# Kamala Harris Rises as the 2024 Election Front Runner: A Poll-of-Polls Forecast*

## How Poll Aggregation Reveals US Presidental Race Trends[!!!UPDATE IT TO MAIN RESULTS!!!]

Yunkyung Ko

October 21, 2024

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

## Table of Contents



*Code and data are available at: https://github.com/koyunkyung/us_election_2024.

# 1 Introduction

The US presidential election is an event that receives a lot of international attention due to its far-reaching implications beyond the country's borders. The effects of US elections are not only related to economy and international relations around the world, but also link to social and environmental issues such as climate change (Bijune and Ha 2024). In the anticipation of the 2024 US electoral competition, this paper is aimed at predicting possible outcomes of the election by analyzing the level of support that Kamala Harris will gain.

We forecast the support of Kamala Harris based on the polling results at the national and state levels, and apply a linear regression and a Bayesian approach. The main parameter of interest is the proportion of vote or support that Harris received in surveys, which is traced over time. By considering the effect of changes in poll-making organizations and geographical distinctions, our objective is to correct for variation across different voter bases with the pooling the polls approach (Jackman 2024).

Our initial linear model examining the support for Harris over time suggests that the rate of support remained relatively stable around 47%, with no significant increase or decrease. However large variability was detected, as seen in the spread of points around the fitted line. Consequently, pollster-specific effects and state-level random effects were added to the model but resulted in even higher variations. Some pollsters or states consistently reported higher or lower support for Harris compared to others, proposing that [!!!!!MEANINGFUL RESULT ADDED LATER!!!!!].

These prognostics provide much more than merely forecasting the election in question. Hinting the trajectory the US might take in matters of foreign policy, economics, and global politics, predictions enable stakeholders worldwide to take preemptive measures for the resulted changes of a newly elected government (CSIS 2024). As such, this study not only contributes to the domestic political discourse but also provides a valuable tool for global actors seeking to navigate the uncertainty surrounding the 2024 US presidential election(CSIS 2024).

The paper is structured as follows. Initially, Section 2 and Section 3 explores the data and methodology used, including filtering and modeling techniques applied to the polling data. Following that, Section 4 presents the results from the linear and Bayesian models, while the next section Section 5 discusses the broader implications of these findings. Finally, the paper concludes with remarks on future directions for research and applications of these models Section 5.

# 2 Data

## 2.1 Overview

We use the statistical programming language R (R Core Team 2023) to analyze US presidential polling data from FiveThirtyEight (FiveThirtyEight 2024), focusing on support for Kamala Harris. The dataset includes 15831 poll results from various national and state-level polls, with key variables such as pollster, sample size, percentage of support for Harris, and end date of the poll. Following the guidance of Alexander (2023), we compiled the results of each opinion poll over a period of time and compared them taking into account the methodological peculiarities of polling by pollsters and geographical scope of the conducted polls.

To ensure data quality, we filtered the dataset to include only polls that measured Kamala Harris' support, with a numeric grade of pollster 2.7 or higher for reliability. We also limited the analysis to polls conducted after July 21, 2024, when Harris officially declared her candidacy, and excluded pollsters with fewer than 35 polls to focus on those with sufficient data for robust results.

In performing the analysis, we utilized several R packages. Wickham et al. (2019) was used for data manipulation and visualization and Goodrich et al. (2024), Arel-Bundock (2022) was respectively used for Bayesian modeling and generating model summaries. For visualizing results, Wickham (2016) was used and Zhu (2024) helped format tables for presentation. These packages provided a framework for efficient data processing, modeling, and reporting.

## 2.2 Measurement

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

## 2.3 Outcome variables

### 2.3.1 pct: the percentage of the vote or support that the candidate received in the poll

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

Some of our data is of penguins (**?@fig-bills**), from (**palmerpenguins?**).

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.)

Talk way more about it.

### 2.4 Predictor variables

#### 2.4.1 end_date: the date the poll ended

#### 2.4.2 pollster: the polling organization that conducted the poll

mention that it was filtered by numeric_grade

#### 2.4.3 pollscore: the score or reliability of the pollster in question

Add graphs, tables and text.

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

## 3 Model

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

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

### 3.1 Model set-up

Define $y_i$ as the number of seconds that the plane remained aloft. Then $\beta_i$ is the wing width and $\gamma_i$ is the wing length, both measured in millimeters.

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

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2024). 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 $\theta$.

# 4 Results

Our results are summarized in Table **??**.

# 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 Additional data details

# B Model details

## B.1 Posterior predictive check

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

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

Examining how the model fits, and is affected by, the data

## B.2 Diagnostics

**?@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

## References


Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bijune, Aiste, and Lan Ha. 2024. "US 2024 Election: Implications for the Global Economy." *Euromonitor International*. https://www.euromonitor.com/article/us-2024-election-implications-for-the-global-economy.

CSIS. 2024. "The Global Impact of the 2024 u.s. Presidential Election." https://features.csis.org/2024-us-election-global-impact/.

FiveThirtyEight. 2024. "Dataset: US Presidential General Election Polls." https://projects.fivethirtyeight.com/polls/data/president_polls.csv.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Jackman, Simon. 2024. "Pooling the Polls over an Election Campaign." chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cdn-uploads.piazza.com/paste/ket0p3a9re9628/8a55272a479e1aaeabbd01b3030aa1d38c4a7fc323a9c3feee3bd5fec2fdca44/Pooling_the_polls_over_an_election_campaign.pdf.

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

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

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

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.