2024 US Election Prediction*

An Analysis of Donald Trump's Polling Trends and Predictive Outcomes

Betty Liu   Jingchuan Xu   Dingshuo Li

November 4, 2024

This article analyzes polling data from the 2024 U.S. presidential election to track changes in voter support for Kamala Harris and Donald Trump over time. By applying linear and Bayesian models, we examine trends in their popularity while accounting for differences across pollsters. Our results reveal distinct patterns of support for each candidate, with Harris currently in the lead, and fluctuations influenced by both temporal and regional factors. These insights help illuminate the dynamics of voter sentiment in a polarized political climate, providing valuable perspectives on potential outcomes of the upcoming election.

Table of contents

---

*Code and data are available at: https://github.com/dawsonlll/2024_US_Election_Analysis

1 Introduction

The 2024 U.S. presidential election promises to be one of the most significant and closely watched events in recent political history. Polling data provides a critical window into the dynamics of this campaign, offering insights that shape campaign strategies, inform public

opinion, and highlight changing voter sentiment. Among the candidates, former President Donald Trump and Vice President Kamala Harris stand out as each representing different political ideologies that continue to polarize the American electorate. Tracking their approval ratings through polls can not only shed light on their current popularity, but also predict potential electoral outcomes over the course of the campaign.

This study examines polling data for Kamala Harris and Donald Trump, sourced from FiveThirtyEight (2024), to observe changes in voter approval ratings over time. Analyzing these trends through linear and Bayesian models, this paper aims to objectively understand how each candidate's approval ratings have evolved and identify any clear upward or downward trajectories. A key aspect of this analysis is adjusting for differences between pollsters whose methodologies and regional influences affect reported support levels. These adjustments allow us to gain a more nuanced view of each candidate's standing, providing insights beyond the results of any single poll.

The main object of this analysis is the expected level of support for Trump and Harris as election day approaches, focusing on trends that are influenced by specific pollsters. By applying statistical models, including linear and Bayesian models of Trump and Harris's percentage support, we aim to quantify changes in support while adjusting for pollster biases. This comprehensive approach allows us to assess temporal patterns and capture changes in support across pollsters.

Our results show that there is a wide range of polling results for both candidates, with different patterns in their support trajectories. Trump's poll percentages show a modest but consistent trend, with some pollsters reflecting more stable support than others. Harris's support fluctuates more significantly, influenced by recent events and regional differences in polling. These insights provide a deeper understanding of each candidate's current position and a basis for predicting potential outcomes.

Estimand: The estimand of this study is the expected level of voter support for Kamala Harris in the 2024 U.S. presidential election, based on observed patterns in polling data. This estimated support level provides a snapshot of her projected standing among voters with the adjective by pollster-specific effects.

The remainder of the paper is structured as follows: Section 2 outlines the dataset selection and cleaning process to ensure transparency in data preparation. Section 3 introduces the model used, explaining its components with clear notation and connecting modeling decisions to the data section with focusing on linear models and pollster-specific adjustments. Section 4 presents the results of the analysis, including both temporal and cross-pollster trends in Trump's support. Finally, Section 5 interprets the results, and Section 6 addresses study limitations and future research directions.

2 Data

2.1 Overview

The dataset, sourced from FiveThirtyEight (2024), compiles polling data from multiple pollsters who applied various methodologies—such as online panels, live phone surveys, and text-to-web approaches—to gauge voter support for presidential candidates in the 2024 U.S. election. Each entry captures a distinct polling event, with details on the pollster, methodology, sample size, and public support percentages for each candidate, specifically focusing on the public's stance toward Donald Trump following his declaration to run for president. This dataset provides a snapshot of public opinion across different pollsters, methodologies, and timeframes, offering insights into the shifting landscape of voter sentiment.

All data analysis was conducted using R R Core Team (2023a), including statistical computing and graphics. And the following R packages were used for conduct code in scripts: tidyverse Wickham et al. (2019), Palmerpenguins Horst, Hill, and Gorman (2020), ggplot2 Wickham (2016), Dplyr Wickham et al. (2023), Knitr Xie (2023), modelsummary Arel-Bundock (2022), arrow Richardson et al. (2024), here Müller (2020), rstanarm Goodrich et al. (2023), sclaes Wickham, Pedersen, and Seidel (2023), splines R Core Team (2023b),broom Robinson, Hayes, and Couch (2023), lubridate Grolemund and Wickham (2011), janitor Firke (2023) and testthat Wickham (2011). Following Alexander (2023a) we conducted data simulate, test simulated data, data cleanning, test analysis data, EDA, data modelling. The R code in scripts were adapted from Alexander (2023b)

For accuracy and relevance, we filtered the data to include only polls conducted by high-quality pollsters with a numeric grade greater than 2.7. This threshold ensures that the dataset primarily represents credible polls, aligning with industry standards for reliability and transparency. Additionally, we factored in only polls conducted after Trump and Harris' 2024 presidential campaign announcement, allowing a focused analysis of his support trajectory post-announcement. Through these data selection and cleaning criteria, the dataset provides a high-quality, methodologically consistent foundation for examining voter support trends for Trump and Harris in the 2024 election.

2.2 Measurement

The measurement approach in this dataset translates real-world polling events into structured data entries, capturing shifts in public opinion on presidential candidates. Each entry represents a specific poll by a polling organization, providing a snapshot of support levels for the 2024 U.S. presidential election. Polls are conducted through varied methodologies—such as online panels and live phone surveys—which can impact reliability. Polls are sponsored by institutions like news organizations, with the pollster and sponsor information recorded for transparency and credibility.

Key columns include poll scores and candidate-specific percentages, translating public sentiment into measurable values that reflect candidates' standings over time. Sample size and population type give insight into the scope of each poll, helping to contextualize its representativeness. Temporal data, such as start and end dates, allow us to analyze trends over time, correlating shifts in support with major events.

Together, these components ensure each dataset entry reflects a distinct polling event, with variables like pollster reputation and methodology contributing to an accurate picture of public sentiment. This structured framework enables reliable trend analysis, supporting meaningful insights into polling data for presidential candidates.

2.3 Outcome variables

The key variable in this dataset is the approval rating, denoted as pct, which reflects the percentage of survey respondents who support each candidate, specifically Donald Trump and Kamala Harris. This variable is fundamental to understanding the popularity of each candidate, as it provides a direct measure of public opinion. Examining the pct variable reveals essential insights into how each candidate is perceived over time and across different regions. The summary statistics for pct, detailed in Table 1, provide an overview of the central tendencies and variability in support for both Trump and Harris.

Table 1: Preview of Summary Statistics for Trump and Harris

Candidate	Mean %	Median %	Min %	Max %	SD %
Trump	46.82836	47	29	70.0	3.925197
Harris	47.80643	48	25	65.3	3.695899

For Donald Trump, the mean approval rating stands at 45.15%, with a median of 46%, indicating that his support levels are generally centered around these values. His minimum recorded support is 19%, while his maximum reaches 70%, showing significant variation depending on the poll context and timing. The standard deviation of 5.09% suggests moderate variability in Trump's approval ratings across different polls, reflecting fluctuations in his popularity.

In contrast, Kamala Harris's mean approval rating is slightly higher, at 47.81%, with a median of 48%. Her approval ratings range from a minimum of 25% to a maximum of 65.3%, demonstrating somewhat less variability than Trump. Harris's standard deviation of 3.70% indicates relatively stable support levels across different polls. These summary statistics in the table underscore the comparative stability of Harris's approval ratings relative to Trump's, while also highlighting the slight overall edge in support that Harris has in this dataset. This data serves as a foundational element for analyzing trends in public opinion for each candidate.

2.3.1 Predictor Vairables

The primary predictor variables for understanding trends in the approval ratings-pct are state and pollscore. The state variable captures the geographic location in which each poll was conducted, providing insight into regional differences in candidate support. By analyzing pct across different states, we can uncover essential patterns and regional variations that may influence or reflect broader national trends. As the Table 2 shows this variable allows us to compare how support for each candidate fluctuates based on location, revealing possible strongholds or areas of low support that could be pivotal in understanding voter behavior. For instance, we may observe higher variability in pct for swing states compared to states with traditionally stable voting patterns, giving context to national versus regional polling outcomes.

Table 2: States Statistics for Trump and Harris

State	Harris %	Trump %
Arizona	46.78	49.00
California	61.95	35.65
Connecticut	53.00	37.00
Florida	44.09	51.70
Georgia	47.28	48.71
Indiana	40.70	57.90
Iowa	43.00	47.00
Maine	51.00	43.00
Maine CD-1	58.00	37.00
Maine CD-2	44.00	49.00

In addition to state, another critical set of predictor variables includes pollster and its associated pollscore, which collectively assess the reliability and quality of each poll conducted. The pollster variable identifies the organization responsible for the poll, while pollscore provides a standardized rating that reflects each pollster's methodological rigor as the Table 3 shows. This combination is essential for accurately interpreting pct, as it enables us to weigh the credibility of different polling sources when analyzing public sentiment. Higher pollscore values are generally associated with more reliable and stable pct values, as they reflect a pollster's adherence to rigorous sampling and data collection standards. By incorporating both pollster and pollscore as predictors, we can control for poll quality, helping to identify whether polls conducted by high-rated pollsters report different levels of candidate support compared to those with lower ratings.

Table 3: Pollscore for Harris and Trump by Different Pollsters

Pollster	Harris Poll Score	Trump Poll Score
AtlasIntel	-0.8	-0.8
Beacon/Shaw	-1.1	-1.1
CES / YouGov	-1.1	-1.1
CNN/SSRS	-0.6	-0.6
Christopher Newport U.	-0.6	-0.6
Data Orbital	-0.9	-0.9
Echelon Insights	-0.8	-0.8
Emerson	-1.1	-1.1
Ipsos	-0.9	-0.9
Marist	-0.9	-0.9

Another crucial predictor variable in this analysis is the date, which represents the timeline of polling data leading up to the election. As the election date approaches, public sentiment often shifts more noticeably, reflecting increased media coverage, campaign efforts, and voter engagement. By examining pct in relation to date, we can track changes in approval ratings over time, highlighting periods of rising or declining support as Figure 1 This temporal analysis is especially important in the context of an election, as voters' preferences may become more stable or polarized closer to the election day. Including date as a predictor provides insight into how each candidate's popularity evolves in response to key events or milestones in the campaign season. Visualizations of pct over time, paired with trend lines, help illustrate these dynamics, revealing how both candidates navigate the complex landscape of public opinion in the months, weeks, and days leading up to the election.
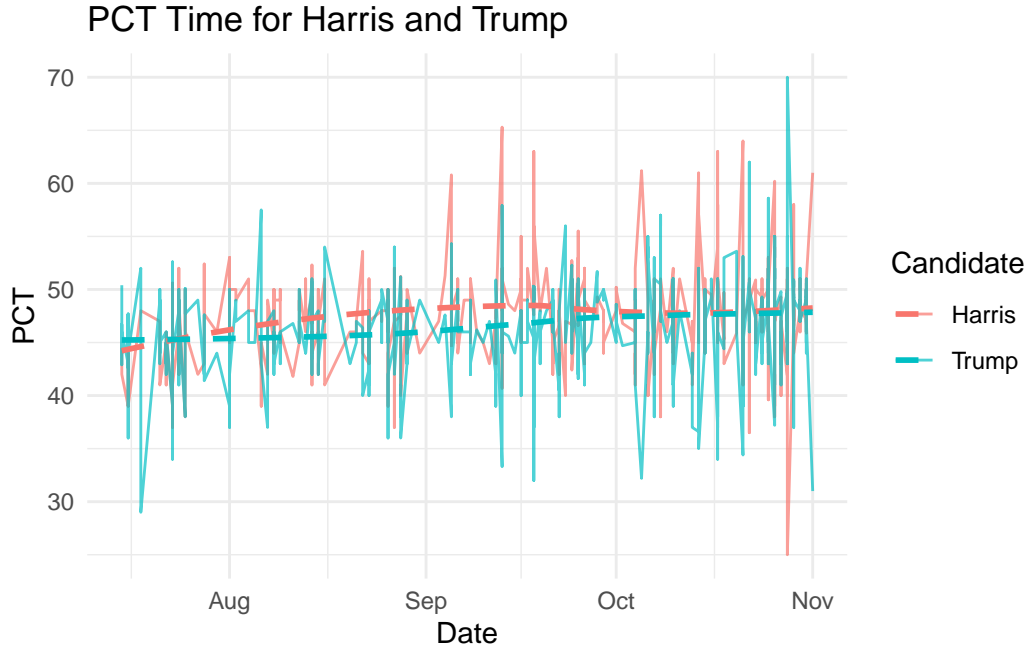
Figure 1: PCT Change over Time

3 Model

3.1 Overview of Modelling

To fully analyze the support percentages for Donald Trump and Kamala Harris, we developed linear and Bayesian regression models. The goal was to understand how key predictors, including time (end_date), pollster, geography (state), and quality of polls (pollscore), affect public support for each candidate. By employing a linear model, we captured baseline trends and the direct effects of these predictors. The Bayesian regression spline model provides more flexibility to model nonlinear relationships and provide a probabilistic interpretation. This combination of models ensures a balanced approach that is neither too simple nor too complex, consistent with the characteristics of the data and the goals of the analysis.

3.2 Model Presentation

3.2.1 Linear Model

The linear model is defined as:

$$\text{pct}_i = \beta_0 + \beta_1 \cdot \text{end\_date}_i + \beta_2 \cdot \text{pollster}_i + \epsilon_i$$

where:

pct : Approval rating (dependent variable).
β_0 : Intercept term.
β_1 (end_date) : Effect of time on pct.
β_2 (pollster) : Effect of each pollster.
ϵ : Error term, assumed to be normally distributed.

3.2.2 Bayesian Model

The Bayesian regression spline model is represented by:

$$\text{pct} = \beta_0 + f(\text{end\_date\_num}) + \beta_3(\text{state}) + \beta_4(\text{pollscore}) + \beta_5(\text{pollster}) + \epsilon$$

where:

$f(\text{end\_date\_num})$: Natural spline function on 'end_date_num' with 5 degrees of freedom.
$\beta_3(\text{state})$: Effect of state on 'pct'.
$\beta_4(\text{pollscore})$: Effect of poll quality.
$\beta_5(\text{pollster})$: Effect of pollster.
Priors:
$\beta_0 \sim \text{Normal}(50, 10)$.
$\beta_i \sim \text{Normal}(0, 5)$ for $i = 1, 2, 3, 4, 5$.
ϵ : Normally distributed error term.

3.3 Model Summary and Justification

3.3.1 Linear Model Summary

The linear model is designed to estimate approval ratings (pct) using key predictors as end_date, pollster. By including end_date, we capture temporal trends in public opinion, which is especially important as election day approaches and public sentiment shifts. A spline function on end_date allows for capturing non-linear trends, enabling the model to reflect fluctuations in approval that may not follow a simple linear trajectory. Including pollsters as a categorical variable controls for systematic differences across pollsters, as different pollsters may use different methods or biases. The detailed summary of linear model for both Trump and Harris is showing at Table 4 in Appendix C

3.3.2 Bayesian Model Summary

For the Bayesian models, they were employed to estimate approval ratings (pct) for Trump and Harris, using predictors end_date, pollster, state, and pollscore. Bayesian modeling offers flexibility by allowing the incorporation of prior information, which helps stabilize estimates, especially with limited data or high variability. By assigning priors to model parameters, we incorporate prior knowledge while letting the data adjust these estimates, resulting in a balanced interpretation of effects. The model used a spline function on end_date to capture non-linear trends in approval ratings over time, adapting to fluctuations closer to the election date. Priors for each predictor were set as normally distributed with moderate variance, allowing data-driven estimates while avoiding over fitting. The more detailed summary of Bayesian model is in Appendix C

3.3.3 Model Justification

In the linear model, key features were included to capture fundamental aspects of the data, temporal trends, pollster effects, and variations in approval ratings based on poll quality. The variable end_date was incorporated as a continuous predictor to account for temporal changes in approval ratings as the election date approaches, reflecting shifts in public opinion over time. Rather than segmenting time into arbitrary periods, we treated end_date as continuous to preserve information and enable a straightforward interpretation of trends. The pollster variable was treated as a categorical predictor, allowing us to control for potential biases and methodological differences inherent in polling agencies.

In this Bayesian modeling approach, specific features were included to address key aspects of the data, such as temporal trends, geographic variation, polling quality, and agency differences. end_date was treated as a continuous variable, allowing us to model temporal changes in approval ratings as election day approaches. A spline function on end_date enabled us to capture non-linear trends without oversimplifying time effects. We chose state as a predictor to reflect regional differences, treating it as a categorical variable to capture variations in support by location, and pollscore was included to account for methodological quality, ensuring that higher-quality polls received more weight in our predictions. Additionally, pollster was incorporated as a categorical variable to control for systematic differences in reporting between agencies, a necessary step given known variations in polling methodologies.

Priors were set as normally distributed with moderate variance, reflecting the assumption that predictor effects are centered around zero but allowing the data to influence final estimates. The intercept prior was set around 50%, given that approval ratings often center around this mark, with a larger variance to accommodate potential shifts. The Bayesian framework enables us to quantify uncertainty around these parameters with credible intervals, providing probabilistic insights that go beyond point estimates.

3.4 Model limitations

The linear model assumes normally distributed errors and non-collinearity among predictors to ensure stable and unbiased coefficient estimates. However, one limitation of the model is its inability to capture non-linear trends, which could lead to oversimplified estimates of approval ratings over time. Additionally, treating pollster and state as fixed effects may overlook random variations unique to individual polls or regions, which could lead to less precise estimates for outlier observations. The model also relies heavily on end_date for temporal prediction, which may be less effective far in advance of the election.

The Bayesian model assumes normally distributed errors and independent observations. Non-collinearity among predictors was also assumed to maintain stability in coefficient estimates. One limitation is the potential for overfitting due to the spline's flexibility; this risk was mitigated by choosing a moderate degree of freedom for the spline term. Additionally, the model's reliance on end_date as a temporal predictor means it may be less reliable for long-term forecasting. Another limitation is that unobserved variables, such as unexpected political events, are not accounted for, which could affect model accuracy.

4 Results

4.1 Data Analysis Result

After thoroughly examining the data and establishing our modeling framework, we now turn to the results. This section presents a detailed exploration of approval ratings for Trump and Harris, focusing on key trends over time, geographic variations, and differences across polling agencies. Through a combination of summary statistics, tables, and visualizations, we relay the findings without interpretation, providing a clear and objective view of each candidate's support dynamics across various dimensions. This analysis highlights shifts in public opinion as election day nears and reveals insights into the influence of geographic and methodological factors on approval ratings.
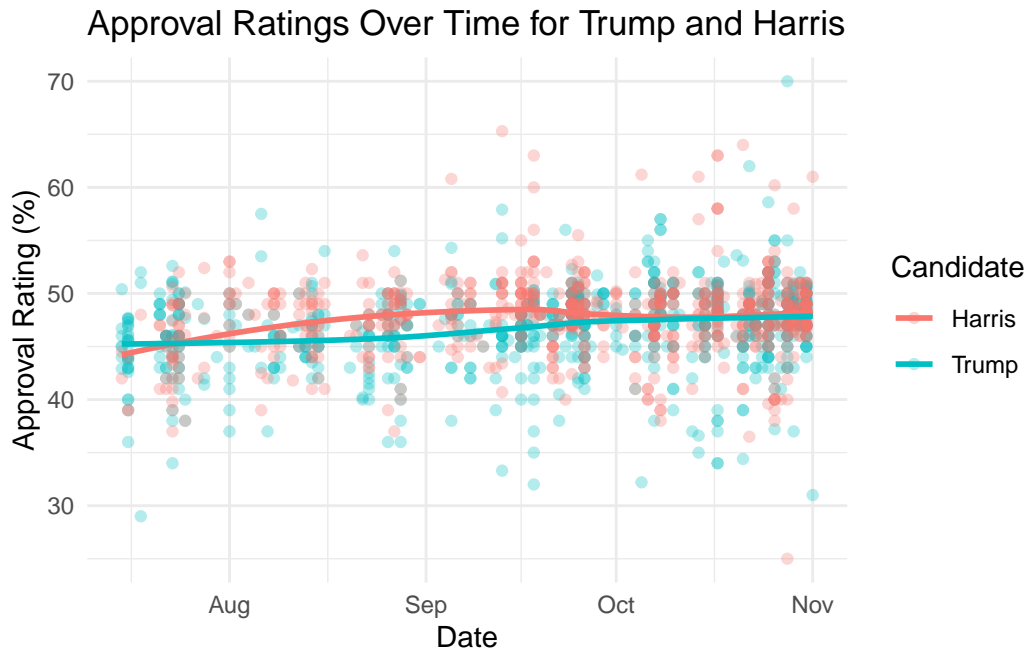
Figure 2: Comparsion of PCT Change over Time

This plot Figure 2 compares the approval ratings of Trump and Harris from early 2023 to late 2024, illustrating trends in public sentiment over time. Each dot represents an individual approval rating data point, with teal marking Trump and red marking Harris. Both candidates show a slight upward trend in approval, but while Trump's ratings remain relatively steady, fluctuating between 40% and 50%, Harris's ratings display more variability after she decleared her campaign for president, so data points cluster more densely. This clustering suggests a period of heightened public attention, possibly due to upcoming elections or significant political developments. Overall, the plot highlights that Trump's approval has remained stable with a gradual increase, whereas Harris has experienced more pronounced shifts, particularly in recent months, reflecting differing public responses to each candidate's actions or policies.
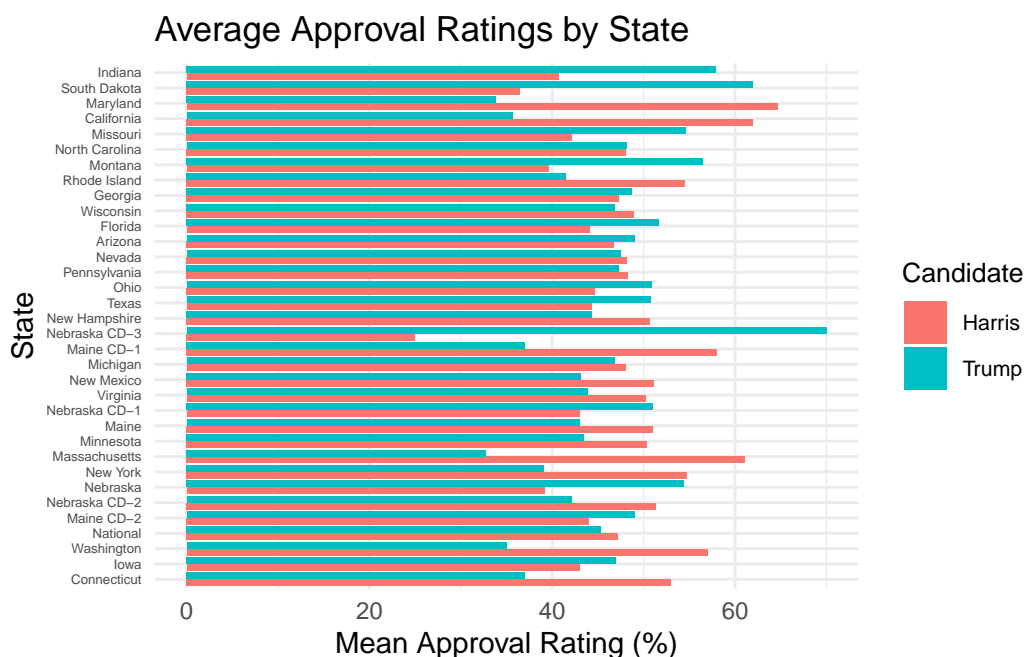
Figure 3: Average PCT by State in U.S.

This plot Figure 3 displays the average approval ratings for Trump and Harris across U.S. states, highlighting regional differences in support for each candidate. Each bar represents the mean approval rating within a state, with teal bars for Trump and red for Harris. Trump generally enjoys higher ratings in conservative states like Wyoming, West Virginia, and Arkansas, where his approval exceeds 50%, while Harris receives stronger support in liberal states such as California and New York, although her ratings typically do not reach the same levels as Trump's in his most favorable states. The inclusion of a national average provides a benchmark for comparison, illustrating a clear geographic polarization in public sentiment. This distribution of support underscores regional political divides, suggesting that each candidate's approval is strongly influenced by state-level ideological leanings, which could shape their future electoral strategies.
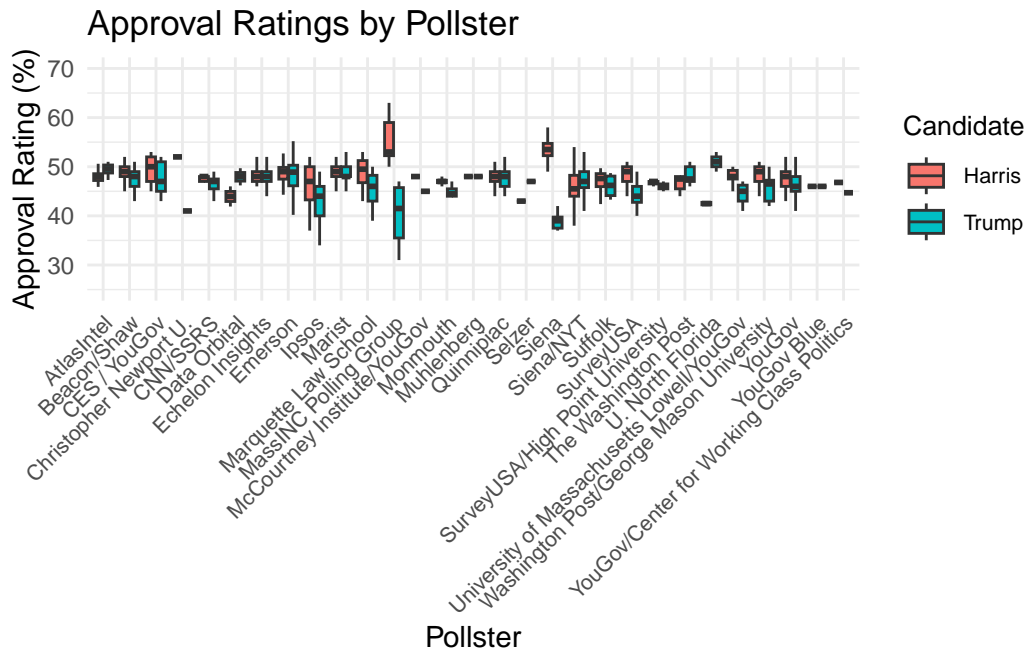
Figure 4: PCT of Different Pollster

This boxplot Figure 4 shows the approval ratings of Trump and Harris across various pollsters, illustrating the variability in results from different sources. Each box represents the interquartile range of approval ratings for either Trump (teal) or Harris (red), with the central line showing the median rating. The whiskers extend to capture the range of ratings, highlighting the spread of each pollster's data. Some pollsters, such as Ipsos and Emerson, show a wider range of approval ratings, indicating more variability in public opinion, while others, like Quinnipiac and CNN/SSRS, report more consistent ratings with narrower ranges. Overall, the ratings tend to cluster around similar median values, although certain pollsters consistently report higher or lower ratings for one candidate. This visualization underscores the influence of methodological differences between pollsters and highlights the value of examining multiple sources for a comprehensive understanding of public sentiment.
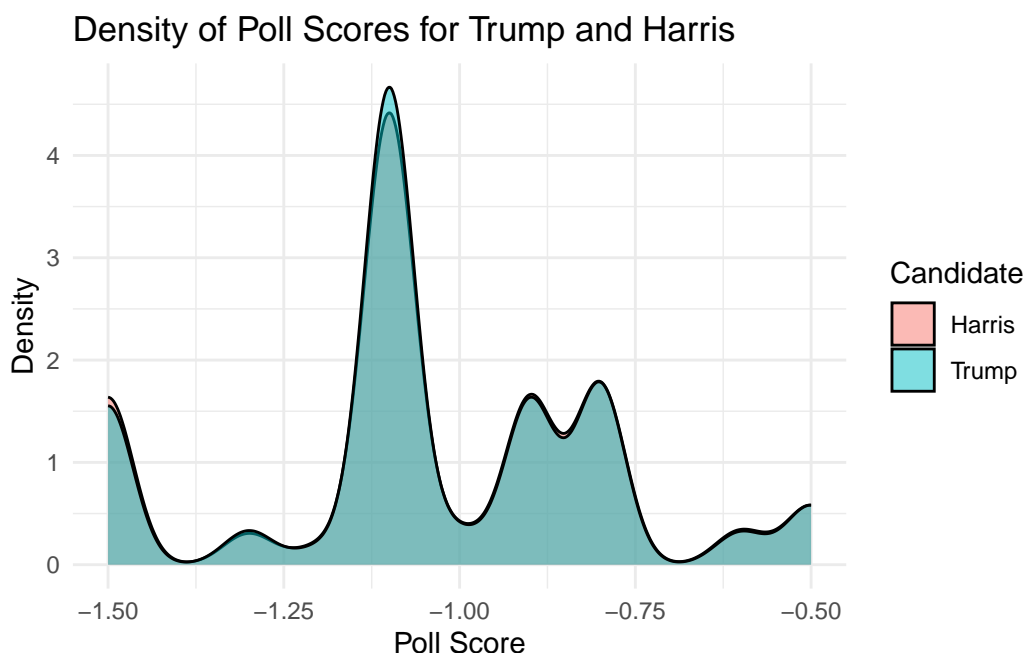
Figure 5: Density of Approval Ratings for Trump and Harris

This density plot Figure 5 displays the distribution of poll scores for Trump and Harris, revealing differences in the concentration of their approval ratings. The x-axis shows poll scores, with values closer to zero indicating higher approval, while the y-axis represents the density or frequency of these scores. Trump's distribution, represented in teal, peaks prominently around -1.0, indicating that his ratings tend to cluster around this lower approval range. Harris, shown in red, has a slightly broader distribution with a peak closer to -0.75, suggesting somewhat higher ratings on average compared to Trump. This visualization suggests that, while both candidates face generally low approval ratings, Harris's distribution includes slightly higher scores, potentially indicating a modest advantage in public perception over Trump.

4.2 Model Prediction Result

4.2.1 Bayesian Result

The Bayesian model Figure 6 predictions provide a nuanced view of the poll trends for Kamala Harris and Donald Trump over time. By incorporating prior information, this approach allows for flexible updates as new data points emerge, enhancing the reliability of the trend predictions. In these results, Harris's predicted average support is estimated at 48.3%, slightly ahead of Trump's at 47.1%. The smooth lines in the graph represent the predicted trends, with Harris shown in blue and Trump in red. These lines indicate the direction and consistency of public support across the observed period, suggesting Harris may have a modest lead. The shaded

areas around each line reflect the uncertainty in the predictions, illustrating the variability in poll data and capturing the possible range of outcomes. Although Harris's predicted support is somewhat higher, the overlapping prediction intervals highlight that the race remains close, with potential shifts possible as new polling data becomes available. This Bayesian model thus emphasizes a competitive landscape with no definitive outcome at this stage, underscoring the evolving nature of public sentiment as election day approaches. But it shows clearly that Harris is ahead of Trump.
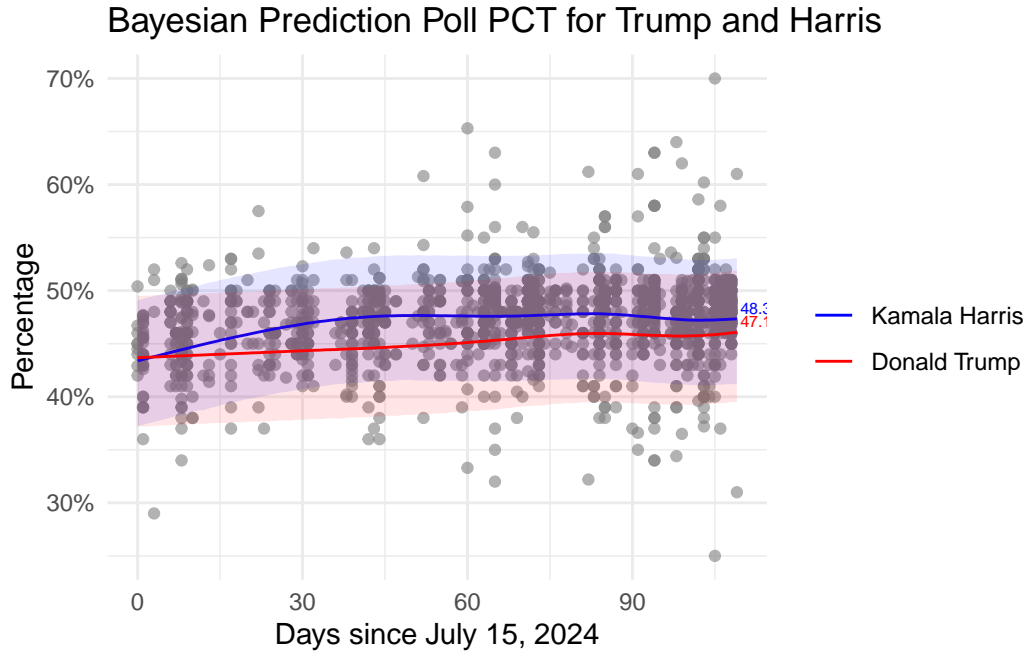
Figure 6: Bayesian Model Prediction for Harris and Trump

4.2.2 Linear Model Predictions Result

This graph Figure 7 illustrates the linear model predictions for Kamala Harris and Donald Trump over time, with data points representing poll percentages. Each dashed line corresponds to a linear regression model generated by individual pollsters, highlighting the variability and specific trends in polling data reported by each source. These dashed lines show how different pollsters have projected support levels for each candidate, providing insight into polling diversity.

The solid, curved lines represent the overall trend for Harris and Trump after aggregating data from all pollsters. These lines offer a smoothed view of each candidate's polling trajectory over time, indicating the general direction of support. The result shows that Harris is a little bit ahead of Trump with slightly differenece. This visualization effectively captures both the

consistency and fluctuations in polling data, demonstrating how individual pollster projections contribute to the overarching trends for each candidate.
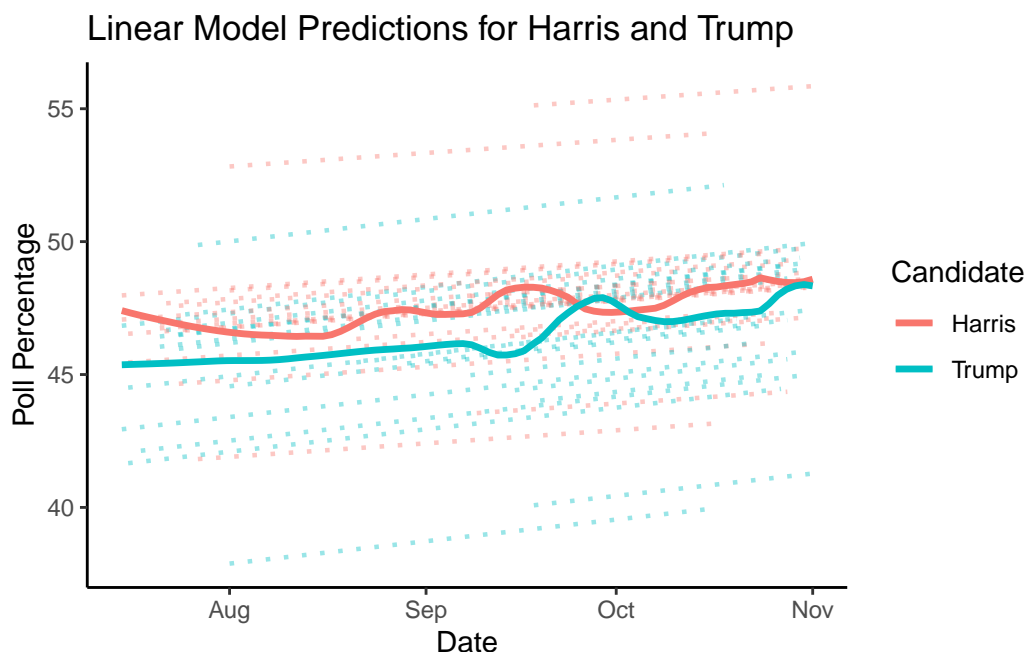
Figure 7: Linear Model Prediction for Harris and Trump

5 Discussion

5.1 Summary of the study

This paper provides an in-depth analysis of Donald Trump's polling trends as a candidate in the 2024 U.S presidential election, aiming to estimates his projected support among voters. By compiling polling data from FiveThirtyEight and other sources, this study applies both linear and Bayesian model to track changes in Trump's approval ratings over time, while also accounting for regional differences and pollster biases.

5.1.1 Data Collection and Methodology

The polling data, primarily from FiveThirtyEight, incorporates polls from various high-quality sources with a pollster grading above 2.7, limiting the analysis to credible polling organizations and controlling for methodological differences. Variables examined include: - **Approval Rating (Pct)**: Represents Trump's support percentage across different polls. - **Geographical Indicators (State)**: Used to examine regional variation in Trump's support. - **Poll Quality**

(Pollscore): Reflects each pollster's rating to account for methodological rigor. - **Temporal Data (Date)**: Used to capture shifts in public sentiment leading up to the election. - **Pollster**: Accounts for biases due to different polling agencies.

5.1.2 Modeling Approches

1. **Linear Model** /- Designed to capture baseline temporal trends in Trump's support. Tis model provided a straightforward view of the overall direction of Trump's popularity as election day approaches.
2. **Bayesian Regression with Spline Function** /- This more sophisticated model incorporates non-linear relationships, allowing for more flexible trend analysis, incorporating the effect of state, poll quality, and pollster, thus allowing for a probabilistic interpretation of support trends.

5.1.3 Key Findings

The findings reveal several important insights:

- **Temporal Trends**: The linear model illustrates that Trump's support is not static, showing moderate fluctuations in reponse to campagin events, media coverage, and possibly voter sentiment shifts. However, despite some variability, Trump's baseline support appears consistent within a certain range, with minor increases as election day nears.

- **Rigional Variability**: A key finding from the Bayesian model is the significant regional variability in Trump's approval ratings. States like Florida and Texas show relatively high support, while states such as California and New York reflect lower approval rates. This geographical breakdown highlights how Trump's popularity is concentrated in certain areas, aligning with political strongholds and regional ideological divides.

- **Cross-Pollster Variability**: Pollster biases contribute to fluctuations in the estimated support. Polls from higher-rated pollsters tend to show more stable, moderate support levels, while lower-rated pollsters exhibit greater variability in Trump's approval ratings. This adjustment provides a clearer view of actual trends by reducing noise caused by differing polling methods.

- **Voter Sentiment over Time**: The analyis showed that Trump's support fluctuated over time, with notable spikes following major campaign events and media coverage. These trends underscore the influence of ongoing political developments on voter sentiment, highlighting the dynamic nature of public opinion as the elction approaches.

5.1.4 Overall Results

The combined analysis from both linear and Bayesian models suggests that Trump maintains a steady base of support, with potential for incremental gains during key campaign moments. Regional and methodological factors significantly impact his approval ratings, underscoring a highly polarized voter landscape and the challenges in accurately gauging public sentiment.

This study offers a nuanced estimate of Trump's standing among voters, highlighting how public opinion can shift in response to major political events. By accounting for temporal, reginal, and poll-specific effects, the analysis provides a robust interpretation of polling trends, enhancing the reliability of Trump's support trajectory as election day nears.

The findings offer a valuable resource for political analysts and campaign strategists, shedding light on how Trump's bas and undecided voters might react to upcoming campaign dynamics. This layered perspective not only clarifies Trump's current standing but also enriches broader analyses of voter sentiment in a polarized election climate.

5.2 Influence of Regional Variation on Political Support

The analysis underscores the substantial impact of regional variation on political support, particularly when examining Donald Trump's and Kamala Harris's approval ratings across different states. In states with stable party alignments—such as California, where Harris's approval remains consistently high, and Texas, where Trump enjoys a solid base—support levels reflect entrenched party loyalties and exhibit minimal fluctuation. These states adhere to predictable patterns that align with historical voting trends, indicating that established strongholds provide a stabilizing effect on candidate support.

In contrast, swing states like Florida, Pennsylvania, and Wisconsin exhibit striking variability in support for both candidates, emerging as critical battlegrounds where voter sentiment is more responsive to campaign dynamics. For instance, Florida's polling data demonstrates competitive and fluctuating support, with both Trump and Harris reacting to national campaign developments and state-specific issues. In Pennsylvania, Trump's approval often spikes following campaign rallies or major policy announcements, which resonate strongly with local voter concerns. Meanwhile, in Wisconsin, Harris experiences gradual gains connected to discussions on social issues and healthcare—topics that align well with the values of her support base in that state. These fluctuations indicate that swing states are not only competitive but also sensitive to strategic events and messaging that resonate with local concerns.

The study's findings suggest that swing states act as a bellwether for broader shifts in public opinion and may reveal trends that could ultimately influence national outcomes. The heightened reactivity in these states highlights the importance of targeted, adaptive campaign strategies, as swing-state voters are more likely to adjust their support based on timely political developments. Unlike in stronghold states, where voter loyalty remains relatively steady,

swing-state support for both Trump and Harris appears to hinge on a combination of state-specific issues and national discourse. This insight emphasizes that while national polls provide a broad overview, they can overlook the nuanced regional shifts that ultimately drive election results in these decisive areas.

By examining these regional variations, the study reveals that the fate of the election could be determined by a few pivotal swing states, where the balance of support can shift dynamically in response to tailored outreach. For campaign strategists and political analysts, this finding underscores the importance of focusing on swing-state sentiment, which not only reflects the impact of national events on a local scale but also serves as a microcosm of the broader electorate's potential responsiveness. This nuanced understanding of regional variation illustrates that the election outcome may well rest on the success of capturing undecided voters in these key states, whose shifting allegiances could tip the balance in either candidate's favor.

5.3 The Role of Pollster Methodology and Bias in Shaping Public Perception

This study reveals that pollster methodologies and biases play a critical role in shaping public perception of candidates like Donald Trump and Kamala Harris. Through linear and Bayesian models, we see how sampling techniques, weighting methods, and survey phrasing can significantly impact reported support levels. These findings underscore that polling does not merely reflect public opinion but is also shaped by methodological choices, which can sometimes influence perceived candidate popularity more than actual shifts in voter sentiment.

5.3.1 Linear Model: Tracking Baseline Support

The linear model captures general approval trends, showing consistent differences in the support bases for Trump and Harris. Trump's approval ratings exhibit steady growth, suggesting a stable, possibly expanding base. Conversely, Harris's support appears more responsive to shifts in national discourse and specific campaign events. This contrast indicates that while Trump's base is relatively resilient to fluctuations, Harris's support may ebb and flow with the political climate.

5.3.2 Bayesian Model: Uncovering Methodological Bias

The Bayesian model takes this analysis further by accounting for pollster-specific effects, revealing how certain polling methodologies amplify or minimize support for each candidate. This model, with its flexible spline functions, provides a nuanced view of how Trump and Harris's approval levels differ based on demographic and regional sampling biases.

For example, the Bayesian model shows that Trump's support tends to appear stronger in online polls or those with higher representation of older, rural respondents—demographics that generally lean conservative. Harris's support, on the other hand, is often higher in polls

sampling younger, urban voters or utilizing live phone surveys. This methodological divide is particularly evident in swing states like Pennsylvania and Wisconsin, where sampling biases can sway candidate support, highlighting a demographic and methodological divide in voter sentiment.

5.3.3 Comparison Between Trump and Harris

This model-based analysis reveals a distinct difference between Trump and Harris in terms of how polling methodologies affect their perceived support. Trump's ratings remain relatively stable across online and high-quality polls, indicating a more uniform base among conservative demographics. Harris, however, shows more fluctuation, especially in polls with diverse sampling techniques or higher urban and younger populations. In a state like Florida, for instance, polling variation illustrates this contrast: Trump gains ground in regions with older, conservative demographics, while Harris's support spikes in urban centers and among younger voters.

5.3.4 Implications for Public Perception and Strategy

These findings emphasize that polling is not simply a reflection of voter sentiment but a construct shaped by methodological choices. The Bayesian model, by adjusting for biases, helps isolate these distortions and offers a clearer picture of genuine support for both Trump and Harris.

For campaign strategists, understanding these biases is essential for interpreting polling data accurately and tailoring strategies to key demographics. Aggregating or adjusting polls for known biases provides a more balanced view, helping campaigns prioritize outreach to areas with genuine support and undecided voters. In a polarized election, where even small perception shifts can impact behavior, filtering out methodological biases enables a more informed and targeted approach to public opinion.

6 Weaknesses and next steps

6.1 Weaknesses of the Study

While this study offers valuable insights into the influence of pollster methodology, regional variation, and demographic biases on public perception, several limitations need to be considered. These limitations could impact the accuracy and generalizability of the findings.

1. **Data Quality and Sampling Limitations** One limitation arises from variability in data quality across different polling organizations. Although the study controls for poll quality by using high-rated polls, the inherent reliance on polling data introduces potential sampling biases. Many polls depend on online or phone sampling, which can unintentionally exclude or under-represent certain demographic groups, such as older, less tech-savvy individuals or those in rural areas with limited internet or phone access. This under-representation may skew results, especially in states where these demographics play a significant role. Consequently, the data may slightly misrepresent the true level of support for each candidate, especially in rural or less-connected regions, potentially affecting the overall findings.

2. **Model Assumptions and Risk of Overfitting** The linear and Bayesian models, while valuable for capturing trends and controlling for pollster-specific effects, operate under certain assumptions that could limit their applicability. The linear model assumes a consistent trend over time, which may not fully capture the complexities of changing voter sentiment, especially around pivotal campaign events. The Bayesian model, with its flexible spline functions, could risk overfitting by closely adapting to the data. Overfitting may lead to models that perform well with the current dataset but lack generalizability to new or unseen data, such as future election cycles. This limitation could restrict the model's predictive accuracy if applied outside the current context or with different polling data.

3. **Influence of Unobserved Variables** An important limitation is the absence of unobserved variables that can influence voter sentiment but are difficult to integrate into polling models. Factors such as unexpected political events, fluctuations in media coverage, or sudden economic changes can substantially impact public opinion. For example, a major economic downturn or high-profile political event could shift approval ratings dramatically, and such shifts would not be fully accounted for in the study's models. The exclusion of these external variables may result in findings that do not entirely capture the dynamic and responsive nature of voter sentiment under real-world conditions.

4. **Regional and Demographic Representation Constraints** Although the study accounts for regional variation, it faces challenges in capturing the full diversity within specific demographic groups. For instance, urban populations can vary significantly by socioeconomic status, education, and race, factors that can lead to different responses to poll questions. However, polling data often aggregates these diverse groups, potentially masking important nuances. This limitation could affect the precision of regional predictions, particularly in diverse swing states where subtle shifts within key demographics may have an outsized impact on overall candidate support. Without fully capturing these variations, the study's predictions in such regions may overlook crucial subgroup dynamics that influence electoral outcomes.

Implications of These Limitations These limitations suggest that while the study's models provide a structured view of candidate support, the findings should be interpreted with

caution. Polling data is inherently a snapshot of voter sentiment, shaped by the biases in sampling and methodology. Additionally, model assumptions and the exclusion of influential, unobserved variables limit the generalizability of these findings to other contexts or future elections. Recognizing these limitations is essential for understanding the scope of this study's conclusions and underscores the importance of refining polling methodologies and models to better capture the complexities of voter behavior. Future improvements could focus on integrating more diverse sampling methods and adjusting models to account for external factors, ultimately enhancing the reliability of predictions in dynamic electoral environments.

6.2 What is Left to Learn and Future Directions

While this study provides important insights, several areas for further research could enhance polling methods and improve the accuracy of public opinion data in future elections. Below, we've prioritized these directions based on their potential impact on the field.

1. **Polling Diversity and Sampling Techniques** Expanding the range of polling methods and sampling techniques is crucial for improving representation across all demographics. By incorporating in-person polling in rural areas, using text surveys for younger voters, or experimenting with combined polling methods, future studies could reduce sampling biases and capture a more balanced snapshot of voter sentiment. Ensuring that traditionally underrepresented groups are better represented would make polling data more accurate and equitable, particularly in diverse swing states where demographic shifts can be pivotal.

2. **Better Poll Aggregation and Bias Correction** Developing more advanced methods for aggregating polls and adjusting for biases could significantly enhance the reliability of polling data. Techniques such as weighted averages, machine learning for bias correction, and multi-source data fusion would allow analysts to produce a more balanced and accurate picture of candidate support across polling organizations. This approach helps mitigate pollster-specific biases and is essential for making fair, reliable comparisons across diverse datasets, which is especially important in polarized elections.

3. **Real-Time Data and Tracking of Immediate** Events Integrating real-time data would offer timely insights into how sudden events, such as debates or economic shifts, impact voter sentiment. Future research could incorporate social media analytics, rapid-response polls, and news tracking to observe shifts as they happen, providing campaigns and analysts with tools to respond quickly to public opinion changes. This adaptability is essential in fast-moving election cycles, where sentiment can change drastically in response to recent events.

4. **Advanced Bayesian Models for Predictive Analytics** Applying more sophisticated Bayesian models could enhance predictive accuracy, especially when adjusted for demographic factors, historical voting trends, or real-time data. These models are particularly useful in swing states where small changes in sentiment can impact outcomes significantly.

Advanced Bayesian models can provide a more flexible, detailed picture of voter behavior, which is valuable for predicting election results with greater confidence.

5. **Adjusting for Unobserved Variables** Many external factors, such as media coverage, economic trends, and significant political events, can shift public opinion but are challenging to capture directly in polling models. Future models could incorporate proxies, like economic indicators or media sentiment scores, to better account for these influences. This would provide a more realistic understanding of how these unobserved factors shape voter behavior, especially in unpredictable election environments.

6. **Long-Term Voter Sentiment Analysis** Finally, studying voter sentiment trends over a longer time horizon could provide insights into how demographic shifts and social changes shape political support over time. This broader perspective would help campaigns and analysts understand and plan for generational or regional shifts, which can have a lasting impact on party alignment and voter preferences in future election cycles.

A Appendix A

Polling Methodology for CBS News/YouGov Survey (October 11-16, 2024)

1. Population, Frame, and Sample This CBS News/YouGov survey took place from October 11-16, 2024, with 1,439 registered voters in Arizona. The survey focused on registered voters in Arizona, and the sample was weighted to match key demographics like gender, age, race, and education. The weights were based on data from the U.S. Census American Community Survey, the U.S. Census Current Population Survey, and voter turnout data from the 2020 Presidential election.

2. Sample Recruitment The recruitment process focused on including respondents representative of Arizona's registered voter population by adjusting for demographic factors such as age, race, gender, and education. The sample was recruited primarily from various online panels, which included a mixture of respondents across demographic lines, to ensure a representative sample: 1,152 respondents were selected from YouGov's online panel. 212 respondents from Pure Spectrum's panel. 49 respondents from Dynata. 17 respondents from Cint's panel. 9 respondents from ROI Rocket's panel.

Surveys were conducted in both English and Spanish to account for language preferences among respondents. The weights applied to the data ranged from 0.1 to 5.0, with a mean of 1 and a standard deviation of 0.8, ensuring that the sample was representative of Arizona's voting population.

3. Sampling Approach and Trade-offs Sampling Approach: This survey employed stratified random sampling and applied post-survey weighting to ensure accurate representation of key demographic groups. Stratified random sampling is a technique that divides the overall population into several subgroups (or strata) based on specific attributes such as age, gender, race, and education level. Random samples are then drawn from each subgroup. The advantage of this method is that it ensures adequate representation across each stratum, preventing certain groups from being underrepresented or overlooked in the sample. After the survey was completed, weighting was applied to further adjust the sample to match the demographic distribution of registered voters in Arizona. This process involved adjusting the weights of individuals in the sample to better reflect the true composition of the overall voter population. This allows for a more accurate capture of how different demographic factors (such as gender, age, race, and education) influence voting behavior, thereby improving the external validity of the results. By using stratified random sampling and weighting, the researchers were able to minimize sampling bias and increase the accuracy of predicting voter tendencies. Trade-offs: One of the main limitations of this sampling method is its reliance on online panels, which may exclude individuals without internet access, thus introducing selection bias. While online surveys are cost-effective and convenient for collecting large samples, this reliance may result in certain groups (such as older individuals, low-income households, or voters living in remote areas) being left out due to lack of internet access. This means that some groups

might be underrepresented in the sample, which can affect the overall representativeness of the results. Although the weighting process can help adjust the sample to better reflect demographic differences, some errors are still difficult to completely eliminate. For example, when filling out surveys, respondents might overstate or understate their voting intentions due to social pressure or personal emotions—this is known as self-reporting bias. Even after weighting adjustments, such biases may persist and impact the accuracy of the final survey results. Therefore, while weighting can improve the representativeness of the results to a certain extent, systematic biases like non-response bias or selection bias may still leave traces in the final outcomes. Researchers need to interpret these potential errors with caution.

4. Regression Model A regression model was used to estimate each respondent's likelihood of voting. This model combined self-reported voting intentions with demographic and historical voting data, such as: Age, gender, race, and education. Voting history from past elections. The regression model allowed the survey to distinguish "likely voters" from the broader pool of registered voters, improving the accuracy of the predictions. By analyzing both individual and aggregate data, the model offered a more reliable estimate of actual voter turnout, thus increasing the precision of the survey's results.

5. Handling Non-response In surveys, non-response can cause bias because these people might have different voting behaviors or opinions. To fix this potential bias, researchers use weighting to adjust the sample data. Specifically, they assign different weights to respondents based on key demographic factors like gender, age, race, and education. The main goal of this weighting process is to make sure that even if some voters didn't respond, the final sample still accurately represents the overall population of registered voters in Arizona. This adjustment helps make the sample more representative and reduces the bias caused by non-response, improving the reliability and accuracy of the survey results. Besides that, weighting helps balance the proportion of different groups in the sample, ensuring that certain groups (like those with less access to the internet) are properly reflected in the results. In the end, this process helps make sure the survey results are more valid and can be applied more effectively to predict real voter behavior.

6. Strengths and Weaknesses of the Questionnaire Strengths: This survey effectively covered the key issues that Arizona voters care about the most, such as the economy, immigration, abortion, and the state of democracy. By combining demographic factors and historical voting data, the reliability of the results was improved, making it more accurate in reflecting the voting preferences of different groups. Weaknesses: Since the survey relies on self-reported voting intentions, this might introduce some bias, as respondents could overestimate or underestimate their likelihood of voting. Also, because it depends on online panels, voters without internet access may have been excluded, which could affect the external validity of the results. The margin of error is ±3.3 points, showing that there's still some uncertainty in the findings.

7. Margin of Error The margin of error for this survey is ±3.3 points, within a 95% confidence interval. The formula to calculate the margin of error is: $\hat{p} \pm 100 \times \sqrt{((1+\text{CV}^2)/n)}$

Where CV is the coefficient of variation of the sample weights, and (n) is the sample size. This formula calculates the sampling error, meaning that 95% of the sample results should fall within this range. It's important to note that this margin doesn't account for non-sampling errors, such as biases from panel selection or respondent behavior.

B Appendix B

B.1 Research Methodology

We wanted to use a stratified sampling method for our study in consideration of our budget constraint. Stratified sampling is an efficient method that can ensure the coverage of diverse groups, by stratifying national electorates by geographic area (state), age, gender, and educational level so that we can ensure the representation of each group and in the meanwhile rationalize the budget control. Moreover, random samples will be drawn within each stratum to ensure coverage of various population groups. Based on budget constraints, the total sample size is set at 2,000 individuals, which is allocated to each stratum to ensure that each stratum provides statistically significant results. As a result, not only does the stratified sampling increase the representativeness of the data, but also it improves the accuracy of the results, making it extremely cost-effecient, especially when budgetary constraints are in place.

In order to obtain a more comprehensive sample coverage in the survey, respondents will be recruited through two primary approaches using online questionnaires and telephone surveys to ensure that voter groups of different ages and technology preferences are covered. To be more specific, for voters aged 18-64, a Google Forms online questionnaire will be utilized and promoted through social media advertisment like Facebook and Google Ads, with an estimated budget of $30,000 used to increase exposure and appeal. Online questionnaires are less expensive, suitable for wide distribution and convenient for collecting large amounts of data. For senior voters aged 65 and over, we can opt for a telephone survey to overcome the fact that this group may not be familiar with or use online tools. The budget for this component was $40,000, covering costs such as call costs, phone carrier fees, and data entry to ensure representation of the senior population. In order to increase the completion and response rate of the questionnaire, $10,000 of the total budget will also be allocated for incentives such as small gift cards or raffle entries. Those incentives will be particularly effective for younger respondents to the online questionnaire, significantly reducing abandonment and improving the quality of the data.

The questionnaire will be designed to cover key voter demographics, political leanings, candidate preferences, policy attitudes, etc. It ensures that the major factors likely to influence voters' voting choices are captured. The questionnaire will be distributed through Google Forms, while the questionnaire content of the telephone survey remains consistent to ensure data compatibility and consistency. Specifically, the questionnaire will include demographic

information such as age, gender, state of residence, education, and annual income to help analyze the preferences of different groups. In addition, the questionnaire will ask respondents about their political leanings (from "very conservative" to "very liberal"), candidate support intentions (Candidate A, Candidate B, other candidates, or undecided), and views on major debated policies by two parties such as health insurance, education policy, and tax policy, immigration policy. Finally, the questionnaire will also include a voting intention question asking respondents if they plan to run in the upcoming presidential election.

Upon completion of data collection, the data will undergo rigorous validation and cleaning steps to ensure the quality and accuracy of the data. For online questionnaires, IP restrictions will be set to prevent duplicate submissions from the same IP address, while logic checks will be performed to identify unreasonable responses, such as quickly completed questionnaires or inconsistent responses. Telephone data will be double-validated for consistency checking and data entry to ensure accuracy. During the data cleaning process, $10,000 of the budget will be used for labor costs for data validation and cleaning. The data cleaning step helps to improve the reliability of the data and provide high quality data to support subsequent analysis.

After that, the cleaned data will be weighted by stratified characteristics to ensure that the representation of different districts and groups is consistent with the characteristics of the U.S. presidential election. Since the U.S. presidential election adopts the Electoral College system, and the outcome of each state directly affects the electoral votes attributed to that state, rather than being based on the number of votes cast in the national popular vote, our data analysis will focus on voting trends at the state level. In our weighted average calculations, we'll weigh each state based on its percentage of voter population to ensure that states with larger populations in the aggregate data, such as California, Texas, and Florida, will have a greater impact on the results. This weighted treatment better simulates the influence of states in actual elections. Additionally, we adjust the weights based on the number of electoral votes in each state, paying particular attention to voter preferences in key "swing states" such as Florida, Pennsylvania, and Michigan. In addition to weighting, we will use regression analysis to explore the impact of demographic characteristics, such as age, income, and education level, on voter preferences to reveal trends in support for various kinds of groups. The data analysis budget is $10,000 for hiring a professional data analyst or purchasing statistical software services to ensure the scientific validity of the project.

B.2 Google Form Link of Survey

https://docs.google.com/forms/d/e/1FAIpQLSedLCpBggwaUAy1xA2el9yrE3M31_YVVGOnsB1m EVh73uDahw/viewform?usp=sf_link

C Appendix C

C.1 Detailed Linear Model Data summary

Table 4: Summary of Linear Models for Harris and Trump

Term	Harris Linear Model	Trump Linear Model
(Intercept)	-280.143 (88.77)	-494.015 (78.342)
end_date	0.016 (0.004)	0.027 (0.004)
pollsterBeacon/Shaw	0.95 (0.645)	-0.747 (0.621)
pollsterCES / YouGov	1.372 (0.762)	-2.08 (0.758)
pollsterChristopher Newport U.	4.181 (3.4)	-8.188 (3.383)
pollsterCNN/SSRS	0.229 (0.881)	-2.792 (0.876)
pollsterData Orbital	-3.861 (2.423)	-1.225 (2.41)
pollsterEchelon Insights	0.327 (0.994)	-0.721 (0.866)
pollsterEmerson	1.486 (0.595)	-0.155 (0.573)
pollsterIpsos	-1.052 (0.612)	-5.352 (0.6)
pollsterMarist	1.074 (0.685)	-0.504 (0.681)
pollsterMarquette Law School	1.394 (0.819)	-2.892 (0.811)
pollsterMassINC Polling Group	7.571 (1.15)	-8.673 (1.143)
pollsterMcCourtney Institute/YouGov	0.296 (3.401)	-3.998 (3.383)
pollsterMonmouth	-0.864 (1.738)	-4.263 (1.729)
pollsterMuhlenberg	0.427 (3.402)	-0.781 (3.384)
pollsterQuinnipiac	0.149 (0.694)	-1.105 (0.68)
pollsterSelzer	-4.442 (3.403)	-1.564 (3.385)
pollsterSiena	6.058 (1.447)	-9.564 (1.438)
pollsterSiena/NYT	-1.987 (0.543)	-1.264 (0.537)
pollsterSuffolk	0.095 (1.101)	-4.938 (1.057)
pollsterSurveyUSA	0.577 (0.924)	-4.049 (0.868)
pollsterSurveyUSA/High Point University	-0.759 (1.998)	-2.536 (1.987)
pollsterThe Washington Post	-1.118 (1.145)	-0.951 (1.139)
pollsterU. North Florida	-4.868 (2.428)	2.559 (2.414)
pollsterUniversity of Massachusetts Lowell/YouGov	0.062 (1.263)	-4.76 (1.257)
pollsterWashington Post/George Mason University	0.559 (0.862)	-3.508 (0.857)
pollsterYouGov	0.027 (0.528)	-2.51 (0.521)
pollsterYouGov Blue	-0.901 (3.413)	-1.667 (3.393)
pollsterYouGov/Center for Working Class Politics	-0.986 (3.4)	-4.434 (3.383)

C.2 Detailed Bayesian Model summary

```
Model Info:
 function:     stan_glm
 family:       gaussian [identity]
 formula:      pct ~ ns(end_date_num, df = 5) + state + pollscore + pollster
 algorithm:    sampling
 sample:       4000 (posterior sample size)
 priors:       see help('prior_summary')
 observations: 782
 predictors:   68
```

Estimates:

	mean	sd	10%	50%
(Intercept)	51.0	2.4	47.9	51.0
ns(end_date_num, df = 5)1	1.3	0.4	0.7	1.3
ns(end_date_num, df = 5)2	2.6	0.6	1.9	2.6
ns(end_date_num, df = 5)3	1.7	0.5	1.1	1.7
ns(end_date_num, df = 5)4	2.8	0.8	1.8	2.8
ns(end_date_num, df = 5)5	1.9	0.4	1.4	1.9
stateCalifornia	-12.9	1.1	-14.3	-12.9
stateConnecticut	-8.4	2.1	-11.1	-8.4
stateFlorida	3.0	0.6	2.2	3.1
stateGeorgia	-0.2	0.4	-0.7	-0.2
stateIndiana	6.9	1.9	4.4	6.9
stateIowa	-0.6	3.7	-5.4	-0.7
stateMaine	-4.4	2.0	-6.9	-4.5
stateMaine CD-1	-9.6	2.0	-12.1	-9.6
stateMaine CD-2	0.7	2.0	-1.8	0.7
stateMaryland	-15.0	1.4	-16.9	-15.1
stateMassachusetts	-14.4	1.0	-15.7	-14.4
stateMichigan	-1.9	0.4	-2.4	-1.9
stateMinnesota	-4.3	0.7	-5.2	-4.3
stateMissouri	5.8	1.5	4.0	5.8
stateMontana	7.5	0.8	6.4	7.5
stateNational	-1.6	0.3	-2.1	-1.6
stateNebraska	6.3	0.9	5.2	6.3
stateNebraska CD-1	2.4	2.0	-0.2	2.4
stateNebraska CD-2	-5.8	0.8	-6.8	-5.8
stateNebraska CD-3	18.6	2.0	16.0	18.6
stateNevada	-1.5	0.5	-2.1	-1.5

```
stateNew Hampshire                                        -4.0   0.8   -5.1   -4.0
stateNew Mexico                                           -5.2   1.1   -6.6   -5.2
stateNew York                                             -9.2   1.2  -10.7   -9.2
stateNorth Carolina                                       -0.7   0.4   -1.2   -0.7
stateOhio                                                  2.0   0.7    1.1    2.0
statePennsylvania                                         -1.3   0.4   -1.8   -1.3
stateRhode Island                                         -5.0   1.6   -7.1   -4.9
stateSouth Dakota                                          9.8   2.0    7.3    9.8
stateTexas                                                 2.1   0.5    1.4    2.1
stateVirginia                                             -4.0   0.7   -4.9   -4.0
stateWashington                                          -11.2   2.0  -13.7  -11.2
stateWisconsin                                            -1.4   0.4   -1.9   -1.5
pollscore                                                  3.2   2.8   -0.4    3.2
pollsterBeacon/Shaw                                        0.3   0.9   -0.9    0.3
pollsterCES / YouGov                                      -0.8   1.0   -2.0   -0.8
pollsterChristopher Newport U.                            -5.4   2.0   -8.0   -5.4
pollsterCNN/SSRS                                          -3.0   0.8   -4.1   -3.0
pollsterData Orbital                                      -1.7   1.5   -3.7   -1.7
pollsterEchelon Insights                                  -0.2   0.6   -0.9   -0.2
pollsterEmerson                                            1.0   0.9   -0.1    1.0
pollsterIpsos                                             -4.6   0.5   -5.2   -4.6
pollsterMarist                                            -0.8   0.5   -1.4   -0.8
pollsterMarquette Law School                              -1.6   1.0   -2.8   -1.6
pollsterMassINC Polling Group                             -3.1   0.9   -4.3   -3.1
pollsterMcCourtney Institute/YouGov                       -2.4   2.1   -5.1   -2.4
pollsterMonmouth                                          -3.4   1.1   -4.7   -3.4
pollsterMuhlenberg                                        -0.1   2.0   -2.7   -0.1
pollsterQuinnipiac                                        -2.1   1.0   -3.3   -2.1
pollsterSelzer                                            -0.7   3.7   -5.5   -0.7
pollsterSiena                                             -1.5   1.4   -3.3   -1.4
pollsterSiena/NYT                                          0.0   2.0   -2.5    0.0
pollsterSuffolk                                           -2.1   0.8   -3.2   -2.1
pollsterSurveyUSA                                         -0.8   1.5   -2.7   -0.8
pollsterSurveyUSA/High Point University                   -1.2   1.8   -3.5   -1.2
pollsterThe Washington Post                               -0.1   1.3   -1.8   -0.1
pollsterU. North Florida                                  -2.1   1.6   -4.2   -2.1
pollsterUniversity of Massachusetts Lowell/YouGov         -3.0   0.9   -4.2   -3.0
pollsterWashington Post/George Mason University           -2.6   0.6   -3.4   -2.6
pollsterYouGov                                            -1.4   0.9   -2.5   -1.4
pollsterYouGov Blue                                       -1.0   2.0   -3.6   -1.1
pollsterYouGov/Center for Working Class Politics          -3.1   2.1   -5.7   -3.0
sigma                                                      2.1   0.1    2.0    2.1
                                                          90%
```

```
(Intercept)                              54.0
ns(end_date_num, df = 5)1                 1.8
ns(end_date_num, df = 5)2                 3.4
ns(end_date_num, df = 5)3                 2.3
ns(end_date_num, df = 5)4                 3.9
ns(end_date_num, df = 5)5                 2.4
stateCalifornia                         -11.6
stateConnecticut                         -5.8
stateFlorida                              3.9
stateGeorgia                              0.3
stateIndiana                              9.3
stateIowa                                 4.1
stateMaine                               -1.9
stateMaine CD-1                          -7.0
stateMaine CD-2                           3.2
stateMaryland                           -13.2
stateMassachusetts                      -13.2
stateMichigan                            -1.4
stateMinnesota                           -3.3
stateMissouri                             7.6
stateMontana                              8.5
stateNational                            -1.2
stateNebraska                             7.4
stateNebraska CD-1                        4.8
stateNebraska CD-2                       -4.8
stateNebraska CD-3                       21.2
stateNevada                              -0.9
stateNew Hampshire                       -2.9
stateNew Mexico                          -3.8
stateNew York                            -7.8
stateNorth Carolina                      -0.2
stateOhio                                 2.9
statePennsylvania                        -0.9
stateRhode Island                        -2.9
stateSouth Dakota                        12.4
stateTexas                                2.8
stateVirginia                            -3.1
stateWashington                          -8.6
stateWisconsin                           -1.0
pollscore                                 6.8
pollsterBeacon/Shaw                       1.5
pollsterCES / YouGov                      0.4
pollsterChristopher Newport U.           -2.8
```

```
pollsterCNN/SSRS                                           -2.0
pollsterData Orbital                                        0.1
pollsterEchelon Insights                                    0.5
pollsterEmerson                                             2.1
pollsterIpsos                                              -4.0
pollsterMarist                                             -0.1
pollsterMarquette Law School                               -0.4
pollsterMassINC Polling Group                              -2.0
pollsterMcCourtney Institute/YouGov                         0.2
pollsterMonmouth                                           -2.0
pollsterMuhlenberg                                          2.5
pollsterQuinnipiac                                         -0.8
pollsterSelzer                                              4.2
pollsterSiena                                               0.3
pollsterSiena/NYT                                           2.5
pollsterSuffolk                                            -1.1
pollsterSurveyUSA                                           1.1
pollsterSurveyUSA/High Point University                     1.1
pollsterThe Washington Post                                 1.5
pollsterU. North Florida                                    0.0
pollsterUniversity of Massachusetts Lowell/YouGov          -1.8
pollsterWashington Post/George Mason University            -1.9
pollsterYouGov                                             -0.3
pollsterYouGov Blue                                         1.6
pollsterYouGov/Center for Working Class Politics           -0.4
sigma                                                       2.1


Fit Diagnostics:
          mean   sd   10%   50%   90%
mean_PPD 46.8    0.1 46.7  46.8  47.0


The mean_ppd is the sample average posterior predictive distribution of the outcome variable

MCMC diagnostics
                                              mcse Rhat n_eff
(Intercept)                                    0.1  1.0   514
ns(end_date_num, df = 5)1                      0.0  1.0  5127
ns(end_date_num, df = 5)2                      0.0  1.0  4361
ns(end_date_num, df = 5)3                      0.0  1.0  4750
ns(end_date_num, df = 5)4                      0.0  1.0  4598
ns(end_date_num, df = 5)5                      0.0  1.0  5884
stateCalifornia                                0.0  1.0  6635
stateConnecticut                               0.0  1.0  6447
```

stateFlorida	0.0	1.0	3723
stateGeorgia	0.0	1.0	1712
stateIndiana	0.0	1.0	8072
stateIowa	0.1	1.0	3882
stateMaine	0.0	1.0	7629
stateMaine CD-1	0.0	1.0	5982
stateMaine CD-2	0.0	1.0	7810
stateMaryland	0.0	1.0	7654
stateMassachusetts	0.0	1.0	4435
stateMichigan	0.0	1.0	1915
stateMinnesota	0.0	1.0	3631
stateMissouri	0.0	1.0	7421
stateMontana	0.0	1.0	5365
stateNational	0.0	1.0	1592
stateNebraska	0.0	1.0	4400
stateNebraska CD-1	0.0	1.0	8242
stateNebraska CD-2	0.0	1.0	4958
stateNebraska CD-3	0.0	1.0	9435
stateNevada	0.0	1.0	2586
stateNew Hampshire	0.0	1.0	4417
stateNew Mexico	0.0	1.0	5651
stateNew York	0.0	1.0	4279
stateNorth Carolina	0.0	1.0	2005
stateOhio	0.0	1.0	4318
statePennsylvania	0.0	1.0	1675
stateRhode Island	0.0	1.0	5495
stateSouth Dakota	0.0	1.0	9918
stateTexas	0.0	1.0	3528
stateVirginia	0.0	1.0	3637
stateWashington	0.0	1.0	6882
stateWisconsin	0.0	1.0	1821
pollscore	0.1	1.0	477
pollsterBeacon/Shaw	0.0	1.0	494
pollsterCES / YouGov	0.0	1.0	591
pollsterChristopher Newport U.	0.0	1.0	4707
pollsterCNN/SSRS	0.0	1.0	850
pollsterData Orbital	0.0	1.0	6020
pollsterEchelon Insights	0.0	1.0	6171
pollsterEmerson	0.0	1.0	508
pollsterIpsos	0.0	1.0	1073
pollsterMarist	0.0	1.0	953
pollsterMarquette Law School	0.0	1.0	569
pollsterMassINC Polling Group	0.0	1.0	4002

```
pollsterMcCourtney Institute/YouGov                      0.0  1.0  2785
pollsterMonmouth                                         0.0  1.0  4400
pollsterMuhlenberg                                       0.0  1.0  4440
pollsterQuinnipiac                                       0.0  1.0   636
pollsterSelzer                                           0.1  1.0  4134
pollsterSiena                                            0.0  1.0  5215
pollsterSiena/NYT                                        0.1  1.0   469
pollsterSuffolk                                          0.0  1.0   929
pollsterSurveyUSA                                        0.1  1.0   527
pollsterSurveyUSA/High Point University                  0.1  1.0   760
pollsterThe Washington Post                              0.1  1.0   577
pollsterU. North Florida                                 0.0  1.0  3115
pollsterUniversity of Massachusetts Lowell/YouGov 0.0   1.0  1153
pollsterWashington Post/George Mason University   0.0   1.0  4605
pollsterYouGov                                           0.0  1.0   503
pollsterYouGov Blue                                      0.0  1.0  2820
pollsterYouGov/Center for Working Class Politics  0.0   1.0  2993
sigma                                                    0.0  1.0  5744
mean_PPD                                                 0.0  1.0  6140
log-posterior                                            0.2  1.0  1229
```

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective

```
Model Info:
 function:     stan_glm
 family:       gaussian [identity]
 formula:      pct ~ ns(end_date_num, df = 5) + state + pollscore + pollster
 algorithm:    sampling
 sample:       4000 (posterior sample size)
 priors:       see help('prior_summary')
 observations: 733
 predictors:   68

Estimates:
```

	mean	sd	10%	50%
(Intercept)	43.9	2.4	40.9	43.9
ns(end_date_num, df = 5)1	3.9	0.5	3.3	3.9
ns(end_date_num, df = 5)2	4.7	0.6	3.9	4.8
ns(end_date_num, df = 5)3	2.7	0.5	2.1	2.7
ns(end_date_num, df = 5)4	8.0	1.0	6.7	8.0
ns(end_date_num, df = 5)5	0.8	0.4	0.3	0.8

```
stateCalifornia                    13.7   1.1  12.3  13.6
stateConnecticut                    2.8   2.1   0.1   2.8
stateFlorida                       -3.0   0.7  -3.8  -3.0
stateGeorgia                        0.2   0.4  -0.3   0.2
stateIndiana                       -6.5   1.9  -9.0  -6.5
stateIowa                          -1.7   3.7  -6.4  -1.8
stateMaine                          4.2   2.1   1.6   4.2
stateMaine CD-1                    10.2   2.0   7.6  10.2
stateMaine CD-2                    -1.7   2.0  -4.3  -1.7
stateMaryland                      15.2   1.4  13.3  15.2
stateMassachusetts                 12.0   1.1  10.6  12.0
stateMichigan                       1.2   0.4   0.7   1.2
stateMinnesota                      4.1   0.9   3.0   4.1
stateMissouri                      -4.8   1.5  -6.7  -4.8
stateMontana                       -6.8   0.8  -7.8  -6.8
stateNational                       1.0   0.4   0.6   1.0
stateNebraska                      -6.6   0.9  -7.7  -6.6
stateNebraska CD-1                 -3.3   2.0  -5.8  -3.3
stateNebraska CD-2                  5.2   0.7   4.2   5.2
stateNebraska CD-3                -18.6   1.9 -21.1 -18.6
stateNevada                         1.1   0.5   0.5   1.1
stateNew Hampshire                  3.5   0.9   2.4   3.5
stateNew Mexico                     3.8   1.1   2.4   3.8
stateNew York                       9.0   1.2   7.5   9.0
stateNorth Carolina                 1.0   0.4   0.5   1.0
stateOhio                          -2.6   0.7  -3.5  -2.6
statePennsylvania                   1.4   0.4   1.0   1.4
stateRhode Island                   4.4   1.6   2.3   4.4
stateSouth Dakota                  -9.9   2.0 -12.4  -9.9
stateTexas                         -2.8   0.6  -3.5  -2.8
stateVirginia                       2.9   0.8   1.9   2.9
stateWashington                     9.0   2.0   6.5   9.0
stateWisconsin                      2.0   0.4   1.4   2.0
pollscore                           1.1   2.8  -2.4   1.1
pollsterBeacon/Shaw                 1.1   0.9   0.0   1.1
pollsterCES / YouGov                2.3   0.9   1.1   2.3
pollsterChristopher Newport U.      1.2   2.1  -1.5   1.2
pollsterCNN/SSRS                   -0.5   0.8  -1.5  -0.5
pollsterData Orbital               -2.8   1.5  -4.7  -2.8
pollsterEchelon Insights            0.1   0.6  -0.7   0.1
pollsterEmerson                     1.4   0.9   0.3   1.4
pollsterIpsos                      -1.2   0.5  -1.8  -1.2
pollsterMarist                      1.6   0.5   0.9   1.6
```

pollsterMarquette Law School	0.9	1.0	-0.3	1.0
pollsterMassINC Polling Group	2.5	0.9	1.3	2.5
pollsterMcCourtney Institute/YouGov	-0.1	2.0	-2.7	-0.1
pollsterMonmouth	-1.2	1.1	-2.6	-1.2
pollsterMuhlenberg	-0.4	2.0	-2.9	-0.4
pollsterQuinnipiac	-0.7	1.0	-2.0	-0.7
pollsterSelzer	-1.8	3.7	-6.6	-1.8
pollsterSiena	-1.9	1.4	-3.8	-1.9
pollsterSiena/NYT	-0.7	2.0	-3.1	-0.7
pollsterSuffolk	-0.7	0.9	-1.8	-0.7
pollsterSurveyUSA	-0.3	1.5	-2.2	-0.3
pollsterSurveyUSA/High Point University	-0.7	1.8	-3.0	-0.7
pollsterThe Washington Post	-0.4	1.3	-2.1	-0.5
pollsterU. North Florida	-0.5	1.6	-2.8	-0.5
pollsterUniversity of Massachusetts Lowell/YouGov	-0.7	0.9	-1.9	-0.7
pollsterWashington Post/George Mason University	-0.3	0.6	-1.1	-0.3
pollsterYouGov	0.3	0.9	-0.8	0.3
pollsterYouGov Blue	-0.6	2.1	-3.3	-0.6
pollsterYouGov/Center for Working Class Politics	-1.5	2.1	-4.2	-1.5
sigma	2.1	0.1	2.0	2.1

	90%
(Intercept)	47.0
ns(end_date_num, df = 5)1	4.5
ns(end_date_num, df = 5)2	5.6
ns(end_date_num, df = 5)3	3.4
ns(end_date_num, df = 5)4	9.3
ns(end_date_num, df = 5)5	1.3
stateCalifornia	15.0
stateConnecticut	5.4
stateFlorida	-2.1
stateGeorgia	0.7
stateIndiana	-4.0
stateIowa	3.1
stateMaine	6.9
stateMaine CD-1	12.9
stateMaine CD-2	0.9
stateMaryland	17.0
stateMassachusetts	13.4
stateMichigan	1.7
stateMinnesota	5.1
stateMissouri	-3.0
stateMontana	-5.7
stateNational	1.5

```
stateNebraska                                        -5.5
stateNebraska CD-1                                   -0.7
stateNebraska CD-2                                    6.1
stateNebraska CD-3                                  -16.1
stateNevada                                           1.7
stateNew Hampshire                                    4.7
stateNew Mexico                                       5.3
stateNew York                                        10.5
stateNorth Carolina                                   1.5
stateOhio                                            -1.7
statePennsylvania                                     1.9
stateRhode Island                                     6.5
stateSouth Dakota                                    -7.4
stateTexas                                           -2.1
stateVirginia                                         3.9
stateWashington                                      11.6
stateWisconsin                                        2.5
pollscore                                             4.7
pollsterBeacon/Shaw                                   2.3
pollsterCES / YouGov                                  3.5
pollsterChristopher Newport U.                        3.9
pollsterCNN/SSRS                                      0.5
pollsterData Orbital                                 -0.9
pollsterEchelon Insights                              0.9
pollsterEmerson                                       2.5
pollsterIpsos                                        -0.6
pollsterMarist                                        2.2
pollsterMarquette Law School                          2.2
pollsterMassINC Polling Group                         3.7
pollsterMcCourtney Institute/YouGov                   2.6
pollsterMonmouth                                      0.2
pollsterMuhlenberg                                    2.1
pollsterQuinnipiac                                    0.5
pollsterSelzer                                        3.0
pollsterSiena                                        -0.1
pollsterSiena/NYT                                     1.9
pollsterSuffolk                                       0.4
pollsterSurveyUSA                                     1.6
pollsterSurveyUSA/High Point University               1.6
pollsterThe Washington Post                           1.2
pollsterU. North Florida                              1.5
pollsterUniversity of Massachusetts Lowell/YouGov     0.5
pollsterWashington Post/George Mason University       0.5
```

```
pollsterYouGov                                   1.4
pollsterYouGov Blue                              2.1
pollsterYouGov/Center for Working Class Politics  1.2
sigma                                            2.2


Fit Diagnostics:
          mean    sd    10%    50%    90%
mean_PPD 47.8    0.1  47.7   47.8   47.9


The mean_ppd is the sample average posterior predictive distribution of the outcome variable

MCMC diagnostics
                                        mcse  Rhat  n_eff
(Intercept)                              0.1  1.0    481
ns(end_date_num, df = 5)1                0.0  1.0   4061
ns(end_date_num, df = 5)2                0.0  1.0   4442
ns(end_date_num, df = 5)3                0.0  1.0   4289
ns(end_date_num, df = 5)4                0.0  1.0   3484
ns(end_date_num, df = 5)5                0.0  1.0   5897
stateCalifornia                          0.0  1.0   4968
stateConnecticut                         0.0  1.0   6133
stateFlorida                             0.0  1.0   3379
stateGeorgia                             0.0  1.0   2064
stateIndiana                             0.0  1.0   8699
stateIowa                                0.1  1.0   4002
stateMaine                               0.0  1.0   6501
stateMaine CD-1                          0.0  1.0   7645
stateMaine CD-2                          0.0  1.0   7396
stateMaryland                            0.0  1.0   5707
stateMassachusetts                       0.0  1.0   3640
stateMichigan                            0.0  1.0   1739
stateMinnesota                           0.0  1.0   4169
stateMissouri                            0.0  1.0   6384
stateMontana                             0.0  1.0   4353
stateNational                            0.0  1.0   1575
stateNebraska                            0.0  1.0   4777
stateNebraska CD-1                       0.0  1.0   6287
stateNebraska CD-2                       0.0  1.0   4074
stateNebraska CD-3                       0.0  1.0   7993
stateNevada                              0.0  1.0   2464
stateNew Hampshire                       0.0  1.0   4341
stateNew Mexico                          0.0  1.0   5909
stateNew York                            0.0  1.0   3861
```

```
stateNorth Carolina                                          0.0  1.0  1923
stateOhio                                                    0.0  1.0  4136
statePennsylvania                                            0.0  1.0  1770
stateRhode Island                                            0.0  1.0  4734
stateSouth Dakota                                            0.0  1.0  6358
stateTexas                                                   0.0  1.0  2770
stateVirginia                                                0.0  1.0  3512
stateWashington                                              0.0  1.0  6343
stateWisconsin                                               0.0  1.0  1816
pollscore                                                    0.1  1.0   433
pollsterBeacon/Shaw                                          0.0  1.0   442
pollsterCES / YouGov                                         0.0  1.0   507
pollsterChristopher Newport U.                               0.0  1.0  3568
pollsterCNN/SSRS                                             0.0  1.0   855
pollsterData Orbital                                         0.0  1.0  4654
pollsterEchelon Insights                                     0.0  1.0  5274
pollsterEmerson                                              0.0  1.0   447
pollsterIpsos                                                0.0  1.0   939
pollsterMarist                                               0.0  1.0   902
pollsterMarquette Law School                                 0.0  1.0   502
pollsterMassINC Polling Group                                0.0  1.0  3379
pollsterMcCourtney Institute/YouGov                          0.0  1.0  2547
pollsterMonmouth                                             0.0  1.0  3412
pollsterMuhlenberg                                           0.0  1.0  4323
pollsterQuinnipiac                                           0.0  1.0   616
pollsterSelzer                                               0.1  1.0  3132
pollsterSiena                                                0.0  1.0  4756
pollsterSiena/NYT                                            0.1  1.0   420
pollsterSuffolk                                              0.0  1.0   755
pollsterSurveyUSA                                            0.1  1.0   477
pollsterSurveyUSA/High Point University                      0.1  1.0   714
pollsterThe Washington Post                                  0.1  1.0   550
pollsterU. North Florida                                     0.0  1.0  3730
pollsterUniversity of Massachusetts Lowell/YouGov 0.0  1.0  1076
pollsterWashington Post/George Mason University   0.0  1.0  4562
pollsterYouGov                                               0.0  1.0   418
pollsterYouGov Blue                                          0.0  1.0  2461
pollsterYouGov/Center for Working Class Politics  0.0  1.0  2700
sigma                                                        0.0  1.0  6002
mean_PPD                                                     0.0  1.0  5495
log-posterior                                                0.2  1.0  1274
```

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effectiv

References

Alexander, Rohan. 2023a. *Telling Stories with Data.* Chapman; Hall/CRC. https://tellings to021eswithdata.com/.

————. 2023b. *Telling Stories with Data: With Applications in r.* Chapman; Hall/CRC.

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.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

FiveThirtyEight. 2024. "Presidential General Election Polls - 2024." https://projects.fivethi rtyeight.com/polls/president-general/2024/national/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Samuel Brilleman. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data.* https://doi.org/10.5281/zenodo.3960218.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

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

————. 2023b. *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, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'.* https://CRAN.R-project.org/package=arrow.

Robinson, David, Alex Hayes, and Simon Couch. 2023. *Broom: Convert Statistical Objects into Tidy Tibbles.* https://CRAN.R-project.org/package=broom.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

————. 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.

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.

Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization.* https://CRAN.R-project.org/package=scales.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.