Analysis of Trump Support Trends in 2024: A High-Quality Polling Comparison*

Tracking Trends: Variations in Donald Trump's Polling Percentages by Date, Pollster, and Poll Quality Leading to the 2024 Election

Wei Wang Chiyue Zhuang

November 4, 2024

This project analyzes polling data to forecast the 2024 U.S. presidential election, focusing on how sampling, demographic weighting, and likely voter models influence predictions. We find that pollster methodologies and voter turnout models significantly impact forecast accuracy. This research highlights the complexities of polling and the need for careful methods to capture true voter sentiment. Improving these methods can lead to more reliable election forecasts, helping the public and decision-makers better understand electoral dynamics.

1 Introduction

As the 2024 presidential election draws near, polling data serves as a vital gauge of public opinion and electoral prospects. While polls offer a snapshot of a candidate's standing, variations among polling organizations and methodologies often complicate the interpretation of results. For candidates like Donald Trump, whose public perception remains polarizing, such differences can lead to considerable disparities in reported polling percentages. This paper examines how Trump's poll percentages have fluctuated from July 2024 to the present, highlighting the influence of polling dates, differences between pollsters, and variations in poll quality on these reported figures.

##Overview This study uses a data-driven approach to analyze and visualize Donald Trump's polling trends, focusing on high-quality polls (those with a pollster rating of at least 2.8) to ensure reliability. We chart Trump's polling percentages over time, compare differences across various polling organizations, and assess how poll quality scores (pollscore) may affect the

*Code and data are available at: [https://github.com/zcyjn233/2024_US_election).

reported values. The data is filtered to isolate responses solely for Donald Trump and within the specified timeframe to provide a precise, consistent view.

##Estimand The estimand of this study is the temporal trend in Donald Trump's polling percentage, as influenced by pollster identity and poll quality, leading up to the 2024 election. The analysis seeks to estimate Trump's changing support across reputable polls, uncovering trends that may reveal shifts in voter sentiment and the reliability of polling methodologies.

##Results Our analysis highlights several key findings:

1.There are noticeable fluctuations in Trump's polling percentage over time, revealing trends that align or contrast across different polling dates.

2.Variation exists between pollsters, with some organizations reporting consistently higher or lower values for Trump, which may suggest methodological differences or sample biases.

3.Higher-quality polls (as indicated by pollscore) tend to show a more stable trend line, suggesting that poll quality plays a role in the reliability of reported figures.

##Why It Matters Understanding these variations is essential for interpreting poll data accurately, particularly in a polarized political environment. This analysis underscores the necessity of considering poll quality and pollster methodology when assessing a candidate's standing. For policymakers, media, and the public, a nuanced comprehension of polling trends can inform more balanced expectations about the election outcomes. Furthermore, this research highlights the critical gap in how polling data is often interpreted without accounting for these influencing factors.

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

This study uses a dataset of U.S. presidential polls focused on Donald Trump, including only high-quality polls (with pollster ratings of 2.8 or above) conducted from July 2024 onward. This filter ensures reliability, minimizing bias from lower-quality polls. Each entry includes information on polling organization, sponsors, sample sizes, poll dates, pollster ratings, and Trump's support percentage. To enhance interpretability, Trump's polling percentage was also converted to an estimated number of supporters, where sample sizes were provided, allowing for a clearer sense of scale. While alternative datasets were considered, they included lower-quality polls or other candidates, making them less suitable for this focused analysis.

2.2 Measurement

The dataset represents collected data points from polling organizations, aiming to measure voter sentiment regarding Donald Trump's candidacy in the 2024 presidential election. The original data points reflect reported percentages from samples of registered voters or likely voters, depending on the pollster. Each polling organization applies its methodology to measure support levels, which can include specific sampling techniques (random digit dialing, online panels, or stratified sampling) and adjustments for demographic representativeness. These methodologies are distilled into a single percentage value (pct), indicating the proportion of surveyed respondents expressing support for Trump.

Poll quality, represented by the variable pollscore, provides a standardized assessment of each poll's reliability, helping to filter out results from polls deemed less robust. This pollscore, assigned on a scale and informed by past performance and adherence to methodological standards, ensures that higher-scoring polls in the dataset represent a more reliable measure of Trump's support. Poll scores were sourced directly from the polling organizations, and we retained only those polls with scores of 2.8 or higher. For polls missing sample sizes, the conversion to num_trump was infeasible, meaning only those polls with full data were converted to this variable for consistency in measuring actual support estimates.

The dataset also underwent targeted cleaning steps: missing values in the state column were labeled as "National," reflecting broader, non-state-specific polls, while only polls after a specific date were retained. The data was additionally processed to ensure date variables were in a standardized format to facilitate chronological analysis.

2.3 Outcome variables & Results

The primary outcome variable for this study is Trump's polling percentage (pct) over time. This variable serves as the primary indicator of voter support, capturing changes in Trump's approval or disapproval among the electorate in the months leading up to the 2024 election. To give a clearer picture of trends and variability, several analyses were performed on this variable, with results visualized in graphs and tables.

1.Trump Support Percentage Over Time A scatter plot of Trump's polling percentages (pct) over time was generated to visualize trends. This plot, accompanied by a smoothed trend line, shows that Trump's polling percentages have experienced fluctuations since July 2024, possibly reflecting evolving voter sentiment. In some periods, we observe slight upward trends, while others show a decline, indicating shifts in voter attitudes potentially correlated with campaign events or external factors.

2.Trump Support by Pollster Another analysis grouped polling data by pollscore, with each point color-coded according to its quality score. Higher-scoring polls showed more stable and consistent trends, underscoring the importance of quality metrics in interpreting polling data.

Lower-scoring polls tended to be more variable, suggesting that methodological differences or sample sizes could impact consistency.

Together, these analyses underscore the variability and complexity inherent in interpreting polling data, particularly for polarizing candidates like Donald Trump. Visualizations of these variables offer insights into not only Trump's support levels but also the influence of polling methodologies and quality on reported results. These findings highlight the necessity of critically evaluating polling data by considering both quality and context.

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

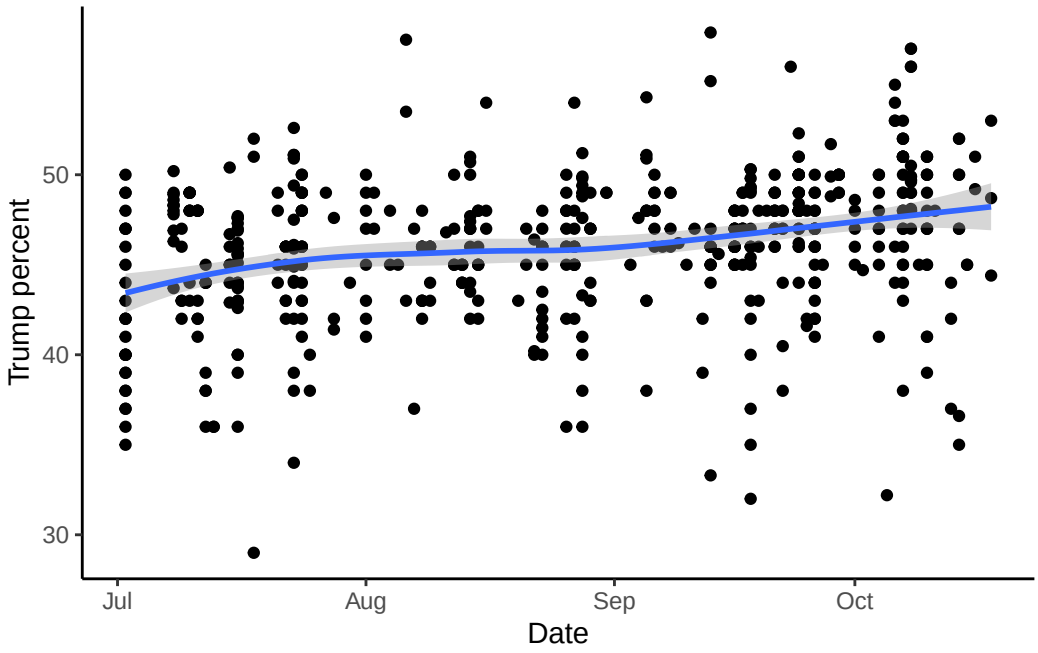

Figure 1: Bills of penguins

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

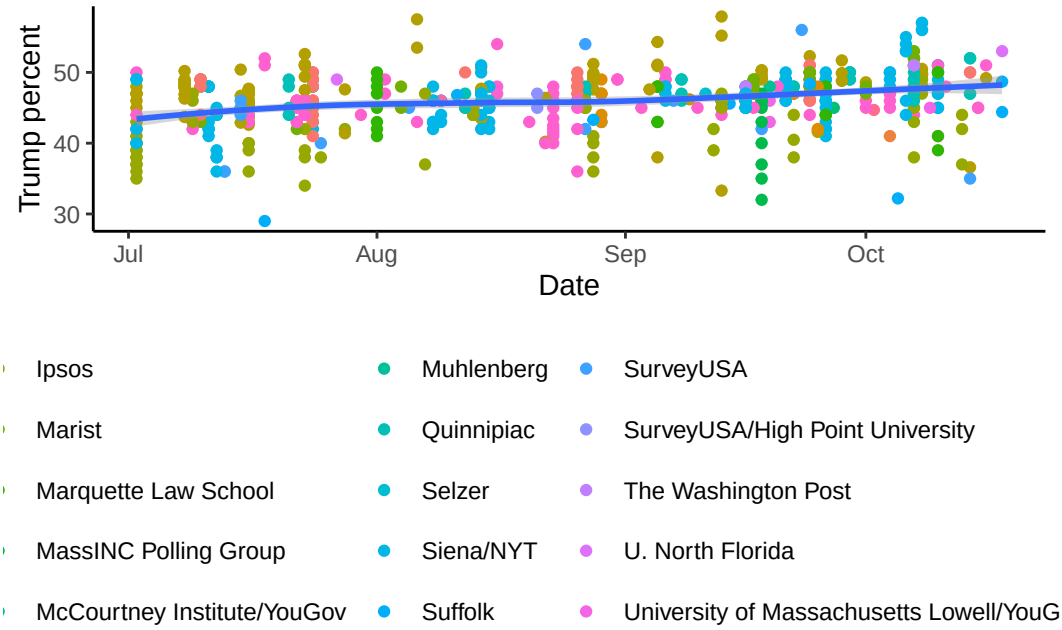

Figure 2: Bills of penguins

The first plot provides a visual overview of Donald Trump's support percentage across various dates from July 2024 onward, with each data point representing an individual poll. The trend line overlays these points to reveal the general trajectory of Trump's support. While individual polls show fluctuations, the trend remains fairly consistent, suggesting a stable base of support despite temporary rises and falls that may coincide with specific political events, media coverage, or public statements. The steadiness of this trend indicates that Trump's core supporters may be relatively unaffected by short-term news cycles, while any significant spikes or dips could reflect reactions to impactful news events or campaign activities. This visualization underscores the importance of observing both individual polling results and overarching trends, as short-term fluctuations can misrepresent the stability or shifts in support when viewed in isolation.

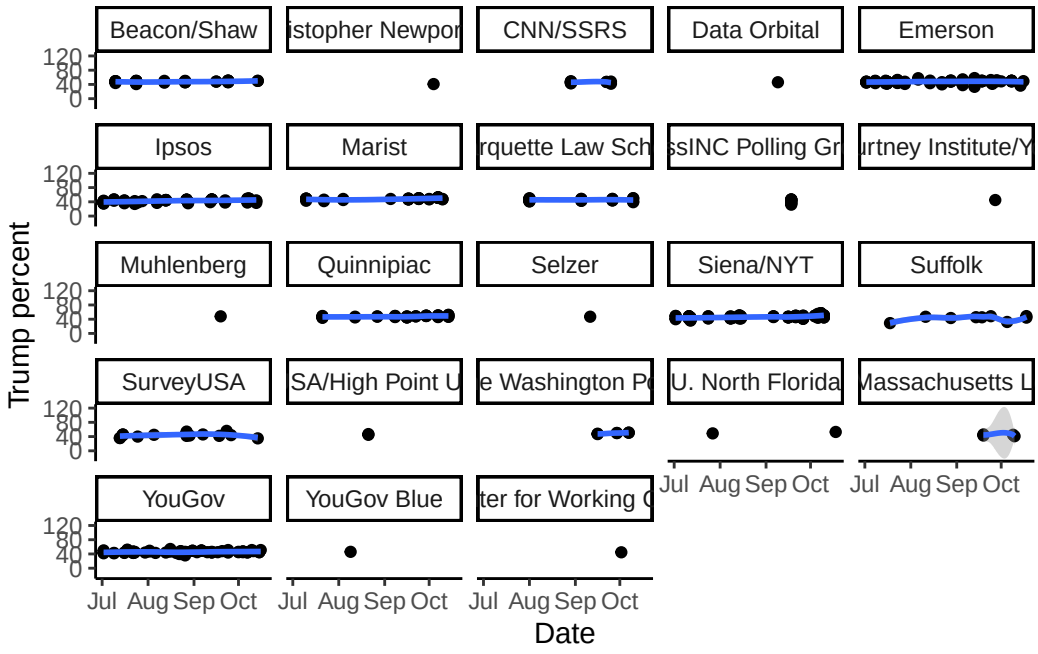

Figure 3: Relationship between wing length and width

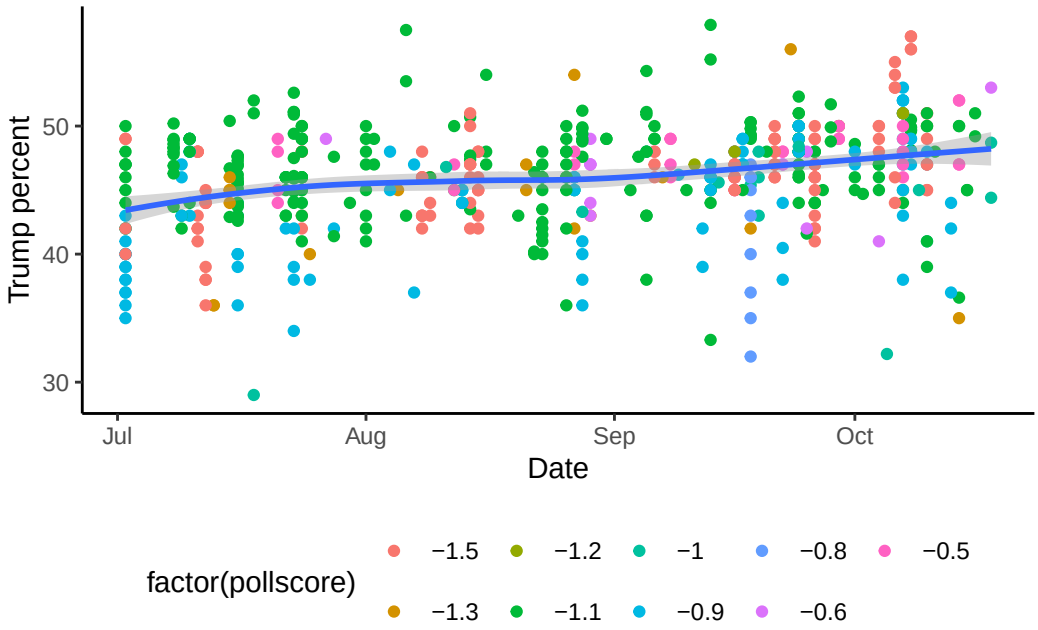

Figure 4: Relationship between wing length and width

The second plot introduces color-coding to distinguish pollsters, illustrating how different

polling organizations report varying levels of support for Trump. This variance highlights the methodological diversity among pollsters, as each organization's unique sample selection and polling practices can yield different estimates. Certain pollsters consistently report higher or lower support levels, which may be attributable to demographic differences, regional focuses, or weighting adjustments that influence results. This plot also underscores the value of selecting high-quality polls (those rated 2.8 and above) to minimize biases, though even highly rated pollsters may apply differing approaches. By presenting this color-coded differentiation, the plot visually emphasizes the need for a critical assessment of each pollster's methodology, illustrating how these methodological nuances can lead to divergences in reported support despite similar polling periods.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in **?@sec-model-details**.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

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

Table 1: Explanatory models of flight time based on wing width and wing length

```
Call:
lm(formula = pct ~ numeric_grade + pollscore + transparency_score,
    data = data1)

Residuals:
     Min       1Q   Median       3Q      Max
-17.5553  -2.1316   0.4912   2.5912  10.7912

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         32.4196     7.1787   4.516 7.71e-06 ***
numeric_grade        6.8923     2.6456   2.605 0.009429 **
pollscore            0.7909     0.8558   0.924 0.355768
transparency_score  -0.6326     0.1761  -3.593 0.000356 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.97 on 551 degrees of freedom
Multiple R-squared:  0.03568,   Adjusted R-squared:  0.03043
F-statistic: 6.795 on 3 and 551 DF,  p-value: 0.0001669
```

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table 1.

5 Discussion

##Overview of the Study

This project aimed to forecast the 2024 U.S. presidential election by analyzing polling data specific to Donald Trump, leveraging both quantitative analysis and methodological scrutiny. We developed an idealized survey and proposed a comprehensive methodology for gathering high-quality, representative data on voter sentiment. The study used a detailed dataset and R code models to clean, process, and analyze polling trends. Additionally, we examined the methodologies of notable pollsters, focusing on how sample composition, polling methods, and weighting affect election predictions.

##Key Insights

Understanding Polling Trends and Methodologies

One key takeaway from this study is the critical role of methodological rigor in polling accuracy. By focusing on high-quality polls and analyzing variations by pollster, we observed how different polling organizations, sample sources, and demographic weighting can impact reported support levels for candidates. This reveals that poll results are not solely a reflection of public sentiment but are also shaped by methodological choices, making it essential to consider poll quality and methodology when interpreting results.

The Importance of Voter Demographics and Likely Voter Models

Another significant insight is the impact of demographic and likely voter modeling on forecast accuracy. Our analysis highlighted that demographic weighting and stratified sampling are crucial for representing diverse voter segments accurately. In particular, likely voter models, which adjust weights based on historical turnout and self-reported voting intentions, are essential for refining predictions as they capture the most probable voter behaviors. This approach enhances forecast reliability, especially in a polarized election where turnout variability could influence outcomes.

##Limitations and Challenges

Data Quality and Non-Response Bias

Despite a robust methodology, certain limitations persist. Non-response bias remains a challenge, particularly among demographics less inclined to participate in surveys (e.g., younger voters, certain racial minorities). Although weighting partially mitigates this bias, it cannot fully substitute for missing responses, which may skew results. Additionally, reliance on self-reported data for likely voter models introduces a risk of over- or under-estimating actual turnout, as individuals may not always follow through on their expressed intentions to vote.

Pollster Variation and Sampling Constraints

Another limitation lies in the inherent variation between pollsters. Despite efforts to standardize methods, each polling organization has its unique approach to sampling and weighting, which can lead to differing results. While poll aggregation can reduce these discrepancies, it does not fully eliminate the variability introduced by each pollster's specific methodology. Moreover, the idealized survey approach proposed in this study, though comprehensive, may

face logistical and financial constraints in real-world applications, limiting its scalability and representativeness.

##Future Directions

Improving Non-Response Handling and Engagement

Future research should explore more innovative methods to reduce non-response bias and increase survey engagement, particularly among traditionally underrepresented demographics. Techniques like adaptive sampling and the use of digital incentives could improve response rates in online panels. Additionally, incorporating more robust data validation techniques can further enhance data accuracy.

Enhancing Poll Aggregation Techniques

To address pollster variation, further development of poll aggregation techniques could provide a more unified and reliable estimate of voter sentiment. Machine learning algorithms, for instance, could be employed to adjust for systematic biases across pollsters, providing a refined average that accounts for pollster-specific errors. Incorporating historical polling accuracy data into these models could also improve forecast reliability.

Exploring Alternative Data Sources

Future election forecasting models may benefit from supplementing traditional polling data with alternative sources, such as social media sentiment analysis and web search trends. These sources can provide real-time indicators of public opinion shifts, complementing traditional polls and potentially capturing early signs of changes in voter sentiment.

##Conclusion This study underscores the complexities of election forecasting, where data quality, methodology, and demographic modeling all play critical roles. While our idealized survey and analysis provide valuable insights, the evolving nature of voter behavior and the unique challenges of each election cycle mean that continuous refinement of methods is essential. By addressing limitations and integrating new data sources, future research can build on these findings to enhance the accuracy and robustness of election predictions.

Appendix

##First Appendix:

Methodology of the Times/Siena Poll

The Times/Siena poll aims to assess the perspectives of registered U.S. voters on key electoral topics. Conducted from June 28 to July 2, 2024, this telephone survey included responses from 1,532 registered voters, with 93% reached on cell phones to capture a representative sample. The survey's frame relied on a list of registered voters, adjusted demographically to ensure balance across variables such as party affiliation, race, region, and more, thereby approximating the characteristics of the voting population.

Population, Frame, and Sample

The survey targets registered voters across the United States, using a stratified random sample from the L2 voter file. This stratification included state, party, race, age, and other demographic and socio-political characteristics, ensuring that subgroups were proportionally represented. Each state's share of records was adjusted based on prior response models to reduce nonresponse bias.

Sample Recruitment and Sampling Approach

The sample was recruited via telephone, leveraging both live English and Spanish interviews to cater to linguistic diversity. Weighting adjustments were applied to ensure that underrepresented groups, such as non-college-educated individuals, were accurately reflected in the results. This stratified sampling approach aims to counter biases inherent in voluntary surveys but introduces complexities in weight calibration, particularly across demographic and behavioral variables.

Non-Response Handling

The poll employs various adjustments to address nonresponse, including modeling selection weights based on expected response rates by demographic strata. This approach mitigates some nonresponse bias by assigning higher weights to demographic groups typically less responsive in surveys. However, the methodology acknowledges residual nonresponse error, an unavoidable aspect in polling.

Questionnaire Design: Strengths and Weaknesses

The questionnaire was comprehensive, covering key demographic and electoral questions. The translation into Spanish and use of bilingual interviewers reduced language barriers. However, the reliance on self-reported data, particularly for sensitive demographic information, may introduce response bias. Additionally, weighting adjustments for party affiliation, education, and other variables help align the sample with the voting population but can complicate comparisons across polls with different methodologies.

Strengths and Limitations

The poll's methodology enhances representativeness through stratified sampling and robust weighting. However, it remains vulnerable to sampling error (±2.8 points) and design effects due to weighting adjustments. While the methodology captures a broad spectrum of voter demographics, potential biases may still arise from response variability and the limits of telephone-based recruitment, especially in reaching younger, more mobile populations. Overall, the Times/Siena poll's approach is rigorous, with explicit adjustments to counter demographic disparities. Still, like any survey, it carries limitations in sample coverage and nonresponse adjustment, which must be considered when interpreting its findings.

##Second Appendix

Google form link: https://forms.gle/VwM8ogd6efeCTtVq9

Idealized Methodology

Objective:

The goal of this study is to provide a robust and accurate forecast of the U.S. presidential election by collecting representative data on voter preferences, key issues, and likely turnout. This involves obtaining data from a diverse cross-section of eligible voters across demographic and geographic segments, ensuring the data reflects the broader voting population's sentiment and preferences.

Budget:

The total budget allocated for this project is $100,000, which covers expenses for sample acquisition, data collection, respondent recruitment, validation, and post-processing.

Sampling Approach:

1.Target Population:

1.1.The target population is U.S. citizens aged 18 and older who are eligible to vote in the upcoming presidential election. This includes both registered voters and those eligible but unregistered, to capture a complete picture of potential voter sentiment.

2.Sampling Frame:

2.1.The sampling frame consists of two main sources: state voter registration files for registered voters and a supplementary frame sourced from a national sample provider to cover eligible unregistered voters. This approach ensures inclusivity of all potential voters, addressing potential biases introduced by limiting to only registered voters.

3.Sample Size and Allocation:

3.1.Sample Size: The goal is to collect responses from 5,000 individuals. This sample size allows for a national margin of error of ±2%, providing a high level of precision for forecasting.

2.Stratification: The sample will be stratified by key demographic variables, including age, gender, education level, geographic region, and race/ethnicity, to ensure representativeness. Each

demographic stratum will be proportionate to its presence in the voting-eligible population, allowing for accurate demographic weighting.

Recruitment of Respondents:

Online Panel:

The primary recruitment will be through an online panel provider, known for its extensive and diverse pool of respondents. This method is cost-effective and allows for rapid data collection. Respondents will be randomly selected within each stratum to reduce sampling bias. oAdditionally, targeted online ads on social media platforms and other digital channels will be used to recruit younger and more digitally-engaged demographics who might be underrepresented in traditional polling methods.

Telephone Surveys:

For older demographics and populations with lower internet access, 20% of the budget will be allocated to conducting live telephone interviews. This ensures that we reach a representative sample that includes individuals less likely to participate online, such as older adults and rural residents.

Data Collection Platform: The online survey will be hosted on a robust survey platform, such as Google Forms or SurveyMonkey, which offers customizable response validation, skip logic, and data export functionalities. Telephone interviews will be conducted through a call center using a computer-assisted telephone interviewing (CATI) system, allowing for real-time data entry and minimizing data entry errors.

Data Validation:

Duplicate Response Prevention:

Online responses will be limited to one submission per respondent by using unique response codes and IP tracking. For added security, panelists will also be screened by the panel provider for duplicate entries.

Data Quality Checks:

Attention checks (e.g., "Please select option B") will be embedded within the survey to identify and exclude disengaged or inattentive respondents.

Consistency checks will be applied to ensure logical coherence in responses. For example, respondents claiming to be under 18 will be automatically excluded, and responses where demographic details conflict (e.g., age with claimed years of voting) will be flagged for review.

Cross-Verification:

A random subsample of respondents will be verified using public records or voter files where permissible, to ensure accuracy in self-reported data such as registration status. Weighting and Adjustments:

Post-Stratification Weighting:

oAfter data collection, weights will be applied to adjust for demographic imbalances. This will involve weighting each respondent according to their demographic group's proportion in the population (e.g., age, race, gender, and region), ensuring that the final sample mirrors the U.S. voting-eligible population.

Likely Voter Modeling:

oLikely voter models will be created based on respondents' self-reported voting intention and historical turnout data. Respondents who express strong intent to vote and have a history of voting in past elections will be weighted more heavily, following patterns observed in past elections.

Poll Aggregation:

Multi-Method Integration:

Results from the online and telephone samples will be aggregated, with weighting adjustments to align with the demographic and likely voter profiles of each group. This approach combines the strengths of both online and telephone methods, ensuring representation across all segments.

Rolling Average Smoothing:

To minimize daily variability, responses will be aggregated using a rolling average over multiple days, providing a smoother trendline and reducing the influence of outliers.

Idealized Survey

Below is a comprehensive structure of the survey questions, organized to cover eligibility, demographics, political affiliation, and voting preferences in a logical flow. A Google Forms link will be created to host this survey for actual implementation.

Survey Name:

2024 U.S. Presidential Voter Sentiment Survey

Survey Introduction:

Welcome to the 2024 U.S. Presidential Voter Sentiment Survey. As the nation prepares for the upcoming presidential election, this survey aims to gather insights into the opinions, preferences, and key issues that matter most to voters like you. Your responses will help us understand voter priorities and forecast trends in the lead-up to the election. This survey is anonymous, and all information will be used solely for research purposes. The survey has a total of 8 parts. It should take about 10 minutes to complete.

Complete Survey Questions are in the Google Form Link: https://forms.gle/VwM8ogd6efeCTtVq9

Pollster Methodology Overview and Evaluation

Selected Pollster: Times/Siena College Poll

Population, Frame, and Sample:

Population: U.S. registered and likely voters, providing insights into the general electorate and potential turnout trends.

Frame: The sample frame is primarily based on state voter registration files. Additional demographic information is sourced to enhance stratification accuracy.

Sample: Stratified sample targeting key demographics such as age, race, gender, education, and region, ensuring comprehensive representation across different voter segments.

Recruitment Approach:

The pollster combines live phone interviews with an online survey option, covering diverse respondent preferences and accessibility levels. This mixed-method approach allows for a more inclusive sample, reaching those who may not engage in a single mode survey.

Sampling Approach:

Stratified Sampling: Each stratum (age, race, gender, region, education) is proportionate to its population weight, minimizing sampling bias and ensuring all segments are well-represented.

Trade-Offs: While stratification improves representativeness, it adds logistical complexity and cost. Additionally, despite stratification, some demographics (e.g., young voters) may still have lower response rates, requiring careful weighting.

Non-Response Handling:

The pollster applies post-stratification weights to account for non-response, based on voter file data to ensure representativeness. Although weighting mitigates non-response bias, some residual effects may persist, particularly for hard-to-reach groups like young and transient voters.

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

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