. Do you plan to vote in the 2024 election?

- ☐ Yes
- ☐ No
- ☐ Not sure

6. If the election were held today, which candidate would you be more likely to support?

- ☐ Kamala Harris (Democratic Party)
- ☐ Donald Trump (Republican Party)
- ☐ Other
- ☐ Undecided

7. Which of the following issues would have the greatest impact on your decision in the 2024 election? (Select one)

- ☐ Economic stability and growth
- ☐ Access to quality healthcare
- ☐ Education reform and funding
- ☐ Environmental sustainability and climate action
- ☐ Addressing social and racial inequalities
- ☐ Other (please specify): _____

8. If a future candidate proposed a policy that aligns perfectly with your concerns, would you consider changing your voting intention?

☐ Yes

☐ No

☐ Not sure

9. If you would like to participate in the reward draw, please provide your email address below. Your contact information will only be used to notify winners. We will reach out to winners via email.

Email: _____

Thank You for Participating

We appreciate your time and input in our survey on the 2024 U.S. Presidential Election. Your responses are valuable to our research. Good luck with the reward draw, and have a great day!

A.3 Clean data

It starts by reading the dataset and standardizing column names. Next, it removes irrelevant columns like `sponsor_ids`, `sponsors`, and etc and filters for high-quality polls (with a rating of 2.7 or higher) that specifically track Trump's support. National polls missing state information are labeled as "National," and dates are formatted and filtered to include only those on or after July 21, 2024 (when Trump declared his candidacy). Finally, the percentage supporting Trump is converted to an actual count for further modeling. The same step also does with the Harris data set which will help us to do comparison later.

A.4 Results of model

A.4.1 Table for summary of model results

In Table 5, we conducted a summary of three models, including a generalized linear regression and a Bayesian model for Trump's data, as well as a Bayesian model for Harris's data. This summary includes the γ_j values for each pollster and monitoring data for the models, such as AIC, BIC, etc.

Table 5: Explanatory models of flight time based on wing width and wing length

[!h]	Model Summary		
	Model by glm	Model by bayes with Trump	Model by bayes with Harris
(Intercept)	-352.728 (86.299)	-390.051	-225.798
numeric_grade	-9.323 (12.348)	-3.038	0.533
transparency_score	0.679 (0.588)	-0.218	-0.025
pollscore	2.490 (12.352)	1.958	0.444
Beacon/Shaw	-1.368 (4.578)	0.780	1.149
pollsterCES / YouGov	-1.243 (4.632)	0.495	1.576
Christopher Newport U.	-9.906 (4.050)	-5.230	3.021
CNN/SSRS	-5.206 (2.432)	-1.836	0.270
Data Orbital	-1.945 (3.668)	0.433	-3.135
Echelon Insights	-2.996 (1.968)	-0.052	0.486
Emerson	1.849 (3.405)	1.562	1.544
Ipsos	-6.559 (2.651)	-3.830	-0.712
Marist	-0.185 (2.336)	0.924	1.138
Marquette Law School	-2.316 (4.657)	-0.493	1.448
MassINC Polling Group	-8.604 (1.489)	-7.745	7.321
McCourtney Institute/YouGov	-0.951 (4.771)	-1.681	0.266
pollsterMonmouth	-4.285 (2.951)	-2.453	-0.635
Muhlenberg	-1.848 (3.706)	-0.451	0.641
Quinnipiac	-3.155 (3.482)	-0.698	0.107
Selzer	-2.937 (5.524)	-1.824	-2.187
Siena	-10.494 (1.518)	-8.715	5.756
Siena/NYT	0.539 (8.401)	1.199	-1.647
Suffolk	-2.876 (2.977)	-2.133	0.209
SurveyUSA	-3.914 (6.513)	-2.714	1.190
SurveyUSA/High Point University	-3.331 (7.661)	-0.351	-0.387
The Washington Post	0.725 (5.228)	1.523	-0.871
University of Massachusetts Lowell/YouGov	-4.507 (3.410)	-2.828	0.228
Washington Post/George Mason University	-5.634 (1.950)	-2.650	0.754
YouGov	-1.250 (4.219)	-0.349	0.190
YouGov Blue	-0.606 (5.347)	0.258	-0.665
End Date	0.021 (0.004)	0.023	0.014
U. North Florida		2.760	-4.044
YouGov/Center for Working Class Politics		-2.217	-0.571
Num.Obs.	803	803	782
R2	0.287	0.280	0.201
R2 Adj.		0.240	0.146
AIC	4213.4		
BIC	4363.4		
Log.Lik.	-2074.682	-2080.530	-2025.723
ELPD		-2106.5	-2050.7
ELPD s.e.		40.8	41.8
LOOIC		4213.0	4101.4
LOOIC s.e.		81.6	83.5
WAIC		4211.5	4099.6
RMSE	3.21	3.52	3.22

^a This table shows the regression models with custom variable names.

A.4.2 Bayesian model for Harris data set

In Figure 12, we applied a bayesian model to predict the percentage of polls Harris according each pollster. Each data point is color-coded by pollster, with support rates ranging from 35% to 65%. The shaded gray area around the trend line represents the confidence interval, indicating the uncertainty in support rate variations. The overall trend line (blue) shows a slight upward trend, increasing from about 46% in August to nearly 48% in October. Compared to Trump's model, the increasing trend of Harris is lower than Trump's. By October, most poll results have converged, with support rates centered between 45% and 50%.



Figure 12: Poll Percentage over Time with Bayesian Fit for Harris

A.4.3 Predictions for both Trump and Harris

In Table 6, we summarized the Bayesian model predictions for Trump and Harris on November 5, 2024, based on 29 pollsters, including the mean, lower bound, and upper bound.

Table 6: predictions for trump and harris by pollster

Predictions Summary					
	Pollster	Candidate	Predicted Mean	Lower Bound	Upper Bound
2.5%	Emerson	Trump	49.74807	43.28219	56.35324
2.5%1	AtlasIntel	Trump	48.17328	41.60123	54.64031
2.5%2	Siena/NYT	Trump	49.42633	42.38901	56.25740
2.5%3	MassINC Polling Group	Trump	40.43736	33.54492	47.35021
2.5%4	Ipsos	Trump	44.35723	37.80983	51.06026
2.5%5	YouGov	Trump	47.88230	41.24154	54.41102
2.5%6	SurveyUSA	Trump	45.38837	38.59699	52.14075
2.5%7	Echelon Insights	Trump	48.12700	41.49321	55.09593
2.5%8	Selzer	Trump	46.31983	38.51049	53.96826
2.5%9	Suffolk	Trump	46.01116	39.33450	52.61313
2.5%10	Marist	Trump	49.15435	42.84796	55.67257
2.5%11	Muhlenberg	Trump	47.67240	40.26288	55.23404
2.5%12	The Washington Post	Trump	49.76048	43.08948	56.59855
2.5%13	Data Orbital	Trump	48.70637	41.12452	56.35085
2.5%14	Monmouth	Trump	45.76013	38.79269	53.12481
2.5%15	Quinnipiac	Trump	47.63194	40.11602	55.05057
2.5%16	Beacon/Shaw	Trump	49.05579	42.57426	55.50880
2.5%17	CNN/SSRS	Trump	46.30395	39.08137	53.75664
2.5%18	CES / YouGov	Trump	48.66837	41.86771	55.45890
2.5%19	Marquette Law School	Trump	47.75041	41.35784	54.47066
2.5%20	University of Massachusetts Lowell/YouGov	Trump	45.37743	38.74817	51.94390
2.5%21	Washington Post/George Mason University	Trump	45.50342	38.62663	52.25494

2.5%22	SurveyUSA/High Point University	Trump	47.87389	40.44454	55.27375
2.5%23	U. North Florida	Trump	50.92362	42.50054	59.16042
2.5%24	Siena	Trump	39.40856	32.12704	46.55104
2.5%25	Christopher Newport U.	Trump	42.89052	34.43363	51.61160
2.5%26	YouGov/Center for Working Class Politics	Trump	46.03093	37.41851	55.04925
2.5%27	McCourtney Institute/YouGov	Trump	46.45987	37.98113	54.93659
2.5%28	YouGov Blue	Trump	48.56749	40.20211	57.01238
2.5%29	Emerson	Harris	49.63418	42.87829	56.14878
2.5%30	AtlasIntel	Harris	48.00457	41.41379	54.68073
2.5%31	Siena/NYT	Harris	46.38911	39.39285	53.52612
2.5%32	MassINC Polling Group	Harris	55.34354	48.59579	62.24625
2.5%33	Ipsos	Harris	47.35787	40.63884	54.03892
2.5%34	YouGov	Harris	48.31199	41.74463	54.82360
2.5%35	SurveyUSA	Harris	49.29559	42.52625	56.07497
2.5%36	Echelon Insights	Harris	48.57784	41.82573	55.68554
2.5%37	Selzer	Harris	45.91051	38.09348	53.55376
2.5%38	Suffolk	Harris	48.22747	41.38776	55.15722
2.5%39	Marist	Harris	49.18951	42.66968	55.59954
2.5%40	Muhlenberg	Harris	48.76800	40.89858	56.70053
2.5%41	The Washington Post	Harris	47.27206	40.58815	54.17179
2.5%42	Data Orbital	Harris	44.88530	37.20791	52.65914
2.5%43	Monmouth	Harris	47.47059	40.31200	54.69014
2.5%44	Quinnipiac	Harris	48.12331	40.51441	55.35946
2.5%45	Beacon/Shaw	Harris	49.22415	42.71583	55.75335
2.5%46	CNN/SSRS	Harris	48.42867	41.17889	55.81444
2.5%47	CES / YouGov	Harris	49.64200	43.16074	56.14841
2.5%48	Marquette Law School	Harris	49.51876	43.05076	56.09031
2.5%49	University of Massachusetts Lowell/YouGov	Harris	48.31432	41.49899	55.01765

2.5%50	Washington Post/George Mason University	Harris	48.79504	41.93451	55.37136
2.5%51	SurveyUSA/High Point University	Harris	47.61006	40.40853	55.18262
2.5%52	U. North Florida	Harris	44.08972	35.93581	52.16615
2.5%53	Siena	Harris	53.71982	46.50436	60.95860
2.5%54	Christopher Newport U.	Harris	50.93937	41.98553	59.89053
2.5%55	YouGov/Center for Working Class Politics	Harris	47.46588	38.68029	55.91440
2.5%56	McCourtney Institute/YouGov	Harris	48.32295	39.86215	56.84292
2.5%57	YouGov Blue	Harris	47.40733	39.10670	55.71866

B Model details

To maintain readability while demonstrating model robustness, we include the following in the appendix:

- **Prior Justification and Sensitivity Analyses:** Alternative priors and their justifications are provided, alongside a sensitivity analysis to examine the impact of these priors on the posterior distributions.
- **Model Validation and Out-of-Sample Testing:** Validation metrics, such as RMSE calculations, out-of-sample testing, and test-train splits, offer evidence of the model’s predictive accuracy, complementing in-sample performance.
- **Alternative Models and Comparison:** An analysis of simpler and more complex models is included, explaining the rationale for selecting this Bayesian model based on performance metrics and interpretability.

References

- Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r*. Chapman; Hall/CRC.
- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- FiveThirtyEight. 2024. “2024 National Presidential Poll Results.” <https://projects.fivethirtyeight.com/polls/president-general/2024/national/>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2023. *Modelr: Modelling Functions That Work with the Pipe*. <https://CRAN.R-project.org/package=modelr>.
- 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>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.