This Election is too Close to Call: Poll of Polls Methodology and Bayesian Modeling Results show a Statistical Tie in the Seven Battleground States*

Vice President Kamala Harris Projected to Win 270 Electoral Votes and 48% of the Popular Vote Based on Poll of Polls and Bayesian Modeling Analysis

Talia Fabregas   Fatimah Yunusa   Aliza Mithwani

October 29, 2024

The U.S. Presidential election will take place on Tuesday, November 5th with Vice President Kamala Harris and former President Donald Trump in a close race for the White House. In this paper, we used the poll-of-polls method and applied a Bayesian model to estimate the winner of the popular vote and the winner of each of the seven key battleground states: Arizona, Georgia, Nevada, North Carolina, Michigan, Wisconsin, and Pennsylvania, and Nebraska's second congressional disctict. Using the results of our poll-of-polls analysis, we can predict that Vice President Harris will win x% of the popular vote and carry the states of Michigan, Wisconsin, and Pennsylvania, plus Nebraska's second congressional district and former President Trump will win x% of the popular vote and the states of Arizona, Georgia, Nevada, and North Carolina. Our electoral college and swing-state predictions both fall within the margin of error;

1 Introduction

Every fourth year, on the first Tuesday of November, Americans head to the polls to choose their president. Next week, on Tuesday November 5th, 2024, Americans will head to the polls to elect their 47th president. At a time of unprecedented division and political polarization, ... The U.S. Presidential Election will take place on Tuesday November 5th. Vice President

---

Kamala Harris and former President Donald Trump will vie to become the 47th President of the United States. Vice President Kamala Harris became the Democratic nominee after 82-year-old President Joe Biden made a historic and unprecedented decision to end his re-election campaign on July 21, 2024. She secured enough delegates to win the Democratic nomination on Vice President Kamala Harris and Former President Donald Trump … President Joe Biden made the unprecedented decision to end his re-election campaign on July 21, 2024 and immediately endorsed Vice President Kamala Harris. She became the presumptive Democratic presidential nominee the same day and secured enough delegates to win the nomination on July 28.

This study only considers polling data from after President Joe Biden ended his re-election campaign and Vice President Kamala Harris became the presumptive Democratic nominee for President. Polls conducted before July 21, 2024 were not considered because they include President Joe Biden, who is no longer running for President, as the Democratic nominee.

The remainder of this paper is structured as follows. Section 2 contains an overview of the polling dataset used, visualizations of different variables, and summary statistics. Section 3 contains the Bayesian regression models used to predict the percentages of voters that will support Vice President Harris and former President Trump. Section 4 contains tables and visualizations that present the results of applying the model. Section 5. Appendix A.

The statistical programming language (R Core Team 2023) and the `tidyverse`, `janitor`, `ggplot`, `kableExtra`, `arrow`, `rstanarm`, and `spline` packages were used to perform this election forecast, clean the dataset, build the Bayesian regression models, create data visualizations and tables, and apply the model.

2 Data

2.1 Overview

This presidential election forecast was performed using the statistical programming language R (R Core Team 2023) and Presidential general election polling data from FiveThirtyEight (FiveThirtyEight 2024). The data was downloaded on October 29, 2024; polling data released after this date was not considered anywhere in this paper. The presidential polls dataset from FiveThirtyEight contains national polls and state-level polls for each of the 50 states and congressional districts in Maine and Nebraska (FiveThirtyEight 2024). We cleaned this dataset to only include high-quality polls at the national and state-level. However, we only analyzed state-level data for the seven swing states that are expected to determine the results of this election: Michigan, Wisconsin, Pennsylvania, North Carolina, Georgia, Nevada, and Arizona.

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

2.3 Outcome and predictor variables

We will use `end_date` (the date that a poll was completed), `state`, `pollster`, and `pollscore` to predict support for Vice President Harris and former President Trump at the national level, at the state level for each of the seven battleground states, and in Nebraska's second congressional district in Section 3.

2.3.1 National and swing state polling averages over time

The percentage of the electorate voting for Vice President Harris and former President Trump varies by state; in some states support for Vice President Harris is higher than the national average and in others, it is lower than the national average. The 2024 presidential election is expected to be decided by seven swing states: Arizona, Georgia, Nevada, North Carolina, Wisconsin, Michigan, and Pennsylvania and Nebraska's second congressional district. Two states, Maine and Nebraska, award electoral votes by congressional district.

Table 1: Polling averages for Harris and Trump at the national level and at the state level for the states included in the polling dataset as of October 29, 2024 show a narrow popular vote lead for Vice President Harris and extremely close races in the 7 battleground states (Arizona, Georgia, Nevada, North Carolina, Michigan, Pennsylvania, Wisconsin).

State	Harris %	Trump %
Popular Vote	50.5	48.0
Arizona	47.0	51.0
Georgia	46.0	51.0
Nevada	51.0	47.0
Pennsylvania	49.0	49.0
Michigan	51.0	46.0
Wisconsin	50.0	47.0
North Carolina	48.0	50.0
Florida	46.0	52.0
Minnesota	53.0	43.0
Missouri	41.0	54.0
Montana	39.5	56.5
Nebraska	39.5	54.0
Nebraska CD-2	53.5	41.5

Table 1: Polling averages for Harris and Trump at the national level and at the state level for the states included in the polling dataset as of October 29, 2024 show a narrow popular vote lead for Vice President Harris and extremely close races in the 7 battleground states (Arizona, Georgia, Nevada, North Carolina, Michigan, Pennsylvania, Wisconsin).

State	Harris %	Trump %
New Hampshire	52.0	45.0
Ohio	45.0	52.0
Texas	41.0	51.5
Virginia	52.0	44.0

In the presidential polling data we obtained from FiveThirtyEight, Vice President Harris leads former President Trump in the popular vote 50.5% to 48.0%. The margins in the seven battleground states are tight, with Vice President Harris leading in Nevada, Michigan, and Wisconsin, and F (FiveThirtyEight 2024). The dataset also includes polling data from Florida, Minnesota, Missouri, Montana, Nebraska, New Hampshire, Ohio, Texas, and Virginia, but those states are not expeced to determine the results of the election.

Since President Biden ended his re-election campaign and Vice President Harris became the Democratic Presidential nominee, the polls have shown a dead heat between Vice President Harris and former President Trump. Polling averages for the six months leading up to election day, including from before President Biden withdrew on July 21, 2024 can be found in Appendix C.2.

2.3.2 Pollster

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

2.3.3 Pollscore

Table 2: Only high-quality datasets with a numeric grade of at least 3.0 were included, so the polls included all have good pollscores. Negative/lower pollscores are better.

Pollscore	Number of Polls
-1.5	112

Table 2: Only high-quality datasets with a numeric grade of at least 3.0 were included, so the polls included all have good pollscores. Negative/lower pollscores are better.

Pollscore	Number of Polls
-1.2	8
-1.1	114

Table 3: All the pollsters included in our analysis dataset have a numeric grade of 3, but Siena/NYT has the best pollscore at -1.5.

Pollster	Candidate	Average Pollscore
CES / YouGov	-1.1	3
Marquette Law School	-1.1	3
McCourtney Institute/YouGov	-1.1	3
YouGov	-1.1	3
The Washington Post	-1.2	3
Siena/NYT	-1.5	3

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 modeling strategy is to use a Bayesian analysis model to investigate the relationship between the percentage of voters in a poll who support Vice President Harris or Former President Trump and the date the poll was conducted, the state (or if it was a national poll), the pollster who conducted the poll, and the pollscore it received. Model details and diagnostics are presented in Appendix D.

We use two Bayesian regression models, one each to model the percentage of voters supporting Vice President Harris, *pct_harris* and the percentage of voters supporting former President Trump, *pct_trump*.

The models to predict *pct_harris* and *pct_trump* both use the following predictors: - `end_date_num`: the number of days since July 21, 2024, when President Biden ended his re-election campaign and endorsed Vice President Harris. This is the spline term; it uses an spline with 5 degrees of freedom to model changes in *pct_harris* (in the Harris model) and *pct_trump* (in the Trump model) over time.

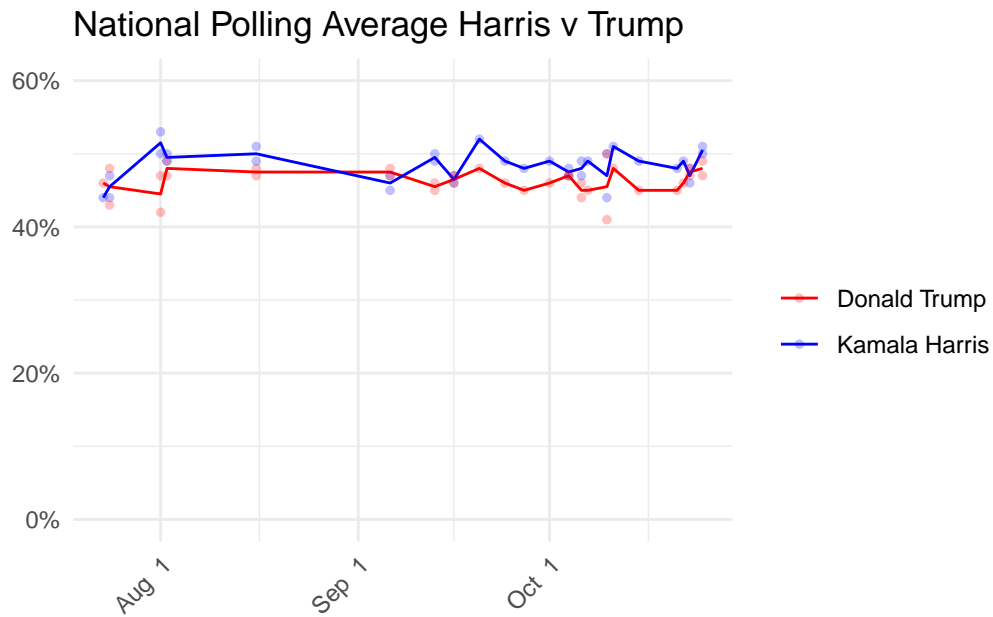## National Polling Average Harris v Trump

Figure 1: National popular vote polling averages for Harris and Trump since Biden ended his re-election campaign on July 21, 2024. As of October 29, 2024, Harris has a narrow lead over Trump.
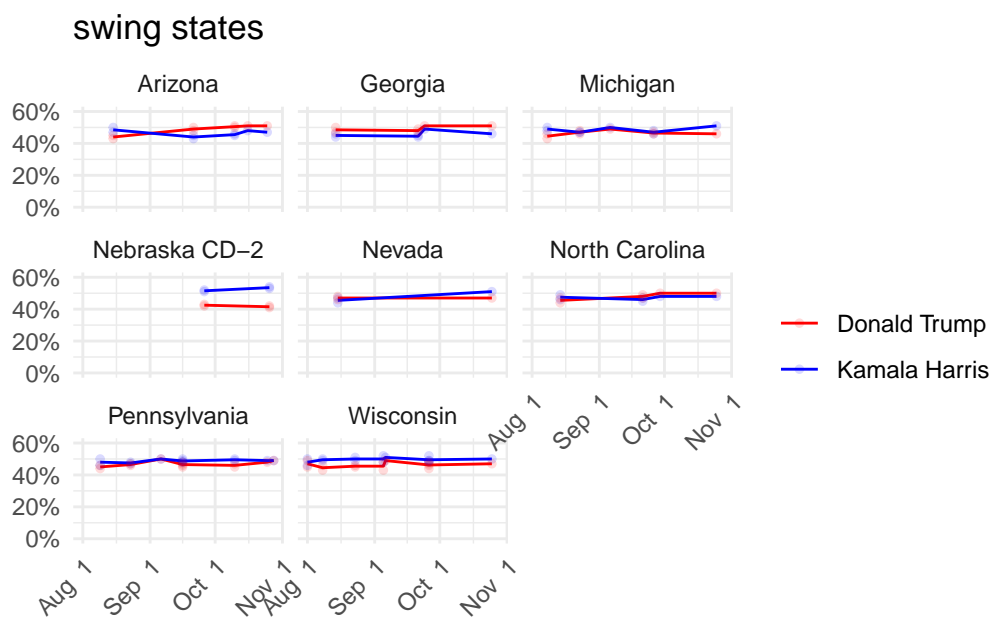
swing states

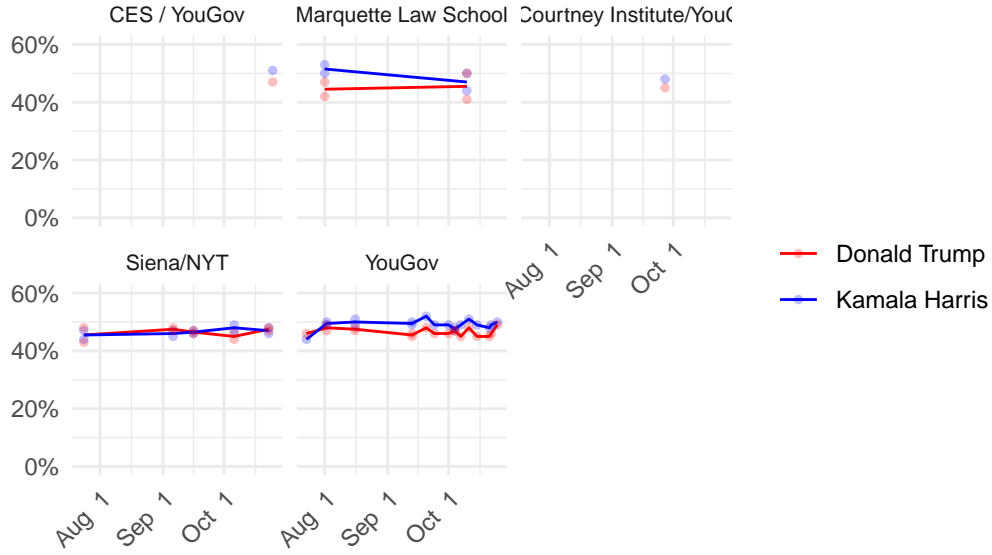Figure 2: seven battleground states with harris

Figure 3: Pollsters caption

- `state`: accounts for the change in *pct_harris* and *pct_trump* at the state-level for a particular state or at the national level.

- `pollster`: accounts for the differences in *pct_harris* and *pct_trump* among different pollsters.

- `pollscore:`adjusts for the pollscore rating, which is the Predictive Optimization of Latent skill Level in Surveys, Considering Overall Record, Empirically, calculated by averaging the predictive error and predictive bias of a poll (Morris 2024).

3.1 Model set-up

Define pct_harris_i as the percentage of voters supporting Vice President Harris and pct_trump_i as the percentage of voters supporting former President Trump.

$$pct\_harris_i = \beta_0 + \beta_1 \cdot ns(end\_date\_num_i, df = 5) + \beta_2 \cdot state_i + \beta_3 \cdot pollster_i + \beta_4 pollscore_i$$
$$\beta_0 \sim \text{Normal}(50, 10)$$
$$\beta_1 \sim \text{Normal}(0, 5)$$
$$\beta_2 \sim \text{Normal}(0, 5)$$
$$\beta_3 \sim \text{Normal}(0, 5)$$
$$\beta_4 \sim \text{Normal}(0, 5)$$

7

$$pct\_trump_i = \beta_0 + \beta_1 \cdot ns(end\_date\_num_i, df = 5) + \beta_2 \cdot state_i + \beta_3 \cdot pollster_i + \beta_4 pollscore_i$$
$$\beta_0 \sim \text{Normal}(50, 10)$$
$$\beta_1 \sim \text{Normal}(0, 5)$$
$$\beta_2 \sim \text{Normal}(0, 5)$$
$$\beta_3 \sim \text{Normal}(0, 5)$$
$$\beta_4 \sim \text{Normal}(0, 5)$$

The model is run in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). The default priors from `rstanarm` are used for both GLM Bayesian models. T intercept normal prior with $\mu = 50$ and $\sigma = 10$ reflects the central tendencies of Harris and Trump's polling percentages, influenced by prior knowledge and the predictors use a normal prior with $\mu = 0$ and $\sigma = 5$ (Goodrich et al. 2022).
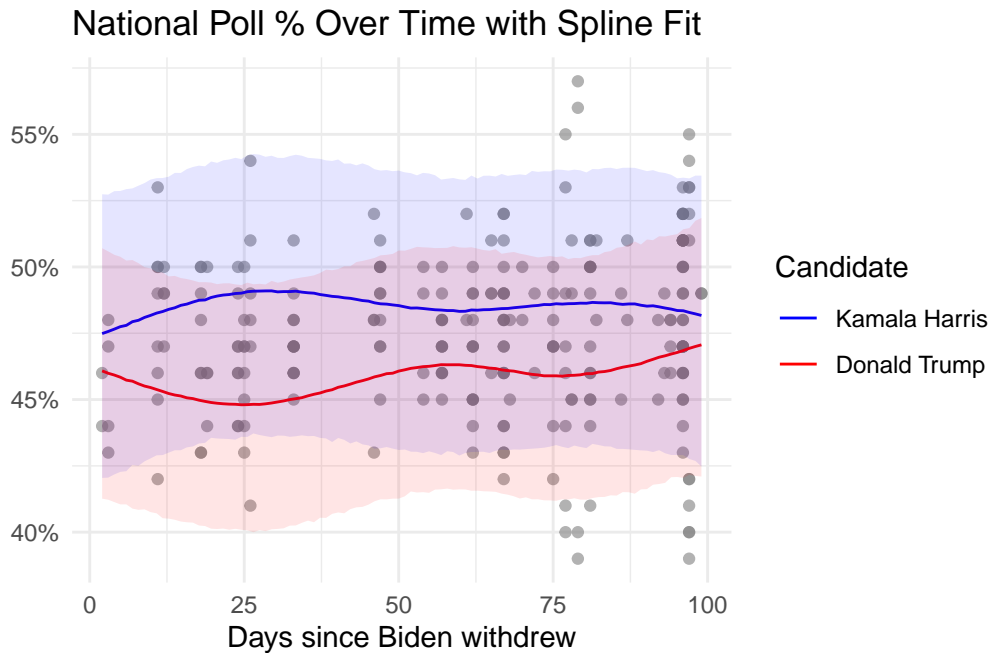
3.1.1 Model justification

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

4 Results

The model results and model summary are presented in Appendix D.

4.1 National Popular Vote Results

Popular vote prediction after applying

National Poll % Over Time with Spline Fit

4.2 State-Level Results for the Seven Battleground States and Nebraska's Second Congressional District

5 Discussion

5.1 Margin of error, close race, why either could win

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

5.2 Polling flaws in 2016, 2020, and 2022

Discuss polling flaws in 2016 and 2020, how support for Trump was underestimated. Discuss how this was overcorrected in 2022, and how support for GOP house and senate candidates was overestimated. Discuss Trump's unique appeal to a certain segment of the electorate. Refer to Nate Silver articles.

(a) Avg Trump lead +1.61

(b) Avg Trump lead +3.01

(c) Avg Trump lead +0.37

(d) Avg Trump lead +0.44

(e) Avg Harris lead +2.22

(f) Avg Harris lead +2.57

(g) Avg Harris lead +3.31
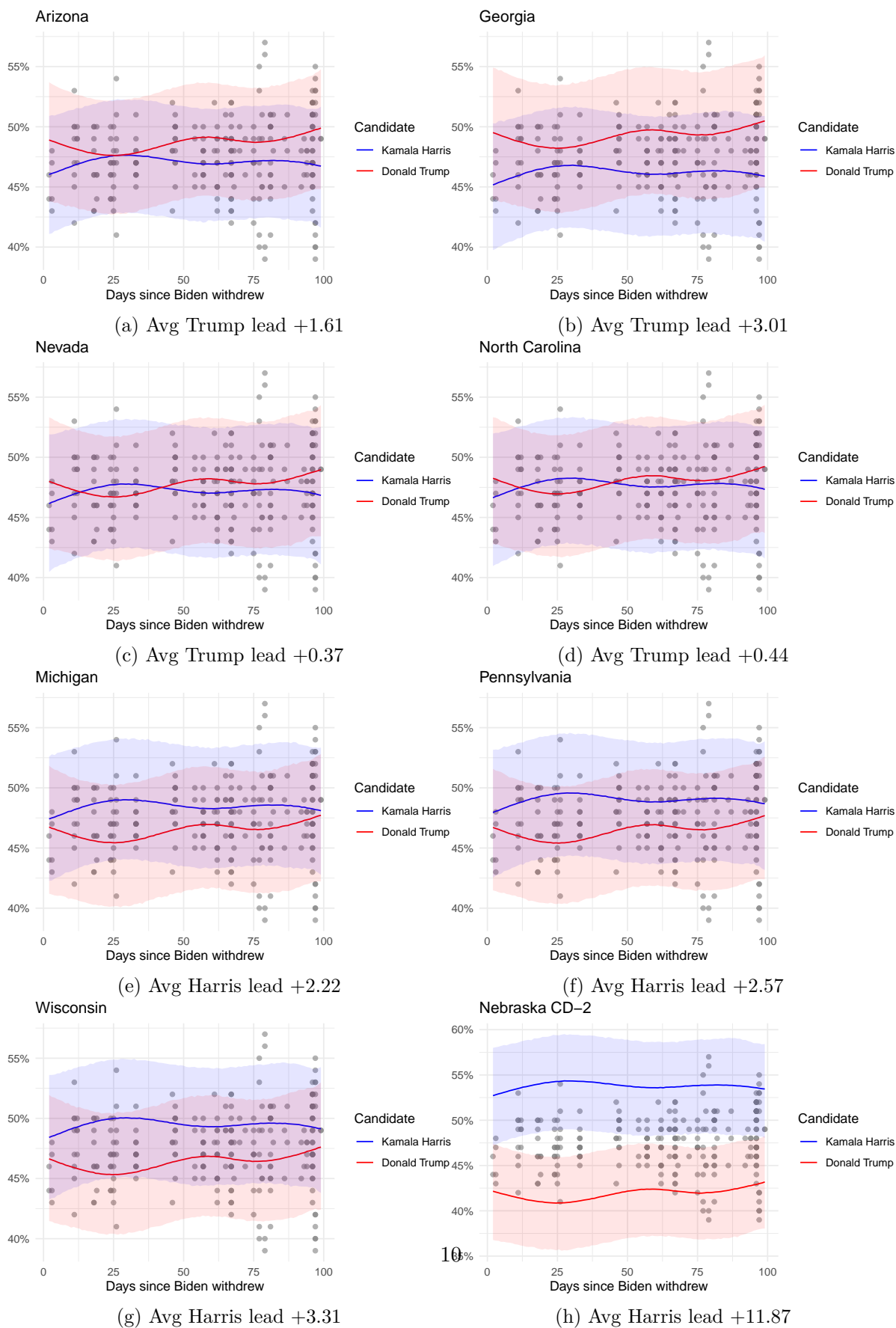
(h) Avg Harris lead +11.87

Figure 4: On average, Trump has had narrow leads in Arizona, Georgia, Nevada, and North Carolina, while Harris has narrow leads in Michigan, Pennsylvania, and Wisconsin.

5.3 Something else about polling, nonresponse, and how a repeat of any of these flaws could swing the race either way

5.4 Weaknesses, limitations, and next steps

Biden's withdrawal, use of two models, no third-party candidates, only national popular vote and swing states were analyzed, we did not consider how the pollsters in our dataset handled nonresponse when modeling. Weaknesses and next steps should also be included.

Appendix

A Idealized methodology

A.1 Survey objectives

A.2 Sampling approach

The target population for our idealized survey is likely voters in the United States. We will use the L2 voter database from L2 (2024) to build a representative sampling frame. The L2 voter database is one of the most trusted sources for enhanced voter data and it includes detailed demographics and voting history data (L2 2024). Using this data set is the first step towards ensuring that our sample aligns with the general electorate and is representative of likely voters so that we can sample precisely.

We will use a stratified sampling approach to closely examine voter demographics. Stratified sampling allows us to look at every stratum and carry out simple random sampling with those strata (Alexander 2024) and its primary goal is to ensure that all strata of the population are represented (Neyman 1934). We selected stratified sampling because it will allow us to have representation within the subgroups that we are interested in and it has a reduced sampling error and improved accuracy rate (Alexander 2024). We considered simple random sampling, but ultimately chose stratified sampling instead because it tends to produce more precise accurates when used to forecast U.S. elections (Pew Research Centre 2024). The U.S. election is decided by the electoral college, not the popular vote so we will oversample from the seven key battleground states that are expected to decide the 2024 election: Arizona, Nevada, Georgia, North Carolina, Wisconsin, Michigan, and Pennsylvania (FiveThirtyEight 2024). Oversampling from the swing states will allow us to put an emphasis on forecasting their results and increase the accuracy of our electoral college estimate. Our target sample size of 100,000 respondents will provide a margin of error of 2% at a 95% confidence level; this will allow us to be precise while still accounting for budgetary constraints (Pew Research Centre 2024).

A.3 Respondent recruitment

A.4 Idealized survey

2024 U.S. Presidential Election Poll This survey collects information about voters' political views and who they support in the 2024 U.S. Presidential Election. The data collected will not be shared with any external parties and will strictly be used for analytical purposes. This survey is completely anonymous and your data will be protected. Any published material regarding the results drawn from this survey cannot be traced back to you. The goal of this survey is to draw conclusions about the 2024 presidential elections held in the United States.

Please answer as accurately as possible. If you have any questions or concerns, please reach out to aliza.mithwani@mail.utoronto.ca (correspondence will not be shared with any external parties).

1. Are you a registered voter in the United States ? Yes No
2. Do you plan to vote in the upcoming presidential election? Yes No Undecided

Demographics Questions

3. Would you consider yourself: White Black or African American Hispanic or Latino Asian American Indian or Alaskan Native Middle Eastern or North African Native Hawaiian or Pacific Islander Prefer not to say Other (specify)
4. What is your age? 18-29 20-44 45-64 65+ Prefer not to say

5. What sex were you assigned at birth, on your original birth certificate? "Female" "Male" Prefer not to say

6. How do you currently describe yourself (select all that apply)? "Female" "Male" "Transgender" Prefer not to say Other (Specify)

7. What is your household income? Less than $20,000 $20,000-59,999 $60,000-79,999 $80,000-99,999 $100,000 or more Prefer not to say

8. In which state do you currently reside? Dropdown list of states

Candidate Evaluation

9. If the 2024 presidential election were held today, who would you vote for?

Donald Trump, Republican Kamala Harris, Democrat Jill Stein, Green Party Write-in Don't know Prefer not to say Other (please specify)

(Optional) If you selected "Write-in" for the last question, please specify below:

10. Do you consider yourself a Democrat, a Republican, an Independent, or a member of another party? Democrat Republican Independent Another Party Don't know Prefer not to say

Key Issues and Concerns

11. Rate these issues in order of importance to you (1 being most important and 7 being the least important): (use multiple choice grid with 7 rows and 7 columns) Abortion Immigration The state of democracy/corruption Foreign policy The economy Character Climate change?
12. If you had to assign a value from 1 to 5 to your level of optimism about the future of the United States, where 1 means highly pessimistic and 5 means highly optimistic, where would you place yourself? 1 (Highly pessimistic) 2 3 4 5 (Highly optimistic)

13. What would you say to someone who is undecided about voting in this election?

Confirmation Message

Thank you for your response! Your answers have successfully been recorded.

Link to survey: https://forms.gle/h7MTA8k21ZbYxahT6

B Pollster methodology overview and evaluation

B.1 Overview

The Siena College/New York Times (Times/Siena) poll is a collaboration between the New York Times and the Siena College Research Institute that aims to capture voter sentiment in key battleground states and at a national level (The New York Times 2024b). Its accuracy has made it a critical poll for predicting U.S. election results. The NYT/Siena poll uses live poll interviews to reach a representative sample of voters and measure their political preferences, views, and issue priorities (The New York Times 2024b).

B.2 Population and Sampling Frame

The target population for the Times/Siena poll is registered voters and likely voters in the United States. Some Times/Siena polls have a target population of registered voters, while others use likely voters (The New York Times 2024a). They put an emphasis on gathering national representations and state-specific samples in the seven key battleground states that decided the 2020 presidential election and are expected to decide the 2024 presidential election: Arizona, Nevada, Georgia, North Carolina, Michigan, Wisconsin, and Pennsylvania (The New York Times 2024b). The sampling frame is drawn from various voter registration databases; this helps ensure precise targeting and that the sample reflects the demographics of likely voters in the upcoming U.S. election (The New York Times 2024b).

B.3 Respondent Recruitment Strategy

Times/Siena recruits respondents for its poll at random, from a national list of registered voters (The New York Times 2024b)

B.4 Sampling Approch and Trade-offs

B.5 Strengths and Limitations of Methodology

B.5.1 Strengths

B.5.2 Limitations

C Additional data details

C.1 Data Cleaning

C.2 Polling averages from before President Joe Biden ended his re-election campaign
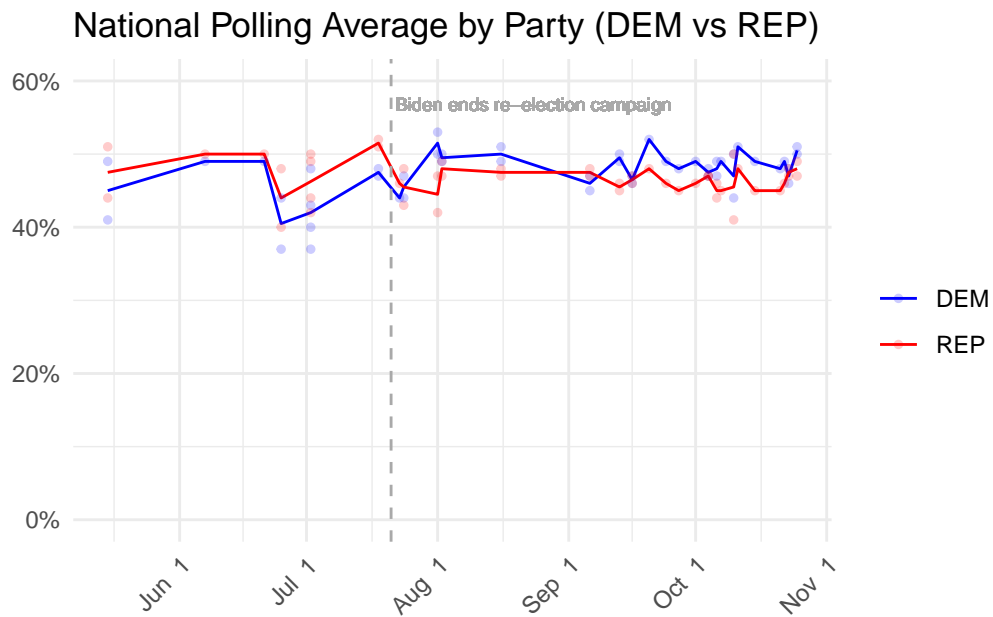

Figure 5: National popular vote averages for the Democratic and Republican presidential nominees since May 5, 2024 (six months before election day).

D Model details

Model summary is shown in Table 4

Table 4: Explanatory models of support for Harris and Trump based on end date, state, pollster, and pollscore

|	Harris	Trump
(Intercept)	48.99	48.59
	(5.02)	(5.21)
ns(end_date_num, df = 5)1	0.49	0.96
	(1.10)	(1.04)
ns(end_date_num, df = 5)2	1.25	−0.58
	(1.41)	(1.44)
ns(end_date_num, df = 5)3	0.83	0.59
	(1.47)	(1.45)
ns(end_date_num, df = 5)4	2.74	−0.98
	(2.22)	(2.11)
ns(end_date_num, df = 5)5	−0.87	2.45
	(1.10)	(1.05)
stateFlorida	−4.13	4.45
	(1.19)	(1.21)
stateGeorgia	−0.81	0.61
	(0.93)	(0.96)
stateMichigan	1.57	−2.18
	(0.86)	(0.88)
stateMinnesota	4.10	−5.70
	(1.82)	(1.79)
stateMissouri	−5.98	4.82
	(1.82)	(1.79)
stateMontana	−5.88	7.63
	(1.41)	(1.37)
stateNational	1.66	−2.81
	(0.71)	(0.72)
stateNebraska	−5.55	4.29
	(1.46)	(1.55)
stateNebraska CD-2	6.75	−6.76
	(1.09)	(1.12)
stateNevada	0.42	−0.91
	(1.20)	(1.16)
stateNew Hampshire	3.28	−3.88
	(1.75)	(1.87)
stateNorth Carolina	0.74	−0.64
	(0.96)	(0.93)
stateOhio	−2.29	1.41
	(1.10)	(1.09)
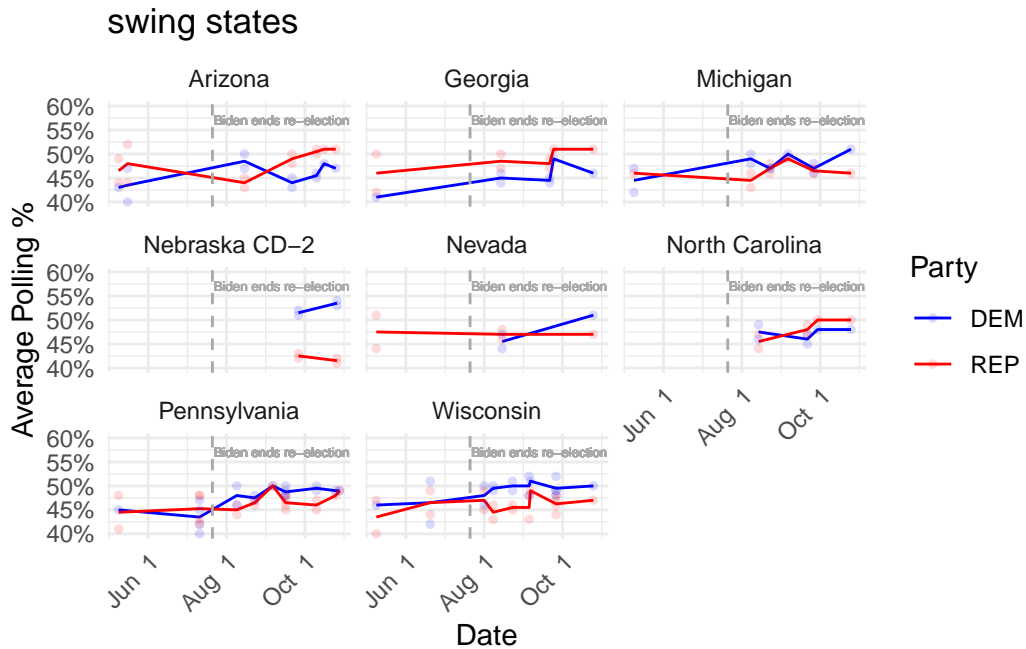
Figure 6: seven battleground states

D.1 Posterior predictive check

D.1.1 Harris Model

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

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

Examining how the model for support for Harris fits, and is affected by, the data

D.1.2 Trump Model

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

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

Examining how the model for support for Trump, and is affected by, the data

D.2 Model Diagnostics

D.2.1 Harris Model

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

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

Checking the convergence of the MCMC algorithm for the Harris model

D.2.2 Trump Model

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

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

Checking the convergence of the MCMC algorithm for the Harris model

References

Alexander, Rohan. 2024. *Telling Stories with Data.* "University of Toronto". https://www.tellingstorieswithdata.com.

FiveThirtyEight. 2024. "National Presidential Polls, 2024." https://projects.fivethirtyeight.com/polls/national/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm/.

L2. 2024. "DataMapping: Winning Strategies Start with the Best Voter Records." L2 Better Data Better Decisions. https://www.l2-data.com/datamapping/.

Morris, G. Elliott. 2024. "How 538's Pollster Ratings Work." ABC News. https://abcnews.go.com/538/538s-pollster-ratings-work/story?id=105398138.

Neyman, Jerzy. 1934. "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection." *Journal of the Royal Statistical Society* 97 (4): 558–625. http://www.jstor.org/stable/2342192.

Pew Research Centre. 2024. "Harris, Trump Voters Differ over Election Security, Vote Counts and Hacking Concerns." Pew Research Centre Election 2024. https://www.pewresearch.org/politics/2024/10/24/harris-trump-voters-differ-over-election-security-vote-counts-and-hacking-concerns/.

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

The New York Times. 2024a. *Cross-Tabs: Late October 2024 Times/Siena Poll of the Likely Electorate.* The New York Times. https://www.nytimes.com/interactive/2024/10/26/us/elections/times-siena-nyc-poll-registered-voter-crosstabs.html.

———. 2024b. "You Ask, We Answer: How the Times/Siena Poll Is Conducted." The New York Times. https://www.nytimes.com/article/times-siena-poll-methodology.html#link-2fc0eb29.