Forecasting the 2024 U.S. Presidential Election: A Poll-Based Approach*

Kamala Harris Projected to Win the Electoral College with 292 Votes to Donald Trump's 246

Xinxiang Gao   Ariel Xing   John Zhang†

November 4, 2024

The upcoming U.S. Presidential Election on November 5th features a closely contested race between Vice President Kamala Harris and former President Donald Trump. This paper uses a poll-of-polls approach, combined with a dynamic linear regression model that updates daily with the latest data, to forecast the popular vote and outcomes in seven battleground states: Arizona, Georgia, Nevada, North Carolina, Michigan, Wisconsin, and Pennsylvania. Our analysis currently projects that Vice President Harris will win the popular vote, 48.19% to 46.01%, and secure the electoral college with 292 votes to Trump's 246, by winning five of these seven swing states: Nevada, North Carolina, Michigan, Wisconsin, and Pennsylvania. By aggregating multiple polls rather than relying on a single poll, our model provides a more robust prediction, supporting the likelihood of a Harris victory.

Table of contents

*Code and data are available at: https://github.com/xgao28/election_forecast.

†The authors are listed in alphabetical order by last name.

1 Introduction

The race between Vice President Kamala Harris and former President Donald Trump for the 2024 presidential election will conclude on November 5, 2024. President Joe Biden's decision on July 21, 2024, to withdraw from this year's race led to Vice President Kamala Harris becoming the official Democratic nominee, supported by Biden and confirmed through a virtual roll call on August 2. Even before Biden's withdrawal, polling organizations had gauged public opinion on a potential matchup between Harris and former President Donald Trump. Early polls showed Harris trailing Trump, but her support increased following Biden's exit, making the outcome of this election uncertain.This paper employs a poll-of-polls approach and a dynamic linear model to project each candidate's likely support on election day, both nationally and within swing states.

The primary estimand in this study is the predicted percentage of support each candidate will receive on election day, both nationally and within critical swing states. This percentage serves as the foundation for forecasting which candidate is likely to win the election, by capturing trends in public sentiment over time and projecting how these trends may manifest in the final vote count.

We apply a dynamic linear regression model named " Baseline Model" to provide a national overview and another dynamic linear regression model named "Primary Model" to assess state-level dynamics, especially in key battlegrounds. The Baseline Model aggregates national polling data to estimate popular vote share, while the Primary Model offers a state-specific analysis aligned with the electoral college structure. This approach addresses a gap in existing models that often overlook the unique impact of swing states on electoral outcomes. Our model projects that Vice President Harris will secure 48.19% of the popular vote, compared to 46.01% for Trump. However, as U.S. elections are decided by the electoral college rather than the popular vote, a candidate can win the presidency without a majority of popular support. There are seven swing states, Arizona, Georgia, Nevada, North Carolina, Michigan, Wisconsin, and Pennsylvania, holding a collective 93 electoral votes, are determining factors to the election's outcome. Excluding swing states, Harris would receive 226 electoral votes, while Trump would receive 219. According to our Primary Model, Harris is expected to capture five of the seven swing states: Nevada, North Carolina, Michigan, Wisconsin, and Pennsylvania, yielding a total of 292 electoral votes and thus securing the presidency. Trump would win two swing states, Arizona and Georgia, ending with 246 electoral votes. Although Harris's lead in North Carolina is narrow (by just 0.036%), which may switch to voting for Trump in the future, but 19 votes in this state in the face of overwhelming vote margins won't have much of an impact on the outcome.

Predicting the next U.S. president holds significant implications for the nation's future and its role on the global stage. The elected leader will influence domestic policy on economic growth, healthcare, and social issues, while also shaping international relations, defense strategies, and trade policies. In an era of geopolitical uncertainty, including conflicts and economic fluctuations, understanding potential election outcomes can help inform the public, policymakers, and global leaders, providing a foundation for navigating the challenges ahead.

The structure of this paper is as follows: Section 2 provides an overview of the dataset, detailing the polling data sources, variables, and preprocessing steps. Section 3 outlines the modeling approach, presenting the rationale behind the Baseline and Primary Models and discussing their respective contributions to the analysis. Section 4 presents the results, including national and state-level predictions, followed by Section 5, which explores the implications of these findings, discusses limitations, and suggests future research directions. The three appendices provide a detailed exploration of key aspects relevant to the 2024 U.S. Presidential Election. Section A explains the distinction between swing states and solid states, highlighting their importance in election outcomes. Section B and Section C delve into the methodology and design of the Economist/YouGov poll and an election forecasting survey, outlining sampling methods, data validation processes, and strategic considerations for accurate and reliable polling results.

2 Data

2.1 Dataset Overview

The dataset by FiveThirtyEight (2024) comprises 16,867 rows and 52 columns, offering a detailed representation of polling data relevant to electoral analysis. To derive the percentage of support for each choice from the pollster, the raw vote counts associated with each candidate or option are normalized by dividing them by the total sample size for that particular poll. This calculation converts the raw counts into percentages (pct), allowing for a clearer comparison of candidate support across different polls and facilitating more straightforward interpretations of public sentiment regarding the electoral options presented.

This study leverages a dataset of polling data for the 2024 U.S. Presidential Election, focusing on predicting support for the two main candidates, Kamala Harris and Donald Trump. Each entry in the dataset represents results from individual polls, capturing both the timing and geographical scope of polling, which is crucial for analyzing national trends and state-specific shifts, particularly in key swing states. By collating data from multiple polling sources, the dataset captures a broad snapshot of voter sentiment across the country, reflecting variations across time, poll methodologies, and geographic regions.

The outcome variable, **Polling Percentage ("pct")**, represents the percentage of respondents in each poll who indicate support for either Harris or Trump. This variable is central to our analysis, as it reflects the changing landscape of voter sentiment and serves as the basis for our forecasts.

2.2 Predictor and Additional Variables

- **Days Toward Election ("days_towards_election")**: The primary predictor in the model, this variable indicates the number of days remaining until the election at the time each poll was conducted. It is essential for capturing time-based trends, enabling the model to adjust for shifts in public opinion as the election approaches. As the election nears, this variable helps illustrate the patterns of support stabilization, often observed closer to election day.

Additional potential predictors include variables reflecting the quality, sample characteristics, and transparency of each poll, which help contextualize the primary predictor and provide an understanding of polling reliability:

- **Poll Quality Score ("pollscore")**: This variable captures the methodological reliability of each poll, including factors such as sampling technique and historical accuracy. Polls with higher scores typically offer more robust estimates, as they reflect the pollster's performance and adherence to reliable sampling practices.
- **Transparency Score ("transparency_score")**: This score measures the level of methodological detail each poll discloses. Higher transparency generally correlates with increased confidence in the data, as it indicates comprehensive disclosure of methods, sample details, and collection practices.
- **Sample Size ("sample_size")**: The number of respondents in each poll is crucial for assessing precision. Larger sample sizes tend to reduce the margin of error, making the estimates more representative of the general population.
- **State ("state")**: This geographic variable distinguishes between national and state-level polls, allowing the model to capture regional variations in voter support. State-specific polling data are particularly valuable for modeling outcomes in swing states, where election outcomes often hinge on narrow margins.

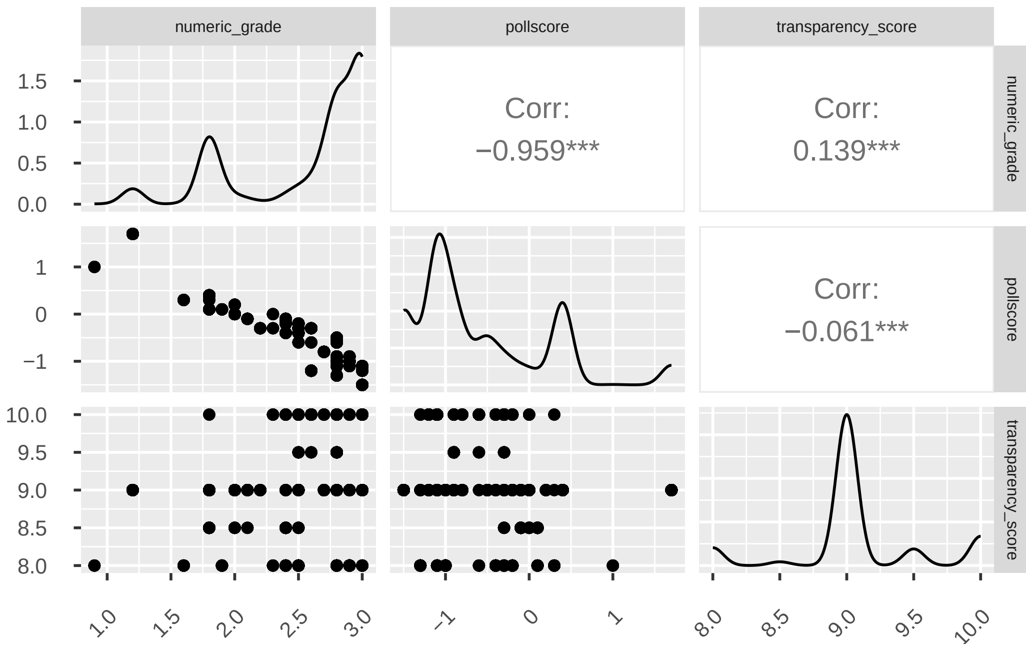
2.3 Measurement and Data Processing

Each variable in the dataset is constructed to accurately reflect polling characteristics across various dimensions of time, geography, and poll quality. Poll quality and transparency scores are derived based on historical pollster performance and disclosure levels, offering an indirect measure of data reliability. The primary predictor, **days toward election**, and the outcome variable, **pct**, are drawn directly from each poll entry. Polling percentages are scaled by sample size, which provides a weighted measure of candidate support, ensuring that polls with larger sample sizes have a proportionally greater impact on model predictions.

To enhance interpretability and prediction accuracy, we adjusted the dataset by scaling polling percentages to reflect each poll's sample size, multiplying average support by sample size, and normalizing by 0.01. This process ensures that larger samples contribute more meaningfully to the model's overall prediction, thus offering a balanced representation of likely voting intentions.

2.4 Exploratory Data Analysis and Summary Statistics

2.4.1 Multicollinearity Considerations



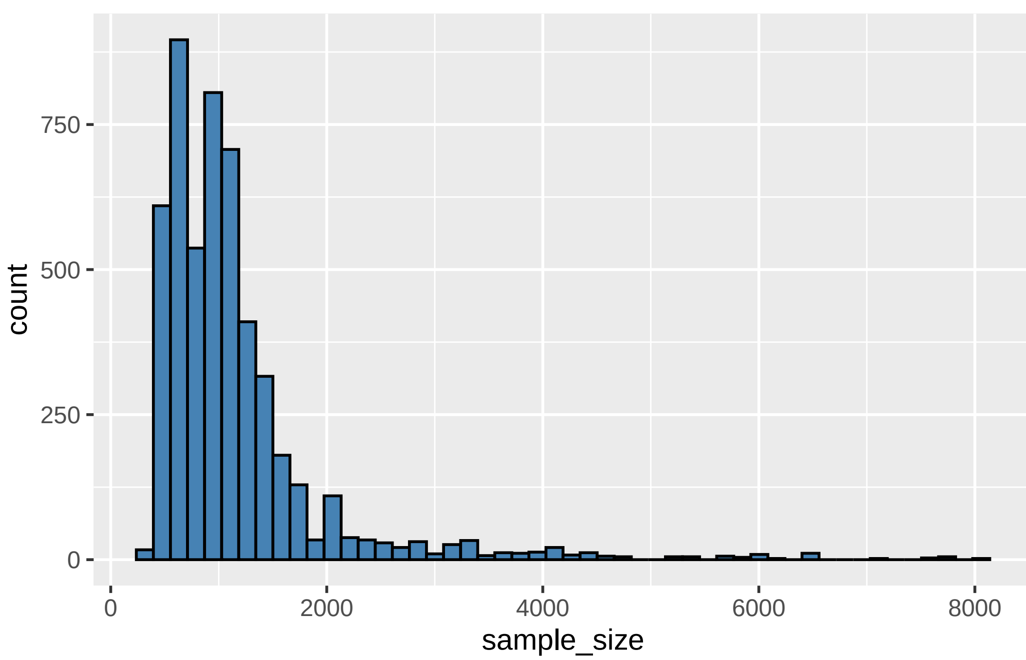
While selecting the predictors for the model, multicollinearity was a key concern. Multicollinearity occurs when two or more predictors are highly correlated, which can inflate the variance of the coefficient estimates and make the model less reliable. In our pair plot analysis, we observed a high correlation between "numeric_grade" and "pollscore," suggesting that they measure similar aspects of polling quality.

```
data_swing <- data %>% filter(state %in% c("Nevada", "Arizona", "Wisconsin", "Michigan", "Pe

data_swing %>%
  group_by(state) %>%
  summarise(count = n(),
            `mean pollscore` = round(mean(pollscore, na.rm = TRUE), 3),
            `mean numeric grade` = round(mean(numeric_grade, na.rm = TRUE), 3),
            `mean transparency score` = round(mean(transparency_score, na.rm = TRUE), 3),
            `mean sample size` = round(mean(sample_size, na.rm = TRUE))) %>%
  kable()
```

state	count	mean pollscore	mean numeric grade	mean transparency score	mean sample size
Arizona	265	-0.649	2.523	9.057	771
Georgia	279	-0.548	2.503	9.032	898
Michigan	318	-0.599	2.566	9.003	789
Nevada	156	-0.658	2.512	9.071	664
North Carolina	273	-0.413	2.401	8.908	836
Pennsylvania	422	-0.684	2.605	8.921	959
Wisconsin	383	-0.857	2.772	9.324	780

```
data %>%
  filter(sample_size < 10000) %>%
  ggplot(aes(x = sample_size)) +
  geom_histogram(bins = 50, fill = "steelblue", color = "black")
```




```
data %>%
  filter(sample_size >= 10000) %>%
  arrange(sample_size)
```

```
# A tibble: 8 x 11
  poll_id transparency_score numeric_grade pollscore state sample_size end_date
    <dbl>              <dbl>         <dbl>     <dbl> <chr>       <dbl> <chr>
1   88104                  9             2       0.2 <NA>        18123 9/4/24
2   88104                  9             2       0.2 <NA>        18123 9/4/24
3   88104                  9             2       0.2 <NA>        20762 9/4/24
4   88104                  9             2       0.2 <NA>        20762 9/4/24
5   88989                 10             3      -1.1 <NA>        48732 10/25/24
6   88989                 10             3      -1.1 <NA>        48732 10/25/24
7   88989                 10             3      -1.1 <NA>        78247 10/25/24
8   88989                 10             3      -1.1 <NA>        78247 10/25/24
# i 4 more variables: party <chr>, candidate_name <chr>,
#   days_towards_election <dbl>, pct <dbl>
```

To understand the general trends in the data, we conducted exploratory data analysis using summary statistics and visualizations. Figure 1 (below) illustrates the polling percentages for each candidate over time, allowing us to observe fluctuations in support as the election day draws near.

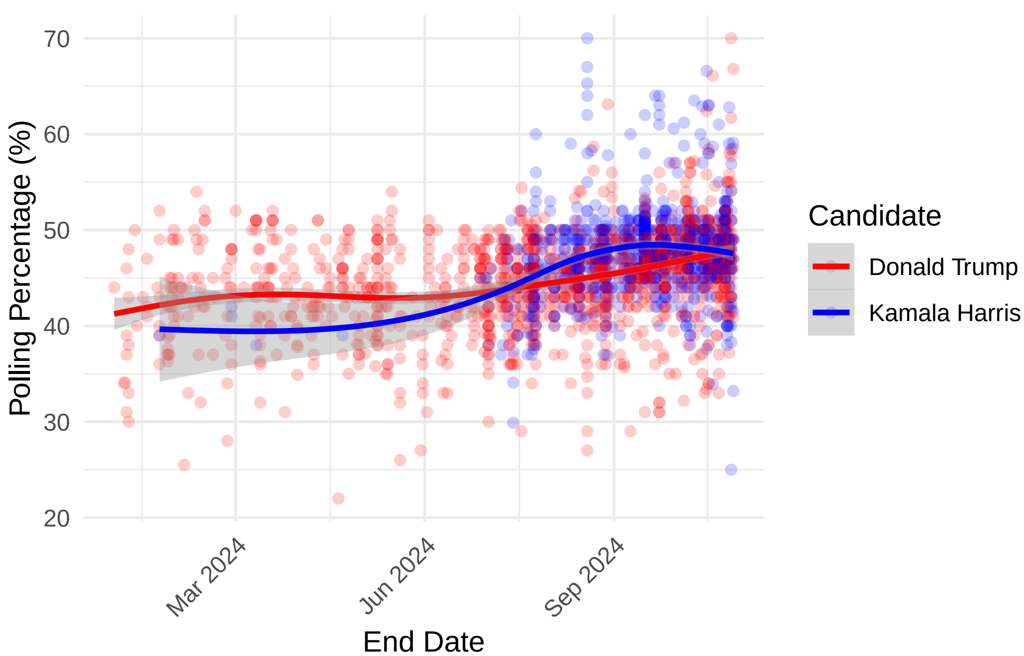


Figure 1: Polling Percentage over Time by Candidate

3 Model

In this study, we employ two main models to forecast the outcome of the 2024 U.S. Presidential Election using polling data for the candidates Kamala Harris and Donald Trump. The models, a Baseline Model (Winner-Take-All National Model) and a Swing State-Based Model, provide two distinct but complementary approaches to assessing candidate support. The Baseline Model aggregates national polling data to project which candidate is likely to win based on national polling trends and assigns all electoral votes to the candidate with the higher projected national percentage on election day. This approach offers a straightforward prediction of overall voter sentiment across the country and captures the aggregate trend of support for each candidate as the election approaches.

On the other hand, the Swing State-Based Model is designed to provide a more granular projection by focusing on polling data at the state level, particularly for key swing states. This model assumes that the winner in each swing state can be determined by state-specific polling trends, and it projects support for each candidate within these competitive regions. By combining outcomes from both swing states and solid states, this model allows for a more nuanced prediction that aligns with the actual mechanics of the Electoral College. Together, these models are well-suited to address both the general national polling trends and the crucial role of swing states in determining the final electoral outcome.

Table 2: Baseline Model

term	estimate	std.error	t.stat	p.value	candidate
(Intercept)	48.189	0.207	232.882	0	Kamala Harris
days_towards_election	-0.012	0.002	-5.119	0	Kamala Harris
(Intercept)	46.010	0.177	260.055	0	Donald Trump
days_towards_election	-0.009	0.001	-10.235	0	Donald Trump

3.1 Baseline Model (Winner-Take-All National Model)

In our baseline model, we predict the election winner by using national polling data and assigning all electoral votes to the candidate with the higher projected percentage on election day. This "winner-take-all" approach assumes that the candidate leading in national polling would theoretically secure all electoral votes.

To build this model, we define each candidate's polling trend over time using simple linear regression:

- Let $P_H(t)$ represent Kamala Harris's national polling percentage at t days before the election, and $P_T(t)$ represent Donald Trump's percentage at the same time.

We express each candidate's polling percentage as a linear function of time:

$$P_H(t) = \alpha_H + \beta_H t$$

$$P_T(t) = \alpha_T + \beta_T t$$

where: - α_H and α_T are the intercepts, representing Harris's and Trump's predicted polling percentages on election day (when $t = 0$). - β_H and β_T are the slopes, capturing the rate of change in each candidate's polling percentage as election day approaches.

Our primary interest is in estimating these intercepts, α_H and α_T, since they give the predicted polling percentages for each candidate on election day. To predict the winner, we compare these intercepts:

- If $\alpha_H > \alpha_T$, we predict that Kamala Harris wins the election and receives all electoral votes.
- Conversely, if $\alpha_T > \alpha_H$, Donald Trump is predicted to win all electoral votes.

This model simplifies the electoral system by assuming that the national polling leader would win the overall election, without accounting for individual state results.

Table 3: Primary Model

term	estimate	std.error	t.stat	p.value	state	candidate
(Intercept)	48.798	0.535	91.297	0.000	Arizona	Donald Trump
days_towards_election	-0.018	0.003	-5.298	0.000	Arizona	Donald Trump
(Intercept)	47.097	0.436	108.050	0.000	Arizona	Kamala Harris
days_towards_election	-0.011	0.005	-2.157	0.036	Arizona	Kamala Harris
(Intercept)	48.585	0.435	111.706	0.000	Georgia	Donald Trump
days_towards_election	-0.012	0.002	-5.089	0.000	Georgia	Donald Trump
(Intercept)	47.824	0.478	100.074	0.000	Georgia	Kamala Harris
days_towards_election	-0.014	0.005	-2.604	0.012	Georgia	Kamala Harris
(Intercept)	46.797	0.541	86.472	0.000	Michigan	Donald Trump

Table 3: Primary Model

term	estimate	std.error	t.stat	p.value	state	candidate
days_towards_election	-0.013	0.003	-4.271	0.000	Michigan	Donald Trump
(Intercept)	47.981	0.472	101.674	0.000	Michigan	Kamala Harris
days_towards_election	-0.010	0.005	-1.939	0.058	Michigan	Kamala Harris
(Intercept)	46.312	0.701	66.022	0.000	Nevada	Donald Trump
days_towards_election	0.002	0.004	0.377	0.708	Nevada	Donald Trump
(Intercept)	48.098	0.686	70.162	0.000	Nevada	Kamala Harris
days_towards_election	-0.020	0.006	-3.252	0.003	Nevada	Kamala Harris
(Intercept)	48.118	0.326	147.531	0.000	North Carolina	Donald Trump
days_towards_election	-0.015	0.002	-7.253	0.000	North Carolina	Donald Trump
(Intercept)	48.154	0.490	98.255	0.000	North Carolina	Kamala Harris
days_towards_election	-0.022	0.009	-2.336	0.023	North Carolina	Kamala Harris
(Intercept)	47.215	0.367	128.546	0.000	Pennsylvania	Donald Trump
days_towards_election	-0.015	0.002	-6.703	0.000	Pennsylvania	Donald Trump
(Intercept)	48.844	0.333	146.547	0.000	Pennsylvania	Kamala Harris
days_towards_election	-0.019	0.004	-4.785	0.000	Pennsylvania	Kamala Harris
(Intercept)	46.667	0.461	101.243	0.000	Wisconsin	Donald Trump
days_towards_election	-0.009	0.003	-2.935	0.004	Wisconsin	Donald Trump
(Intercept)	49.305	0.364	135.402	0.000	Wisconsin	Kamala Harris
days_towards_election	-0.010	0.004	-2.414	0.019	Wisconsin	Kamala Harris

3.2 Swing State-Based Model

This model forecasts electoral outcomes by projecting each candidate's polling results in individual swing states, then combining those results with electoral votes from states that are already solidly in favor of one candidate or the other.

1. **Swing State Polling Projections**
   To predict each candidate's chances in swing states, we model their polling trends over time for each swing state $s \in S$. Let:

   - $P_{H_s}(t)$ be Kamala Harris's polling percentage in state s at t days before the election.
   - $P_{T_s}(t)$ be Donald Trump's polling percentage in state s at t days before the election.

   We express each candidate's polling percentage in each swing state as a linear function of time:

$$P_{H_s}(t) = \alpha_{H_s} + \beta_{H_s} t$$

$$P_{T_s}(t) = \alpha_{T_s} + \beta_{T_s} t$$

   where:

   - α_{H_s} and α_{T_s} represent the intercepts, or the predicted polling percentages for Harris and Trump on election day (when $t = 0$).
   - β_{H_s} and β_{T_s} are the slopes, indicating how each candidate's polling percentage changes over time in state s.

   For each swing state, we focus on the predicted polling percentages on election day, $P_{H_s}(0) = \alpha_{H_s}$ for Harris and $P_{T_s}(0) = \alpha_{T_s}$ for Trump. We predict the winner of state s based on which candidate has the higher projected percentage: if $\alpha_{H_s} > \alpha_{T_s}$, Harris wins the state; otherwise, Trump does.

2. **Electoral Vote Aggregation**
   Once we have determined the predicted winner in each swing state, we calculate the total electoral votes for each candidate by combining:

   - Electoral votes from solid states, as defined by Electoral Ventures LLC (2024).
   - Electoral votes from the swing states predicted to support each candidate.

   In this approach, the candidate who surpasses the 270 electoral votes threshold is predicted to win the election. This model considers the variability of swing states, differentiating it from the winner-take-all national model by capturing the dynamics of individual state races.

3.3 Model justification

The two-model approach is justified by the need to capture both national and state-level dynamics in election forecasting. The Baseline Model's focus on national polling trends is appropriate for understanding the overall sentiment of the voting population, providing an aggregate view that reflects general support levels for each candidate. Since national support can indicate the broader trajectory of an election, the Winner-Take-All National Model captures this sentiment in a straightforward, interpretable way. The inclusion of a temporal component, represented by days toward the election, allows both models to track changes in support levels as the election day approaches. This time-based element is critical, as voter sentiment often fluctuates and solidifies in the lead-up to an election, especially with the influence of campaign events and media coverage.

The Primary Model, focusing on swing states, is particularly valuable because of the unique structure of the U.S. Electoral College, where specific states can disproportionately influence the final outcome. Including state-specific intercepts and slopes enables this model to account for localized variations in support, reflecting the critical importance of state-level outcomes in swing regions. By incorporating state indicators, this model can address the heterogeneity of voting behavior across different regions, capturing the specific dynamics of battleground states where candidate support may diverge significantly from national trends. Aggregating the electoral votes based on projected winners in both swing and solid states makes this model well-aligned with real-world electoral processes, enhancing its practical utility in forecasting the election outcome.

Additionally, the linear relationship assumed in these models is justified by the often gradual and linear shifts in voter preferences over short periods, especially as the election date nears. State projections are treated independently, consistent with the winner-take-all approach in most U.S. states. Although linear models are generally effective for capturing large-scale polling trends, alternative approaches, such as Bayesian methods for quantifying uncertainty, could be considered for highly volatile polling scenarios. Nonetheless, the current models' simplicity and interpretability make them particularly suitable for capturing the aggregate and state-level trends necessary for this election forecast, providing a balanced approach that aligns well with the complexities of the U.S. electoral system.

4 Results

The results from our forecasting models provide insight into predicted support levels for Kamala Harris and Donald Trump, drawing from both national and state-level polling data. Each model presents a distinct view of candidate support and projected outcomes based on trends in polling data as election day approaches.

4.1 National Polling Trends and Baseline Model Results

The Baseline Model evaluates national polling data to project which candidate would win all electoral votes based on the higher national polling percentage on election day. The intercept term, representing predicted support on election day, shows that Kamala Harris is expected to receive approximately 48.19% of the national vote, while Donald Trump is projected to receive around 46.01%. This 2.18 percentage point lead for Harris in the Baseline Model suggests a slight advantage for her on a national scale. This model illustrates overall voter sentiment trends, capturing how each candidate's support changes as election day approaches and indicating a national-level lean toward Harris if these trends hold.

Table 4: Baseline Model Result

Kamala Harris	Donald Trump	difference
48.189	46.01	2.179

4.2 Swing State Analysis and Primary Model Results

In contrast, the Primary Model delves into state-specific polling data, focusing on swing states where vote margins are especially narrow and could decisively impact the electoral outcome. Intercept estimates for each candidate across key swing states show a mixed pattern, with Harris projected to lead in states like Wisconsin (49.30% vs. 46.67%) and Pennsylvania (48.84% vs. 47.21%), where her support surpasses Trump's by 2.64 and 1.63 percentage points, respectively. However, Trump maintains a lead in Arizona (48.80% vs. 47.10%) and Georgia (48.59% vs. 47.82%), reflecting the competitive nature of these swing states.

Table 1 provides a summary of intercept estimates for both candidates across swing states, showing how close these races remain. Harris's estimated advantage in states like Wisconsin and Pennsylvania signals potential strategic wins in the electoral college, while Trump's narrow leads in Arizona and Georgia underscore the contested nature of the 2024 election in these battleground areas.

Table 5: Primary Model Result

State	Donald Trump	Kamala Harris	Difference	Votes
Arizona	48.798	47.097	-1.701	11
Georgia	48.585	47.824	-0.761	16
Michigan	46.797	47.981	1.184	15
Nevada	46.312	48.098	1.786	6
North Carolina	48.118	48.154	0.036	16
Pennsylvania	47.215	48.844	1.629	19

Table 5: Primary Model Result

State	Donald Trump	Kamala Harris	Difference	Votes
Wisconsin	46.667	49.305	2.638	10

4.3 Electoral Vote Projections

Integrating the swing state projections with solid state predictions, the Primary Model estimates the electoral vote totals for each candidate. Kamala Harris is projected to win 292 electoral votes, surpassing the 270-vote threshold required to secure victory. Donald Trump, by comparison, is projected to receive 246 electoral votes. This projection, based on aggregating state-level outcomes, suggests that Harris has a favorable chance of achieving the necessary electoral college majority if current polling trends continue. The model indicates that Harris's projected wins in swing states like Wisconsin, Michigan, and Pennsylvania could be critical to her securing an electoral majority.

These results from both the Baseline and Primary Models provide distinct perspectives on the likely outcome of the 2024 U.S. Presidential Election. The Baseline Model projects a national-level popular vote advantage for Harris, while the Primary Model's state-based approach points to an electoral college path for her, albeit with highly competitive results in key swing states that remain essential to the final outcome.

Table 6: Predicted Electoral Votes for Each Candidate

Candidate	Solid State	Predicted Swing State	Total Predicted Votes
Harris	226	66	292
Trump	219	27	246

5 Discussion

5.1 Summary of Contributions

This paper presents a polling-based forecasting model for the 2024 U.S. Presidential Election, utilizing two complementary approaches to predict candidate support for Kamala Harris and Donald Trump. The **Baseline Model** (Winner-Take-All National Model) offers a straightforward projection of overall national sentiment by aggregating national polling data, while the **Primary Model** (Swing State-Based Model) provides a more refined, state-specific analysis, focusing especially on swing states crucial to the electoral college outcome. The two models together balance broad sentiment tracking with the targeted detail needed for swing-state-specific

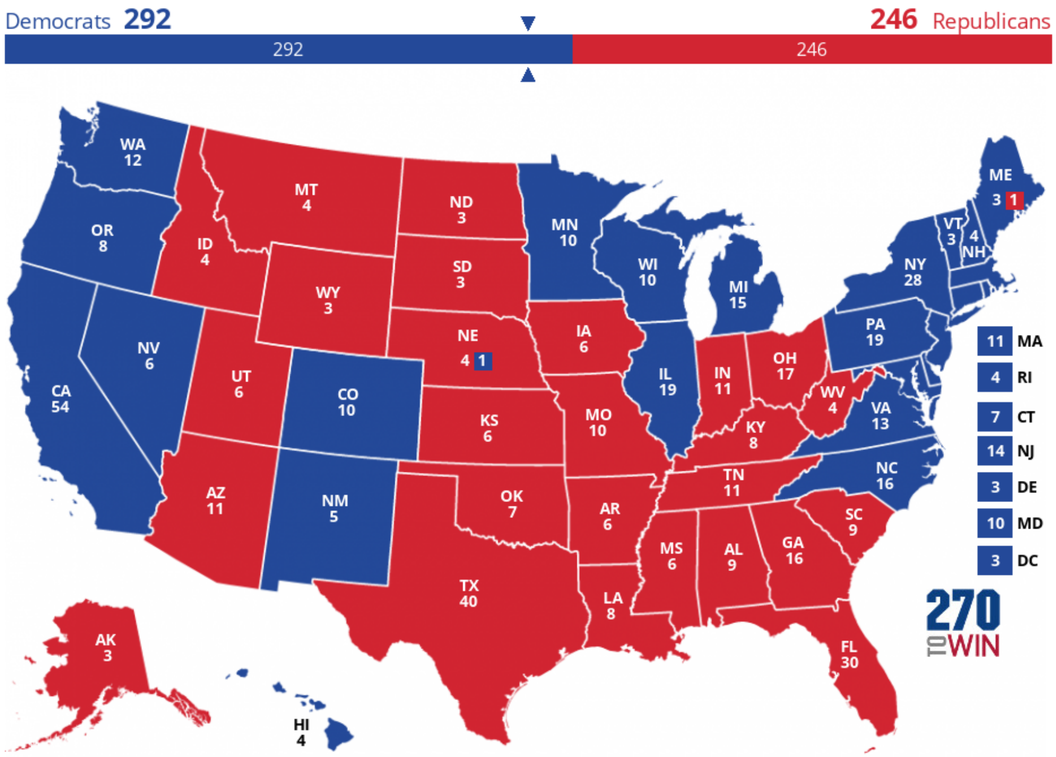


Figure 2: Projected Electoral College Results for the 2024 U.S. Presidential Election, with Kamala Harris Expected to Secure 292 Electoral Votes Against Donald Trump's 246

electoral forecasting. Through this dual approach, the analysis captures both national-level trends and the granular dynamics within pivotal swing states, offering an informative look into the likely distribution of popular and electoral votes as election day approaches.

5.2 Insights into Election Dynamics

The results reveal critical insights into the nature of U.S. elections and the structure of the electoral college system. One key takeaway from this analysis is that winning the national popular vote does not guarantee a candidate's victory in the U.S. presidential election. The Baseline Model indicates a potential popular vote lead for Kamala Harris; however, the ultimate victory depends on success within specific swing states, as highlighted by the Primary Model. This reflects a unique feature of the U.S. election system, where a candidate can win the presidency without securing the majority of the popular vote by focusing on electoral votes, especially in key battleground regions. Therefore, while Harris's projected national lead is notable, the Primary Model's swing-state analysis underscores the importance of targeting resources and strategies in these competitive areas.

5.3 Comparison of Model Effectiveness

Each model brings distinct strengths and weaknesses, serving different forecasting needs. The **Baseline Model** provides a broad, data-rich perspective, aggregating a large volume of polling data to capture general national sentiment. This simplicity allows for a straightforward, easy-to-understand representation of national trends in support levels. However, this model does not account for the state-by-state dynamics critical to winning the electoral college and may oversimplify the complexities of an election where only the electoral votes, not the popular vote, decide the winner.

In contrast, the **Primary Model** offers more precise predictions by incorporating state-level polling data for swing states, which are decisive in the electoral college outcome. By focusing on these battleground areas, the model provides a nuanced picture that aligns more closely with the mechanics of the U.S. election system. This model, however, requires more detailed data for each state and is therefore less broadly applicable to a general sense of popular sentiment but is far more insightful for understanding likely outcomes under the electoral college system.

5.4 Limitations and Areas for Improvement

Despite the insights provided by both models, there are limitations to this approach. The models rely on polling data, which is inherently subject to variability and potential biases, such as non-response bias, sampling errors, and the limitations of data collection methods across different polling organizations. Additionally, both models assume a linear trend in polling changes as election day approaches, which may not fully capture sudden shifts in voter sentiment that can occur due to unforeseen events, campaign dynamics, or emerging social issues.

Furthermore, the Baseline Model's national aggregate approach does not account for the unique political landscape within individual states, potentially underestimating the nuances in regional voting behaviors. Meanwhile, the Primary Model, while more granular, requires extensive state-specific data that may not be consistently available or reliable across all swing states.

5.5 Future Directions

Future research could benefit from integrating more sophisticated methods to address the limitations of linear assumptions and polling variability. For instance, incorporating time-series models that allow for non-linear trends could better capture sudden shifts in voter sentiment over time. Additionally, exploring Bayesian models to handle polling uncertainty more effectively may provide a way to quantify the degree of confidence in each state's predicted support levels.

Further improvements could also be achieved by incorporating demographic and socio-economic data to create more robust and representative models of each state's voter base. Finally, validating these models using historical data from previous elections could enhance their robustness and accuracy, offering a clearer picture of how these methods perform under different electoral conditions.

In conclusion, this study demonstrates the importance of integrating both national and state-level data for election forecasting and highlights how U.S. election outcomes hinge on both popular and electoral vote dynamics. By combining the strengths of a broad national model and a targeted swing-state approach, this paper provides a comprehensive framework for anticipating electoral outcomes in a system where popular sentiment and electoral mechanics are often misaligned.

A Swing State and Solid State

In the context of U.S. presidential elections, states are often classified as either "swing states" or "solid states" based on their voting behavior and predictability. Swing states, sometimes referred to as "battleground states," are those where neither major political party has a consistent or dominant advantage. This makes their electoral outcomes highly competitive and crucial for candidates aiming to secure a majority in the Electoral College. These states can shift between supporting Republican and Democratic candidates from one election to the next, depending on factors such as voter turnout, shifting demographics, and regional issues. Examples of prominent swing states for the 2024 election include Arizona, Georgia, Pennsylvania, Michigan, Nevada, North Carolina, and Wisconsin, as these states have shown significant variability in their support for both parties in recent election cycles (Elliott Davis Jr. (2024)) (Team (2024)).

In contrast, solid states are those that consistently support one political party by a substantial margin, making their outcomes highly predictable in presidential elections. For instance, states like California and New York are considered solidly Democratic, while states like Alabama and Idaho are reliably Republican. These solid states play a lesser role in the competitive dynamics of presidential campaigns, as their electoral votes are typically seen as secure for the party that historically dominates them. Consequently, candidates often focus their campaign efforts, resources, and messaging on swing states where the race is closer, since these states ultimately have the potential to decide the outcome of the election (Elliott Davis Jr. (2024)).

B Methodology Analysis of The Economist/YouGov Poll (October 6–7, 2024)

B.1 Overview

The Economist/YouGov poll, conducted from October 6 to October 7, 2024, involved 1,604 U.S. adult citizens to gather insights into public opinion on the 2024 presidential election and related political issues. YouGov's approach, known for its consistency in methodology, relies on an online sampling model that utilizes an opt-in panel of respondents. This section provides a comprehensive breakdown of the methodological elements that define this survey: the target population, sampling strategy, handling of non-response, and questionnaire design, while also highlighting its strengths and limitations (YouGov (2024a)).

B.2 Population, Frame, and Sample

In survey research, the target population is the group of people to whom the survey results are meant to generalize (YouGov (2024b)). For the Economist/YouGov poll, the target population includes all U.S. adult citizens, making it representative of the broader American public on political issues. This choice of target population is intended to capture a wide array of perspectives, from various demographics and political affiliations, to gauge national sentiment on the election (YouGov (2024a)).

The sampling frame, in contrast, represents a more specific group within the target population from which the sample is actually drawn. YouGov's sampling frame consists of members of its U.S.-based online panel, who have agreed to participate in surveys. Panelists are recruited through various channels, including commercial opt-in lists, online advertisements tailored to attract people with diverse interests, and invitations extended to past respondents from prior research initiatives. Each panelist provides background demographic information upon joining, allowing YouGov to build a versatile sampling frame with significant demographic coverage, aligning with the characteristics of the U.S. adult population (YouGov and The Economist (2024)) (YouGov (2024a)).

The sample, or the subset of individuals selected from the sampling frame to participate in the survey, includes 1,604 respondents who meet specific demographic criteria to mirror the target population's composition. This particular sample size offers a balance between reliability of data and efficiency of resources, as it's large enough to capture statistically significant findings while remaining feasible for rapid deployment online (YouGov (2024a)).

B.3 Sample Recruitment and Sampling Approach

Sampling can follow different designs, such as probability and nonprobability methods, each with unique implications. Probability sampling is the approach where each member of the target population has a known, non-zero chance of being selected, making the sample theoretically representative of the population. This method, which is often employed in high-stakes polling (e.g., election forecasting), is seen as the gold standard for ensuring unbiased, representative data (YouGov (2024b)).

However, YouGov uses a nonprobability sampling method, in which the panelists are drawn from an opt-in online panel rather than randomly from the entire U.S. adult population. Nonprobability sampling doesn't guarantee that every individual has an equal chance of selection; instead, it focuses on constructing a sample that closely matches the population based on key demographic characteristics through weighting. YouGov's approach enables it to conduct surveys quickly and cost-effectively while also allowing targeted sampling of specific groups of interest, which can be beneficial for identifying subpopulation insights or trends (YouGov (2024b)).

While nonprobability sampling offers these efficiencies, it introduces the risk of selection bias, as the individuals who join an opt-in panel may not fully reflect the population's diversity in attitudes or behaviors. For example, those who choose to participate might have higher engagement in political matters than the general public. To mitigate these potential biases, YouGov weights the sample on various demographic variables to match national population benchmarks, making the findings more reflective of the broader U.S. adult population (YouGov (2024b)).

B.4 Non-Response Handling and Weighting

Non-response bias is a concern in all survey methods and arises when certain types of individuals in the target population are less likely to participate, potentially skewing the results (YouGov (2024b)) In the Economist/YouGov poll, respondents who are unwilling or unable to participate may differ in meaningful ways from those who do participate, and these differences could influence survey outcomes if left unaddressed. For instance, if younger adults are less likely to respond, the survey could over-represent older adults' opinions, leading to results that don't accurately reflect the broader population (YouGov and The Economist (2024)).

To address non-response, YouGov employs weighting to adjust the sample composition. Weighting assigns different statistical weights to respondents based on their demographic characteristics — such as age, gender, race, income, and voting history — to ensure these characteristics align with those of the target population as measured by reliable sources like the U.S. Census. This process gives less frequently represented groups a higher weight in the analysis and more commonly represented groups a lower weight, balancing the sample to mitigate non-response effects. Weighting can't entirely eliminate the effects of non-response but substantially reduces its potential bias, making the survey findings more credible (YouGov (2024a)).

B.5 Questionnaire Design and Quality Control

YouGov prioritizes clarity, neutrality, and accessibility in questionnaire design. Clear wording helps prevent misunderstandings, while neutral language avoids leading respondents to particular answers. The survey also incorporates randomization of question and response option orders, a strategy that reduces the likelihood of order bias, where responses may be influenced by the sequence in which options are presented. YouGov's commitment to neutrality and randomization adds to the reliability and validity of the responses, making the findings more trustworthy YouGov and The Economist (2024).

Quality control measures further reinforce the survey's integrity. YouGov's approach includes verifying panelists' identity and monitoring their responses for consistency. For instance, individuals providing contradictory answers or completing the survey unusually quickly are flagged, with their data excluded from the final analysis. This process ensures that only high-quality responses contribute to the survey's conclusions YouGov (2024b).

While the survey's straightforward question format enhances accessibility, it can limit depth in exploring complex topics, as respondents may not have the opportunity to fully express nuanced views. For example, questions about candidate preference capture top-line support but may lack the depth to reveal the underlying motivations or concerns driving these preferences.

B.6 Strengths and Limitations of YouGov's Methodology

YouGov's methodology provides a flexible and efficient way to gather public opinion, allowing rapid data collection across a broad demographic range. The use of an opt-in panel enables YouGov to survey specific subgroups effectively and maintain control over respondent quality. By applying weighting to compensate for nonprobability sampling, YouGov achieves a sample composition that closely resembles the U.S. adult population, allowing generalizations about national opinion.

However, the reliance on nonprobability sampling introduces certain limitations. Since the sample is drawn from an opt-in panel, there is an inherent risk of selection bias, which, despite efforts to control for it through weighting, may affect the representativeness of the findings. The use of online-only surveys also limits participation to individuals with internet access, although this exclusion is less impactful in the U.S., where internet penetration is high.

In conclusion, YouGov's methodology balances efficiency and accuracy well, leveraging demographic weighting and robust quality controls to produce credible insights. While limitations related to nonprobability sampling and selection bias exist, these are mitigated through strategic adjustments, making the poll a reliable source for gauging national opinions.

B.7 Reflection

We created an account on YouGov to gain real experience with the platform, as illustrated in Figure 3 and here is my reflection on the onboarding process and user experience.

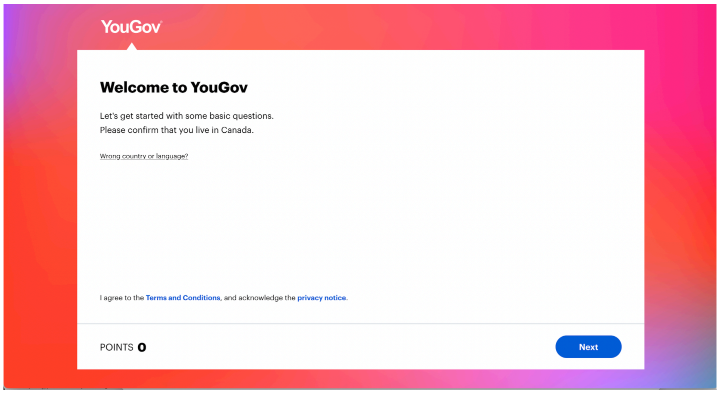


Figure 3: Initial welcome screen during YouGov onboarding, confirming country of residence and introducing users to the platform's privacy policies and terms

The YouGov survey onboarding process is designed to be clear and straightforward, making it easy for new users to sign up and begin participating. The questions asked during onboarding are simple and direct, covering basic demographic details without overwhelming the participant. The first question of the first survey we took at YouGov is shown at Figure 4. This streamlined approach likely helps increase engagement and reduces drop-off rates, ensuring that users can quickly understand what is required of them. The layout and visuals are accessible, contributing to a positive first impression and making it more likely that new panelists will continue engaging with future surveys.

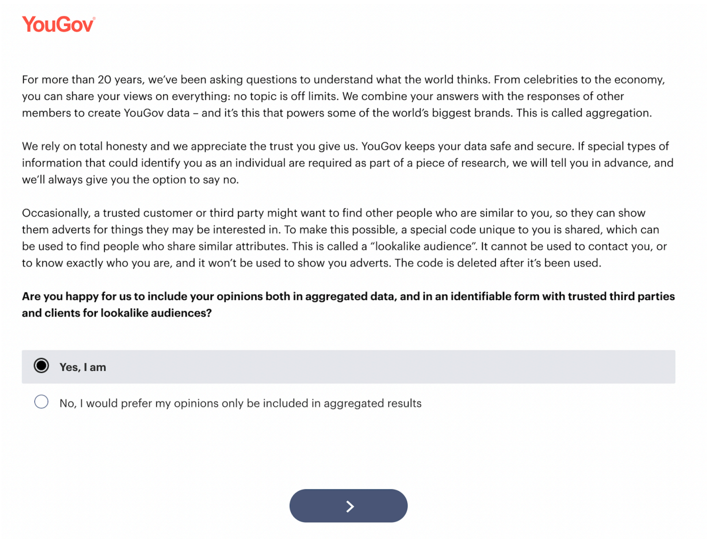


Figure 4: Example of a YouGov survey question, designed to gather demographic information and establish privacy preferences, fostering user engagement through a clear and straightforward layout

To enhance data integrity, YouGov includes verification steps such as confirming users' email addresses through a code. This extra step not only ensures that participants are real and unique individuals but also helps prevent fraudulent or duplicate accounts, as shown in Figure 5, which can otherwise skew survey results. By verifying email addresses, YouGov can maintain a higher quality of data and build a more reliable panel, knowing that each participant is committed enough to complete these security measures. This verification process is a straightforward yet effective method for improving data quality from the outset.

Data privacy is also well-considered in YouGov's onboarding, with clear options allowing users to control how their information is shared. For instance, participants can opt in or out of having their responses shared in identifiable formats with trusted third parties. Figure 6 includes an example of how YouGov is letting users to control their information privacy. This transparency around data usage builds trust, showing participants that their privacy is valued and that they have agency over their data. The option to decide on data sharing likely encourages long-term participation, as users are reassured that their information will be handled responsibly.

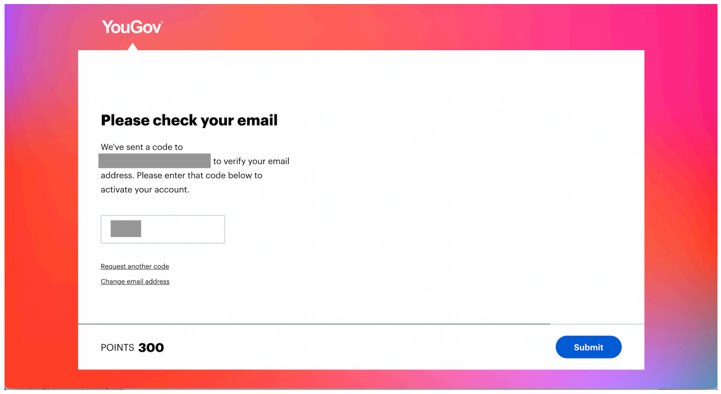


Figure 5: Email verification screen to confirm user identity, enhancing data integrity by preventing duplicate or fraudulent accounts

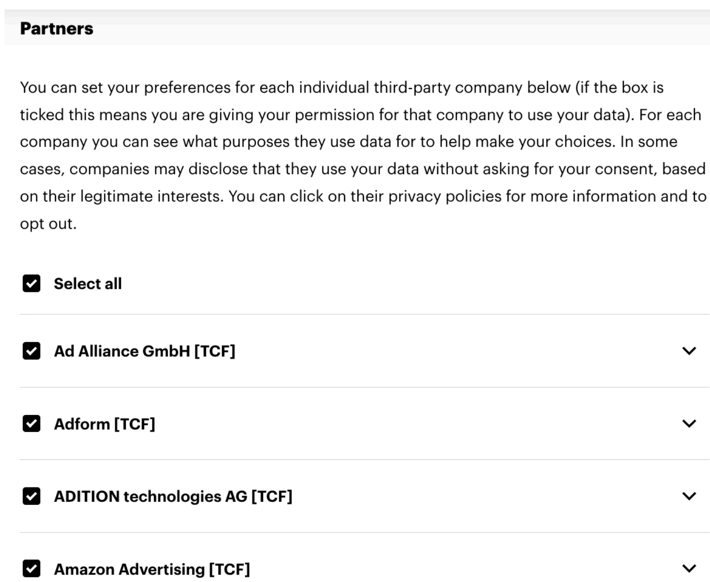


Figure 6: Data privacy options allowing users to control the sharing of their information with third parties, reflecting YouGov's commitment to transparency and user autonomy

However, the onboarding interface in Figure 7 has a gamified feel, awarding points for each step completed. This reward system, while motivating, could encourage participants to treat the surveys more like a task for monetary gain than an opportunity to provide genuine responses. The point-based incentive structure may attract users who are primarily interested in earning rewards, which risks impacting data quality if participants prioritize speed over thoughtful answers. This setup introduces the possibility that some users are engaging more for the rewards than for contributing to research, which could affect the reliability of the responses YouGov collects.

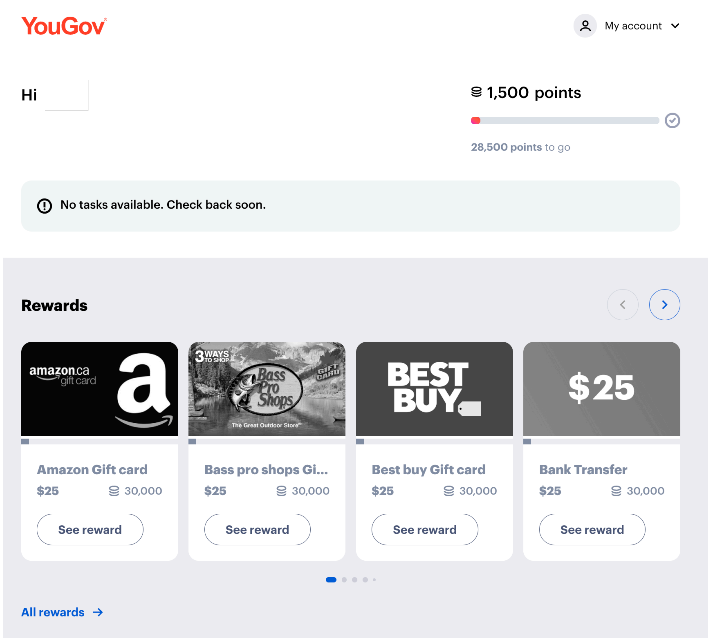


Figure 7: Overview of YouGov's gamified rewards system, where participants accumulate points redeemable for gift cards, potentially motivating user engagement with surveys

C Methodology and Survey Design for 2024 U.S. Presidential Election Forecast

C.1 Overview

This methodology outlines an election forecasting plan with a budget of $100,000, focusing on swing states where election outcomes are uncertain, making them essential for forecasting the Electoral College result. Drawing on research from YouGov and the Annual Review of Economics on survey methods, we selected an online format to improve accessibility and reduce selection bias, allowing respondents to participate at their convenience(Stantcheva (2023)). This approach is particularly suited for students and working-age individuals who may have limited availability during traditional hours. Mobile technology also increases participation, reaching groups that are otherwise difficult to access, such as seniors and individuals who primarily use mobile devices. To encourage participation and reduce drop-offs, we created a concise survey with digital incentives to attract a broad range of respondents across income levels. By incorporating methods from both sources, our approach is designed to be cost-effective and demographically balanced, offering reliable predictions for the 2024 U.S. Presidential Election.

C.2 Sampling Approach

This methodology uses a multi-mode sampling approach, integrating both probability and non-probability methods to achieve broad demographic reach and representativeness. This approach enables us to capture an inclusive snapshot of voter preferences in critical electoral regions.

C.2.1 Target Population

The target population is eligible voters in the seven swing states: Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin. By focusing on these states, the survey aims to forecast potential outcomes in critical areas for the Electoral College.

C.2.2 Sampling Frame and Sample Size

The sampling frame combines both verified voter lists (probability-based) and opt-in online panels (non-probability) to access a comprehensive range of respondents. Our total sample size is 7,000 respondents (1,000 per state), which supports statistically reliable insights across age groups and geographic regions.

C.2.3 Sample Recruitment by Age Group

- 18-24 years: Recruited through university partnerships, with a target of 150 respondents per state (1,050 total).
- 25-60 years: Recruited via online ads on platforms like Facebook and Prolific, targeting 700 respondents per state (4,900 total).
- 60+ years: Reached through live phone calls, with 150 respondents per state (1,050 total).

C.2.4 Sampling Methods

Probability Sampling (Phone Interviews): For the 60+ age group, live phone interviews and Interactive Voice Response (IVR) calls are employed. Using verified voter lists reduces bias and increases data accuracy for less digitally active participants. Non-Probability Sampling (Online Surveys): For participants aged 18-60, online surveys target mobile-first and younger voters through social media and survey platforms. Although not all respondents have equal selection probability, quota sampling and post-stratification weighting enhance representativeness.

C.2.5 Stratified Sampling and Weighting

Stratified Sampling: The sample is stratified by demographic characteristics such as age, gender, race/ethnicity, education, and income, to improve representativeness. Post-Stratification Weighting: To correct for any sample imbalances, post-stratification weighting adjusts the sample to match demographic proportions from U.S. Census data, ensuring balanced representation across demographic categories.

C.3 Recruitment Strategy

Using verified voter lists and opt-in online panels as the sampling frame, this survey focuses on likely voters in the 2024 election. Likely voters are identified based on past voting behavior and registration status. To align with demographic targets, quota sampling is employed. Quota sampling allows us to proportionally represent demographic groups by age, gender, race, and education, ensuring the sample reflects the U.S. electorate's profile. Our primary focus is on controlling age distribution. As each age group reaches its target sample size, we close the survey for that group to prevent over-representation. This method ensures that all age demographics are adequately represented without skewing results due to excess responses from any particular age group. Incentives: To encourage participation, especially for online surveys targeting younger and working-age demographics, we offer a $10 digital gift card to each respondent upon survey completion. This incentive helps attract a more diverse group, increasing the likelihood of representative participation across income levels and ensuring a higher response rate.

C.4 Data Validation

Given the hybrid data collection format, we use several validation techniques to maintain data quality: - Duplicate detection (based on IP addresses and phone numbers) ensures unique responses. - Attention checks help identify respondents who may not be fully engaged. - Conflict detection: Responses showing contradictions (e.g., conflicting demographic details or inconsistencies in voting behavior) are flagged. If respondents provide conflicting answers, their data is excluded, as this may indicate they completed the survey solely for incentives without thoughtful engagement. Both duplicate and inattentive responses introduce bias and are therefore removed from the data set. This filtering process improves data accuracy and reliability. To further align the sample with national demographics, post-stratification weighting is applied. Weights are calculated based on key variables such as race, gender, and education, following Census benchmarks. This weighting corrects for any under- or over-represented groups, resulting in data that is more reflective of the broader electorate.

C.5 Trade-Offs

Data Diversity vs. Recruitment Control: Although quota sampling on age helps ensure demographic diversity, strict quotas for every characteristic (such as race and income) are difficult to impose. This may lead to over- or under-representation of certain groups. Solution: Post-stratification weighting is used to adjust the sample based on Census proportions, improving demographic balance. Accurate Senior Outreach vs. Higher Costs: Phone interviews are essential to reach senior citizens but come at a higher cost. Additionally, some seniors may remain unreachable, potentially skewing representation. Solution: Allocate phone interview resources to maximize outreach while acknowledging potential non-response among unreachable seniors. Incentives vs. Economic Bias: Offering $10 digital incentives may disproportionately attract lower-income individuals, possibly skewing the sample toward lower socioeconomic backgrounds. Solution: Income-based weighting in post-stratification helps correct for any economic biases that might affect the sample's representativeness.

C.6 Conclusion

By integrating YouGov's multi-modal sampling with the Economist's robust weighting approach, this methodology effectively captures a broad cross-section of voter demographics. The use of both probability-based and non-probability-based sampling methods, along with rigorous data validation, ensures a reliable forecast for the 2024 U.S. Presidential Election.

C.7 Survey Link

https://forms.gle/2MGYeZavDsCNuWZ1A

C.8 Copy of Survey

Below is the full content of the survey to be implemented using Google Forms:

1. What is your age?
    - 18-24

- 25-39
- 40-60
- 60+

2. What is your gender?
   - Male
   - Female
   - Other: _____
3. What is your race/ethnicity?
   - White
   - Black
   - Hispanic or Latino
   - Asian
   - Indigenous
   - Other: _____
4. What is your highest level of education?
   - Less than high school
   - High school diploma
   - Some college
   - Bachelor's degree
   - Graduate degree or higher
5. What is your annual household income?
   - Less than $25,000

- $25,000 - $49,999
- $50,000 - $99,999
- $100,000 or more

6. Which state do you currently reside in?
7. Are you a registered voter?
   - Yes
   - No
   - Not sure
8. How likely are you to vote in the 2024 presidential election?
   - 1 (Definitely will not vote)
   - 2
   - 3
   - 4
   - 5 (Definitely will vote)
9. If the 2024 election were held today, who would you vote for?
   - Kamala Harris (Democrat)
   - Donald Trump (Republican)
   - Undecided
   - Other: _____
10. What is the most important issue for you in this election?

- The economy
- Healthcare

- Immigration
- Climate change
- Social Security and Medicare
- Foreign policy

11. In your opinion, how will the majority in your state vote in 2024?

    - Democrat
    - Republican
    - Too close to predict

12. Do you have any additional comments or suggestions?

Thank you for completing the survey! Your input is greatly appreciated and will help provide insights into the upcoming 2024 election.

Here's the revised acknowledgments section using "we" and incorporating `@citeR` for R:

D Acknowledgements

We would like to express our gratitude to the developers and contributors of the R programming language by R Core Team (2023) and its ecosystem of packages, which have been invaluable in conducting this research. The **tidyverse** collection by Wickham et al. (2019), particularly its extensive functionality for data manipulation and visualization, has greatly enhanced the clarity and effectiveness of our analysis, and **arrow** package by Richardson et al. (2024) enables us to work with parquet files.

For generating compact and informative summaries of our data, we utilized the **skimr** package by Waring et al. (2022), which provided a flexible overview of the dataset. The **knitr** package by Xie (2023) facilitated the dynamic report generation, allowing us to seamlessly integrate analysis results into our documentation.

To enhance our visualizations, we employed **GGally** by Schloerke et al. (2024), an extension of **ggplot2**, which allowed for more complex and informative graphics . The **dplyr** package by Wickham et al. (2023) was instrumental in data manipulation, enabling efficient data transformation and summarization. For converting statistical results into tidy formats, we relied on the **broom** package by Robinson, Hayes, and Couch (2023), which simplified the reporting of statistical outputs .

We also utilized **styler** by Müller and Walthert (2023) to maintain a clean and readable codebase, ensuring that our R scripts adhered to best practices for code formatting. The **lubridate** package by Grolemund and Wickham (2011) simplified the handling of date and time data, which was crucial for temporal analyses . Lastly, the **validate** package by van der Loo and de Jonge (2021) played a key role in ensuring the integrity and validity of our data throughout the research process.

The contributions of these packages have greatly enriched our work, and we appreciate the effort of the R community in maintaining and enhancing these resources.

References

Electoral Ventures LLC. 2024. "270toWin - 2024 Presidential Election Interactive Map." https://www.270towin.com/.

Elliott Davis Jr. 2024. *7 States That Could Sway the 2024 Presidential Election*. https://www.usnews.com/news/elections/articles/7-swing-states-that-could-decide-the-2024-presidential-election.

FiveThirtyEight. 2024. "National: President: General Election: 2024 Polls." ABC News Internet Ventures. https://projects.fivethirtyeight.com/polls/president-general/2024/national/.

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

Müller, Kirill, and Lorenz Walthert. 2023. *Styler: Non-Invasive Pretty Printing of r Code*. https://CRAN.R-project.org/package=styler.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

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

Schloerke, Barret, Di Cook, Joseph Larmarange, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg, and Jason Crowley. 2024. *GGally: Extension to 'Ggplot2'*. https://CRAN.R-project.org/package=GGally.

Stantcheva, Stefanie. 2023. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." *Annual Review of Economics* 15 (1): 205–34. https://doi.org/10.1146/annurev-economics-091622-010157.

Team, USA Facts. 2024. *What Are the Current Swing States, and How Have They Changed over Time?* https://usafacts.org/articles/what-are-the-current-swing-states-and-how-have-they-changed-over-time/.

van der Loo, Mark P. J., and Edwin de Jonge. 2021. "Data Validation Infrastructure for R." *Journal of Statistical Software* 97 (10): 1–31. https://doi.org/10.18637/jss.v097.i10.

Waring, Elin, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu, and Shannon Ellis. 2022. *Skimr: Compact and Flexible Summaries of Data*. https://CRAN.R-project.org/package=skimr.

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

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

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

YouGov. 2024b. *Methodology*. https://today.yougov.com/about/panel-methodology.

———. 2024a. *Methodology*. https://d3nkl3psvxxpe9.cloudfront.net/documents/econTabReport_pw9W1fW.pdf.

YouGov, and The Economist. 2024. *YouGov Methodology for the Economist Polls*. https://www.economist.com/media/pdf/Methodology.pdf.