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

Forecasting the 2024 U.S. Presidential Election using FiveThirtyEight Polling Data and Poll of Polls Methodology Resulted in ...

Talia Fabregas Fatimah Yunusa Aliza Mithwani

October 31, 2024

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

1 Introduction

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 code and data used to perform this presidential election forecast can be found at: https://github.com/taliafabs/USPresidentialPollingForecast2024.git.

The remainder of this paper is structured as follows. Section 2 contains an overview of the dataset, summary statistics and visualizations. Section 3 contains the Bayesian regression used to … Section 4 contains the results of applying the Bayesian regression model. Section 5. Appendix A

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

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

2.3.1 National and swing state polling averages over time

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 between Vice President Harris and former President Trump in the seven 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
New Hampshire	52.0	45.0
Ohio	45.0	52.0
Texas	41.0	51.5
Virginia	52.0	44.0

This takes into account date and state.

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.)

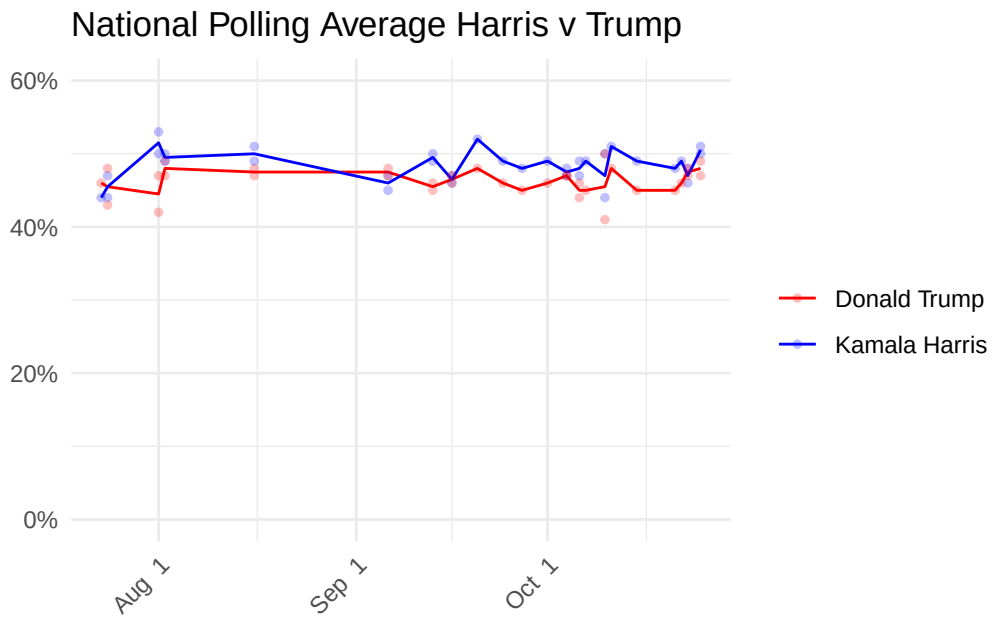


Figure 1: National popular vote polling averages for Harris and Trump since Biden ended his re-election campaign on July 21, 2024. As of October 29, 2024, Harris has a narrow lead over Trump.

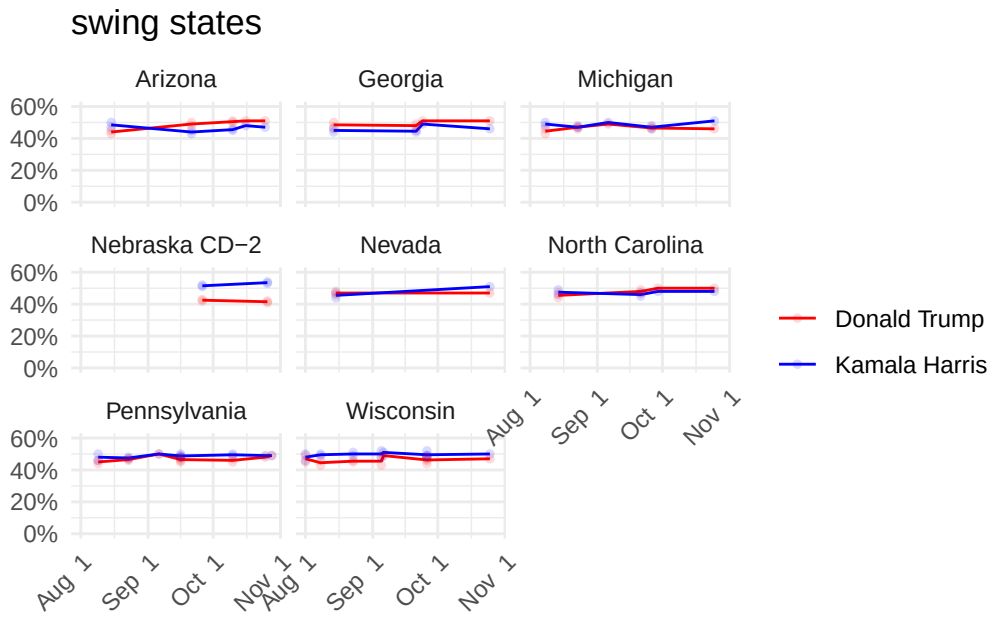


Figure 2: seven battleground states with harris

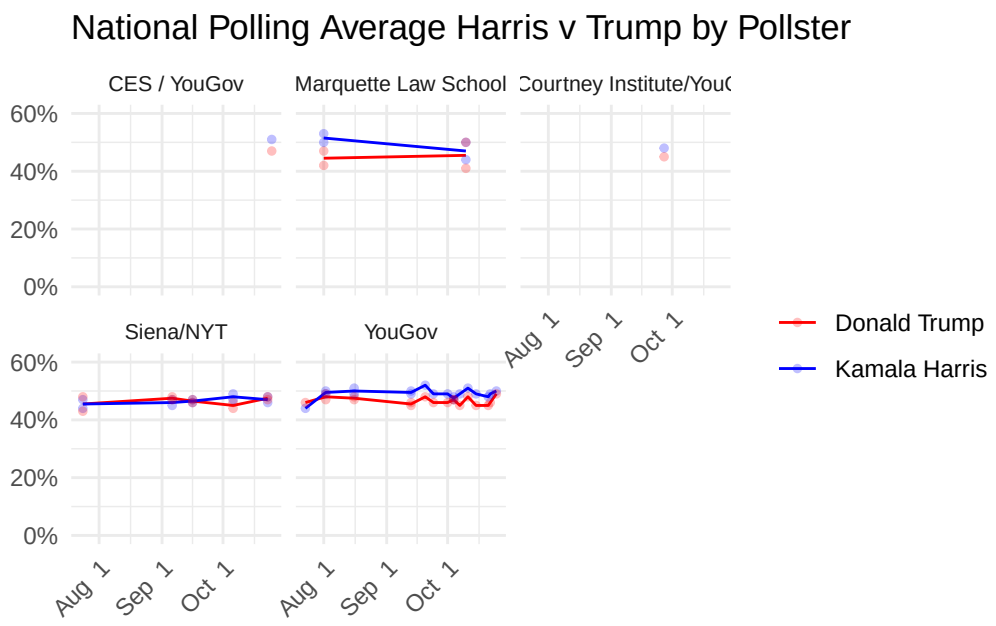


Figure 3: Pollsters caption

2.3.3 Methodology

After conducting further exploratory data analysis, we decided not to include methodology in our model because it is strongly related to pollster. Using pollster as a predictor sufficiently accounts for methodology. We discuss this decision in depth in Section 3. Different pollsters tend to use different methodologies.

2.3.4 Pollscore

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. We 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.

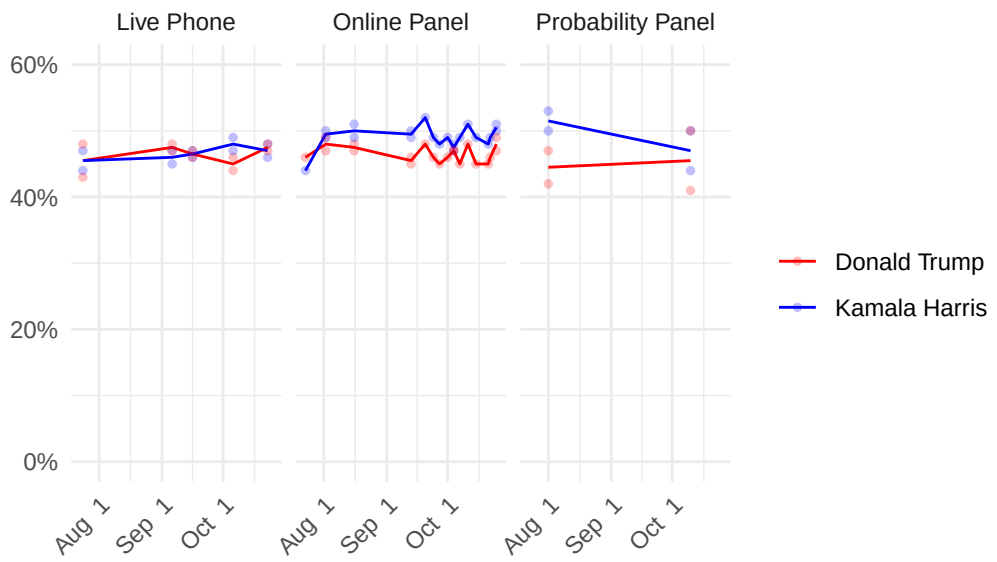


Figure 4: National polling averages for Kamala Harris and Donald Trump since July 21 have varied by methodology. Live phone polls had Trump leading by a narrow margin until mid-September, online panel polls have had Harris in the lead since early August, and probability panel polls have shown a shrinking lead for Harris since shortly after Predient Biden ended his campaign.

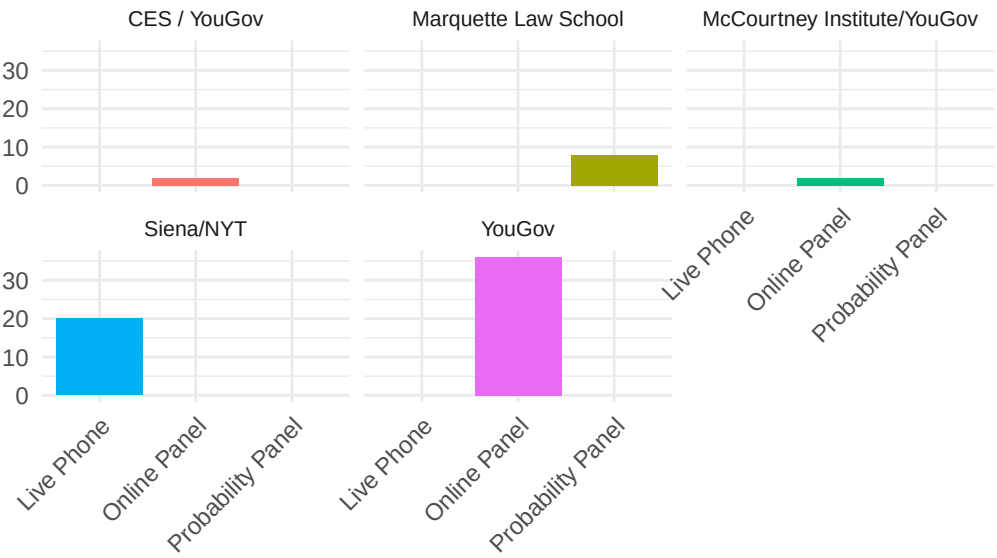


Figure 5: In polls conducted after Vice President Harris became the Democratic nominee, every pollster used a distinct methdology.

Here we briefly describe the Bayesian analysis model used to investigate… Model details and diagnostics are included in Appendix D.

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`.

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 Model results are in the appendix. **?@tbl-modelresults**.

4.1 National Popular Vote Results

Popular vote prediction after applying

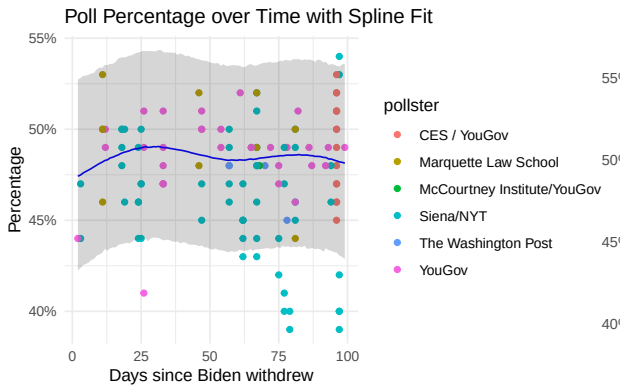


Figure 6: caption

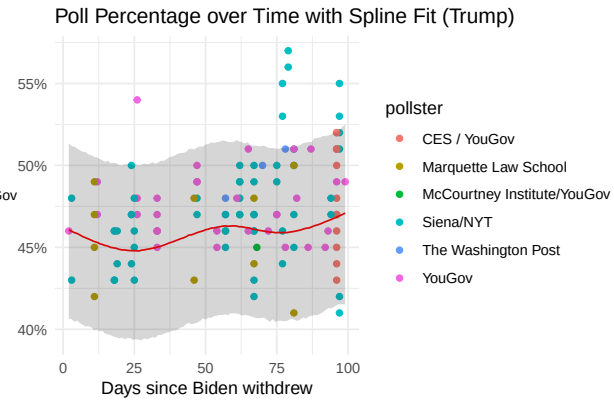
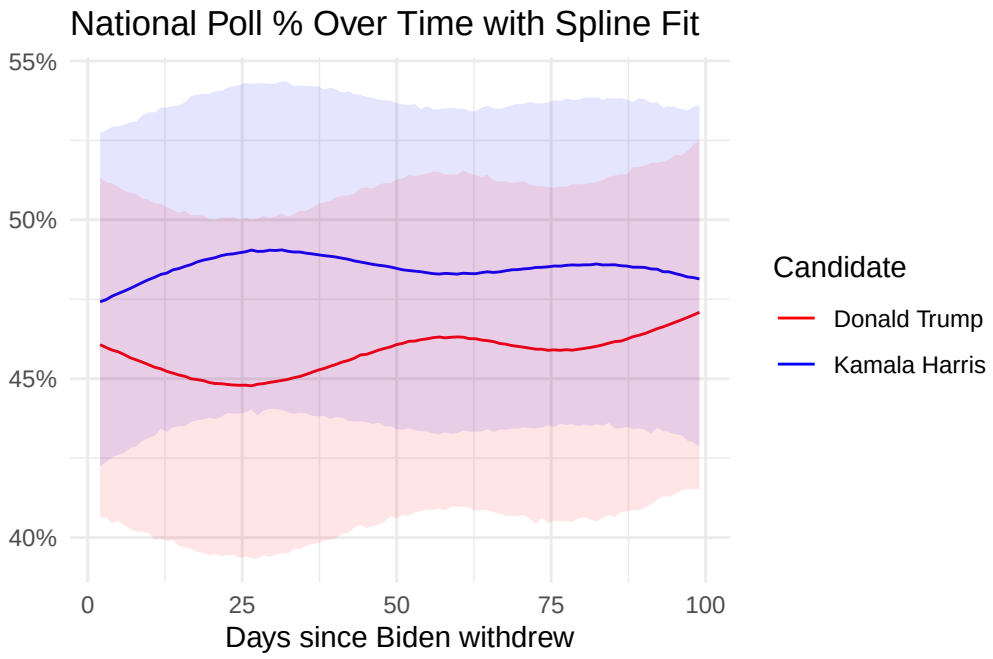
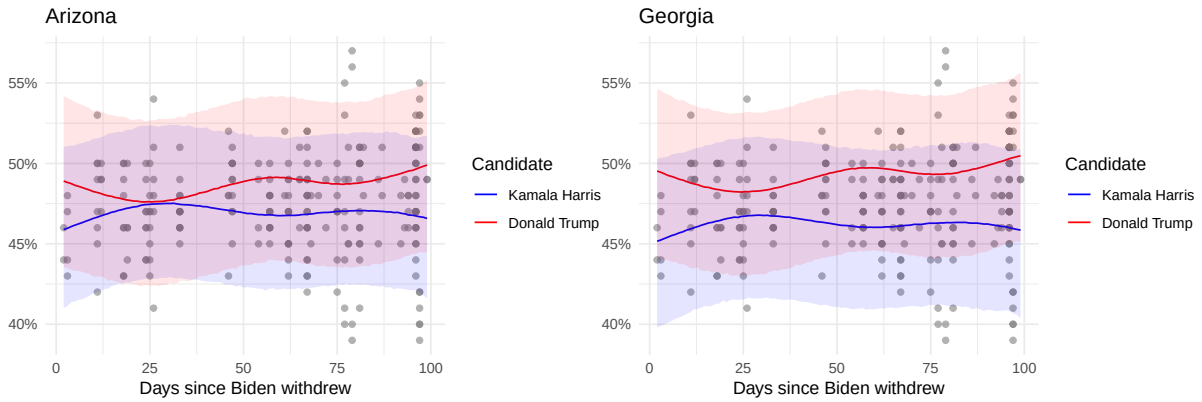


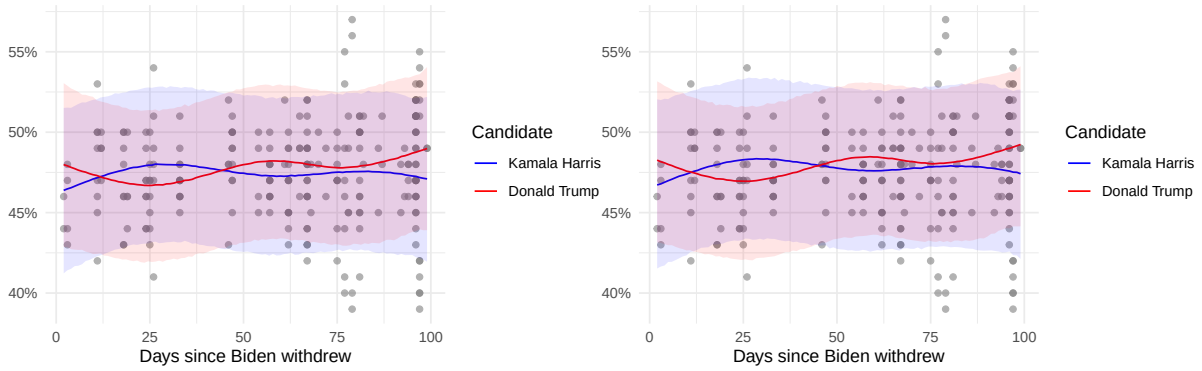
Figure 7: caption





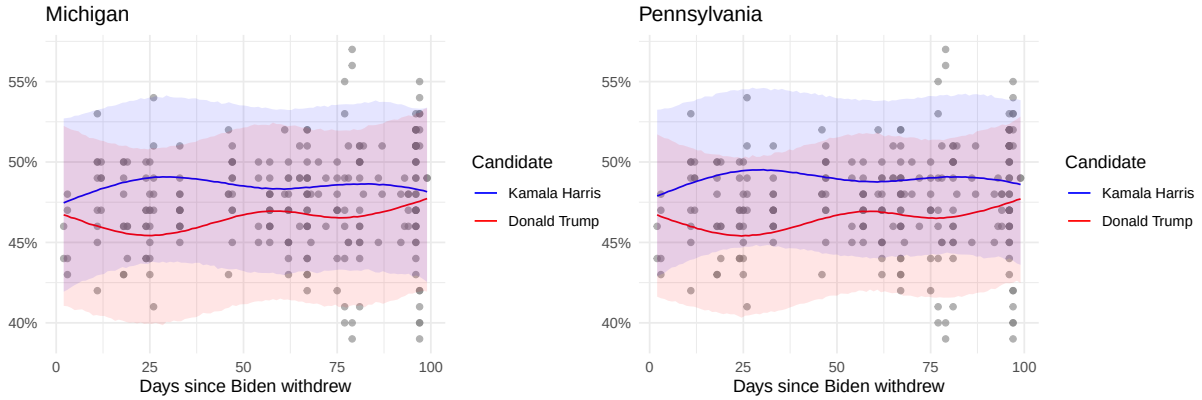
(a) Arizona: Trump +1.61

(b) Georgia: Trump +3.01



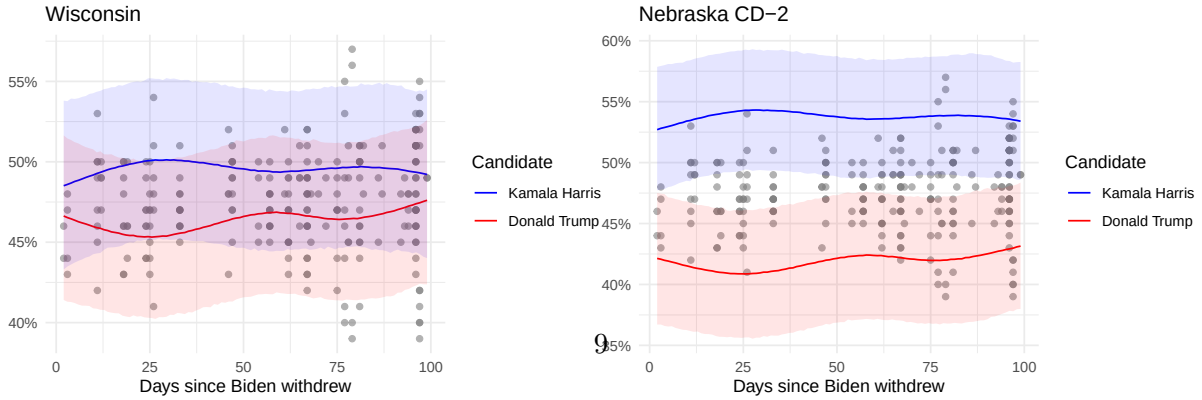
(c) Nevada: Trump +0.37

(d) North Carolina: Trump +0.44



(e) Michigan: Harris +2.22

(f) Pennsylvania: Harris +2.57



(g) Wisconsin: Harris +3.31

(h) Nebraska CD-2: Harris + 11.87

Figure 8: On average, Trump has had narrow leads in Arizona, Georgia, Nevada, and North Carolina, while Harris has narrow leads in Michigan, Pennsylvania, and Wisconsin.

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

5 Discussion

5.1 First discussion point

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

5.2 Second discussion point

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

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Idealized methodology

A.1 Idealized survey

B Pollster methodology overview and evaluation

C Additional data details

C.1 Data Cleaning

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

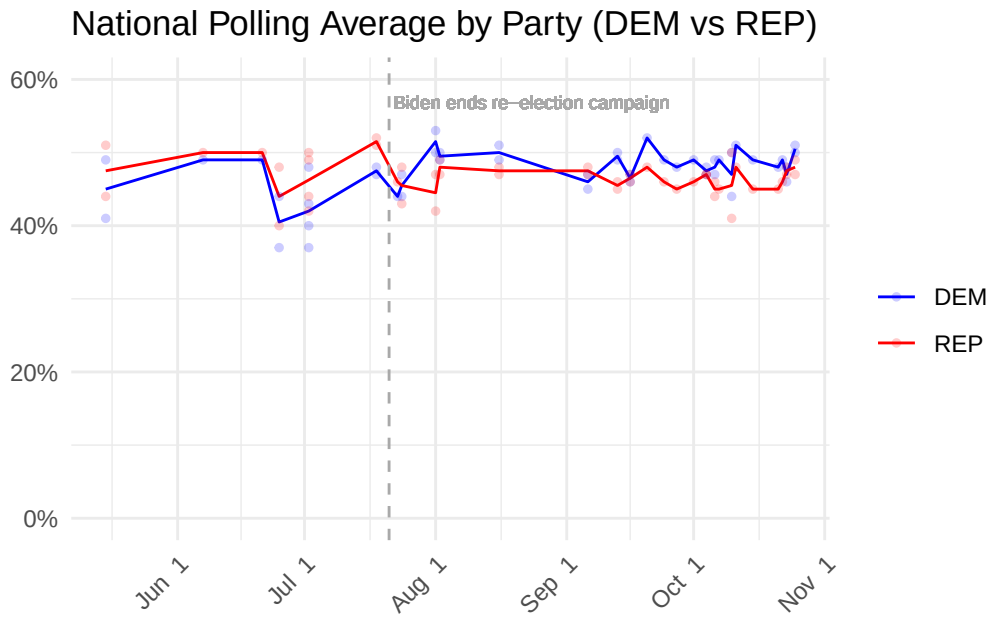


Figure 9: National popular vote averages for the Democratic and Republican presidential nominees since May 5, 2024 (six months before election day).

D Model details

Model summary is shown in Table 2

Table 2: Explanatory models of support for Harris and Trump based on …

	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)
statePennsylvania	2.06	−2.19

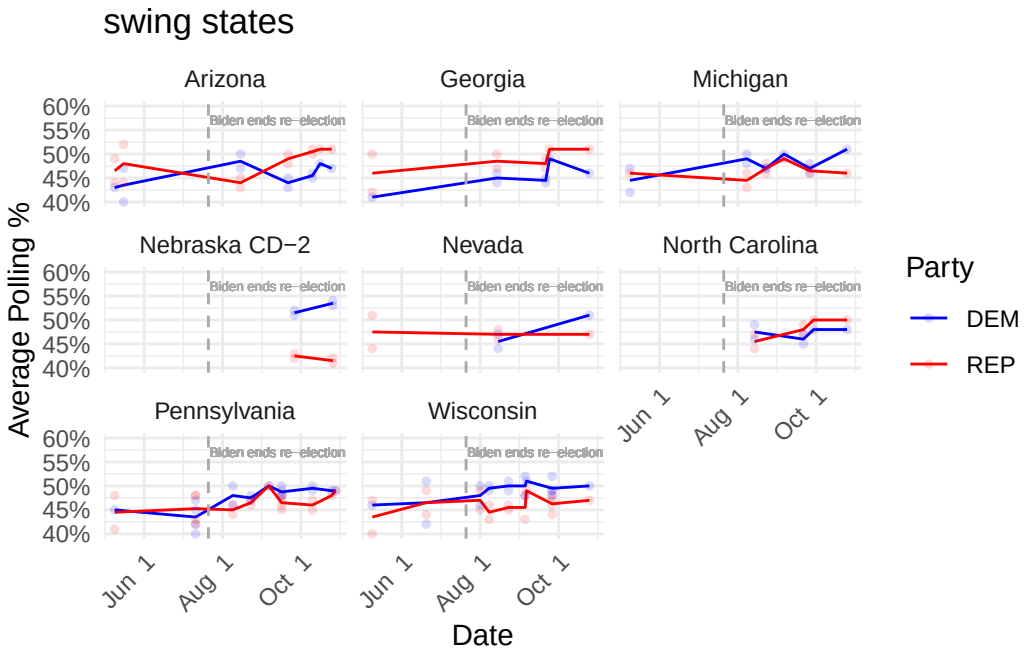


Figure 10: seven battleground states

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

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/.

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