2024 US Election Prediction*

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

Betty Liu   Jingchuan Xu   Dingshuo Li

November 4, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

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 is one of the most closely watched events in recent political history. Polling provides a glimpse into the dynamics of the race, offering insights that can shape campaign strategies and public opinion. Among the candidates, former President Donald Trump consistently draws significant attention and debate from the public across different political perspectives. Tracking Trump's support through the poll reveals his current popularity and potential electoral outcomes. The paper dives into Trump's polling data, analyzing data from various pollsters to see how his support has changed over time. The study explores whether there is an upward or downward trend using linear modelling in Trump's polling percentage, while also accounting for differences in his support reported by various pollsters. By investigating these patterns, the analysis aims to estimate Trump's likely level of support as election day approaches.

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

In this study, we apply statistical models, including a linear model of Trump's percentage over time, to quantify changes in his support level. Additionally, we adjust for pollster variations to provide a more accurate reflection of his support trajectory across different sources. Our findings reveal both temporal trends and variability among pollsters, highlighting the complexities of interpreting polling data in a polarized political environment. These insights contribute to a better understanding of Trump's standing and the potential outcomes in 2024.

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: 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), and splines R Core Team (2023b). 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: Preview of 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: Preview of Pollscore

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.
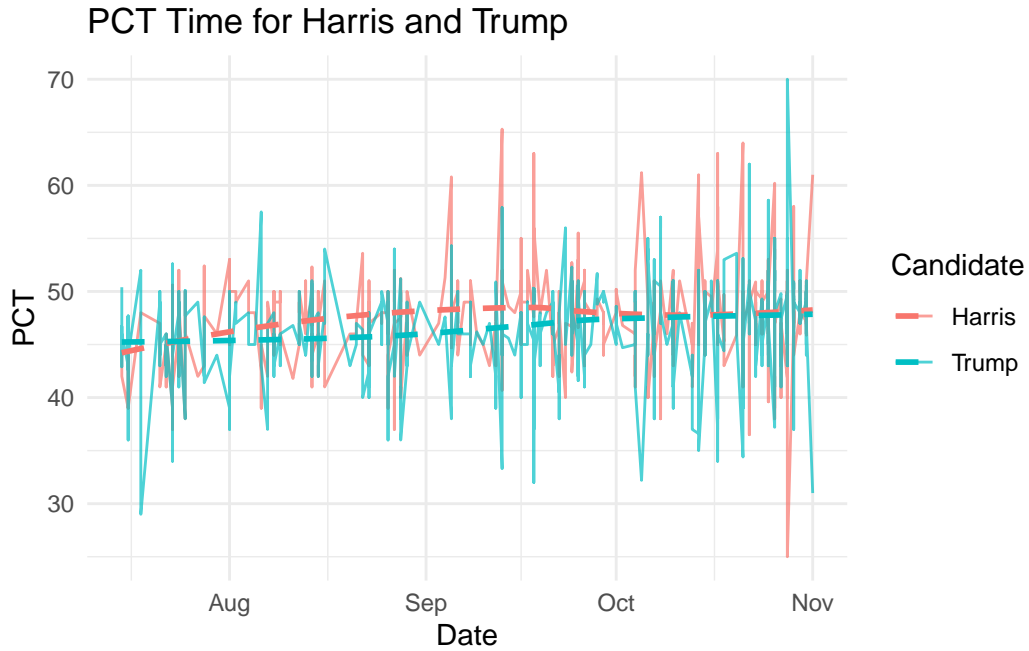
Figure 1: figure of PCT Change over Time

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 summary of linear model for both Trump and Harris is showing at (**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 overfitting. The more detailed summart 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.
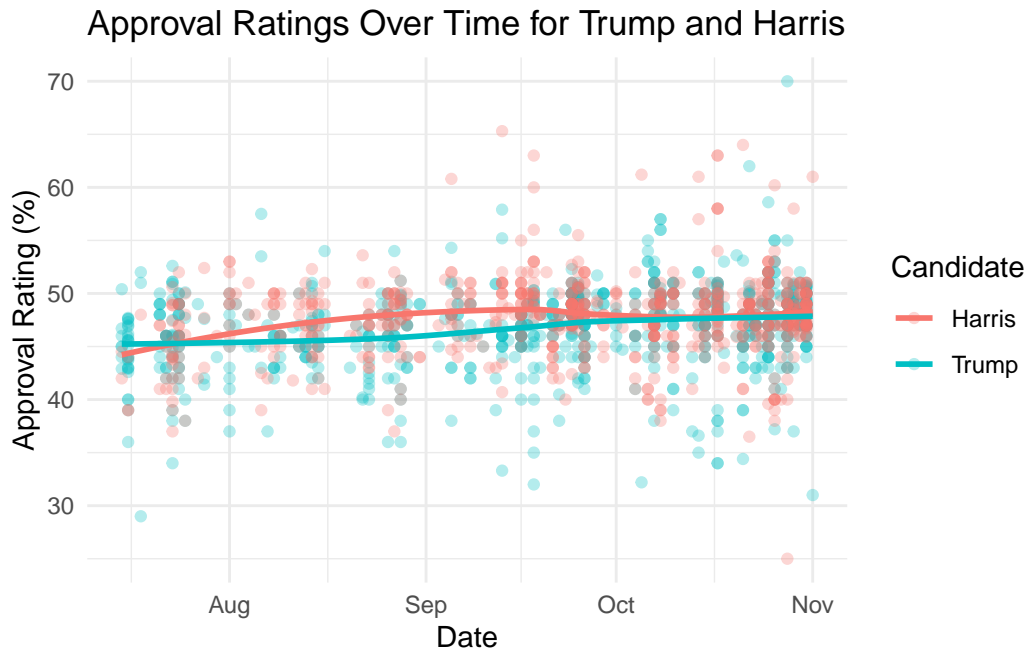
Figure 2: Comparsion of PCT Change over Time

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.
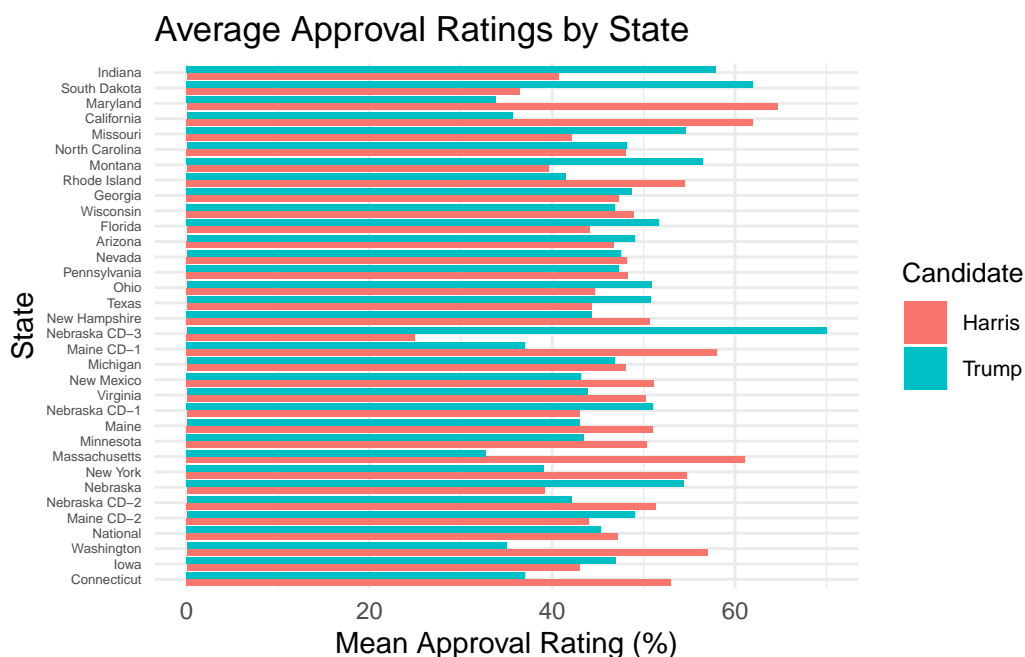
Figure 3: Average PCT by State in U.S.

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.
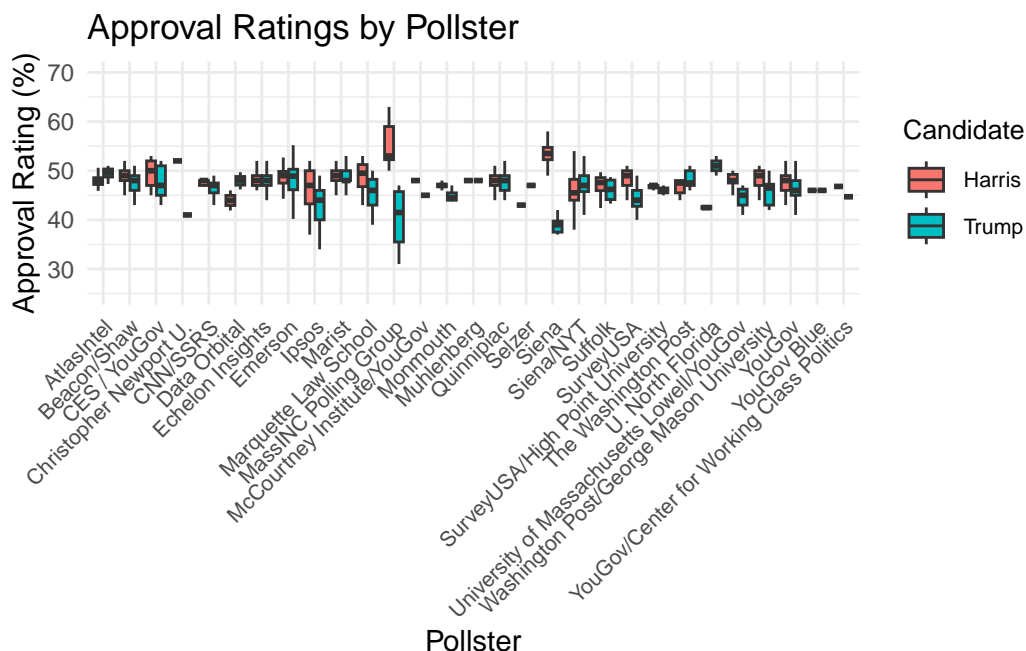
Figure 4: PCT of Different Pollster

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.
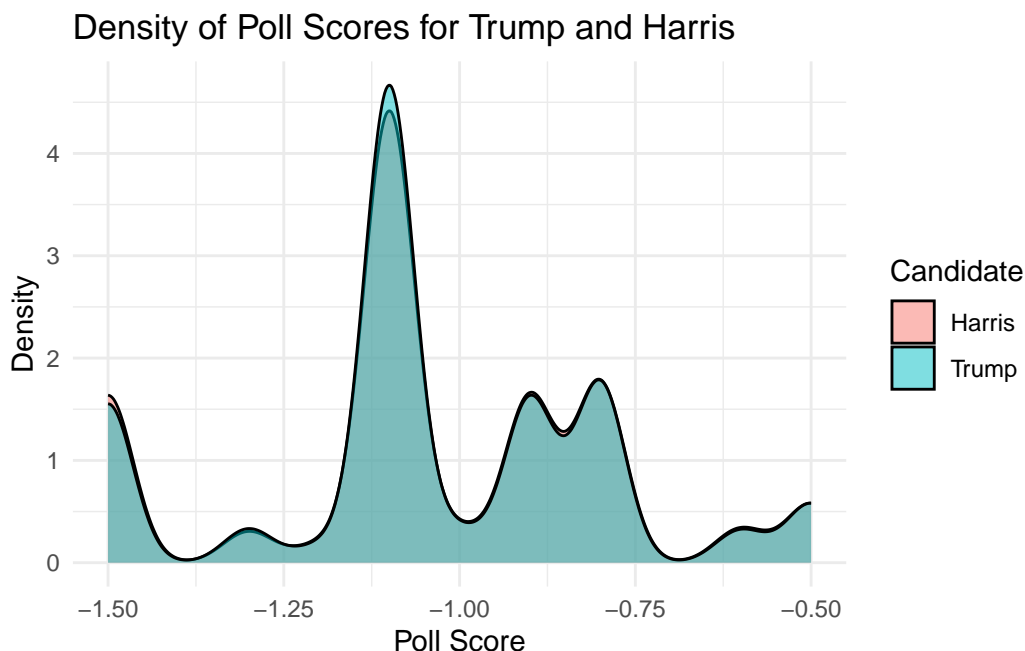
Figure 5: Density of Approval Ratings for Trump and Harris

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.
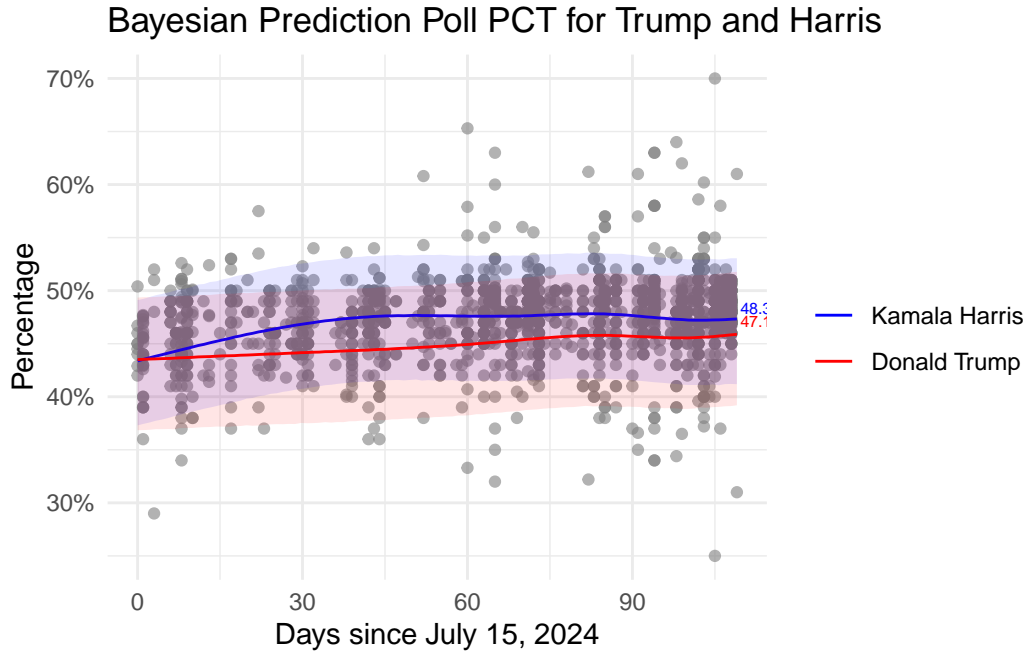
Figure 6: Bayesian Model Prediction for Harris and Trump

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.

Linear Model Predictions for Harris and Trump
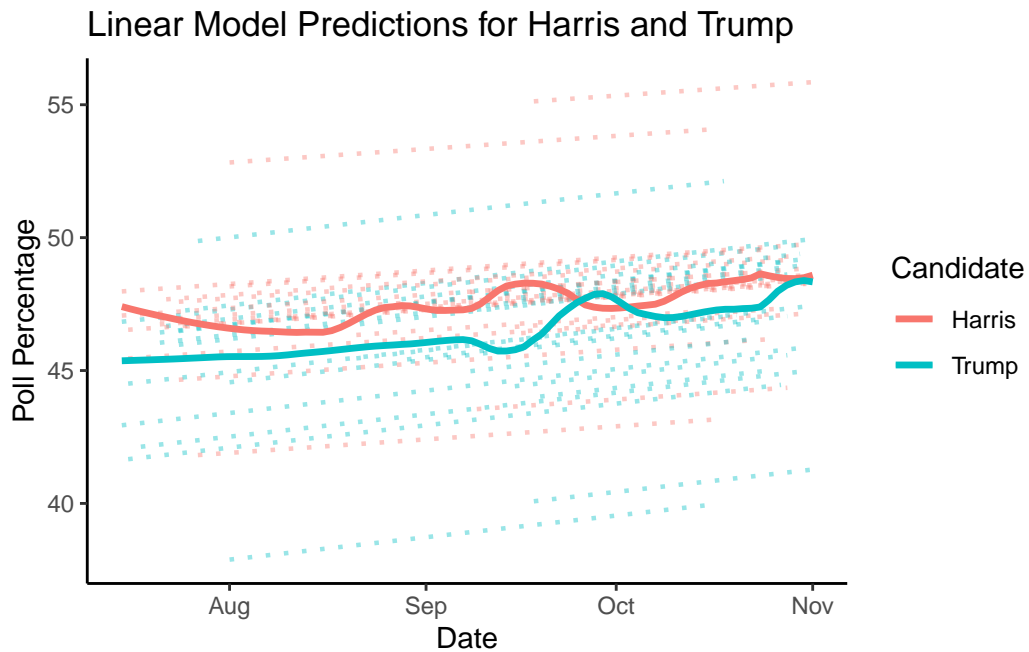
Figure 7: Linear Model Prediction for Harris and Trump

5 Discussion

5.1 First discussion point

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

5.2 Second discussion point

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

5.3 Third discussion point

6 Weaknesses and next steps

Weaknesses and next steps should also be included.

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: $\hat{p} \pm 100 \times \sqrt{((1+CV^2)/n)}$

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

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%	

24

```
(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 effectiv

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 effective

C.3 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

C.4 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

References

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

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.

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/packag e=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

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/.