Analysis of 2024 U.S. Presidential Election Polling Trends*

A Comparison of Trump and Harris

Jiaxuan Song Zien Gao Shuheng (Jack) Zhou

October 21, 2024

This report analyzes the polling trends for Donald Trump and Kamala Harris in the 2024 U.S. Presidential election using data from FiveThirtyEight's 2024 National Presidential Polls. The study focuses on understanding the changes in support for both candidates over time, across different regions, and how polling methodologies affect their respective support rates. The results show that Trump's support base remains relatively stable throughout the election period, particularly in Republican-leaning states, while Harris exhibits greater fluctuations, especially during mid-2023, with notable regional differences. Additionally, the impact of different polling methods is explored, showing how online and phone surveys may influence the results. These findings highlight the importance of swing states and the variability across pollsters in shaping election forecasts. The report concludes with a discussion on the factors that might influence the outcome of the election and the need for continuous monitoring of polling trends as election day approaches

1 Introduction

The 2024 U.S. Presidential election is one of the most closely watched political events, with intense competition between candidates. Donald Trump, as a former president, is seeking to reclaim the presidency, while Kamala Harris, the current Vice President, is a prominent figure in the Democratic Party. Polls play a crucial role in predicting election outcomes, offering insight into public support for each candidate. This report analyzes polling data from FiveThirtyEight's 2024 National Presidential Polls, focusing on the trends in support for Trump and Harris. We will explore how time, regions, and pollsters influence the support rates for both candidates.

*Code and data are available at: https://github.com/Jiaxuan-Song/U.S-Election-Prediction-2024.git.

2 Data

2.1 Overview

The dataset used for predicting the 2024 U.S. Presidential election is sourced from various public opinion polls conducted by reputable polling organizations throughout the United States. These polls provide detailed information such as pollster names, polling methodology, sample sizes, candidate names, political party affiliations, and percentage support for each candidate, specifically for key figures like Kamala Harris and Donald Trump. Additionally, the dataset includes important poll characteristics such as whether the poll was conducted at a national level, the pollster's rating, and whether the poll involved hypothetical matchups between candidates.

We utilize the statistical programming language R (R Core Team 2023) for data cleaning, preprocessing, and analysis. The dataset is processed to remove inconsistencies and missing data, ensuring it is suitable for forecasting election trends. Following Alexander (2023), we consider he codes used and some graphics .

Overview text

2.2 Measurement

The dataset reflects real-world polling data gathered through various methods such as phone surveys, online interviews, and ranked-choice voting reallocation. Each poll entry represents the percentage of respondents supporting a specific candidate at a particular time. Polling organizations, such as TIPP Insights, Quantus Insights, and ActiVote, are included, and the dataset captures details about when and where the poll was conducted, the poll's methodology, and whether it was a general election or hypothetical scenario. Data is collected and categorized based on responses related to key candidates (e.g., Kamala Harris and Donald Trump), with percentage values representing the share of voters supporting each candidate at the time of polling. After cleaning and verifying the data for consistency and accuracy, the final dataset will be used to model voting trends leading up to the election.

2.3 Outcome variables

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 (Figure 1), from Horst, Hill, and Gorman (2020).

Talk more about it.

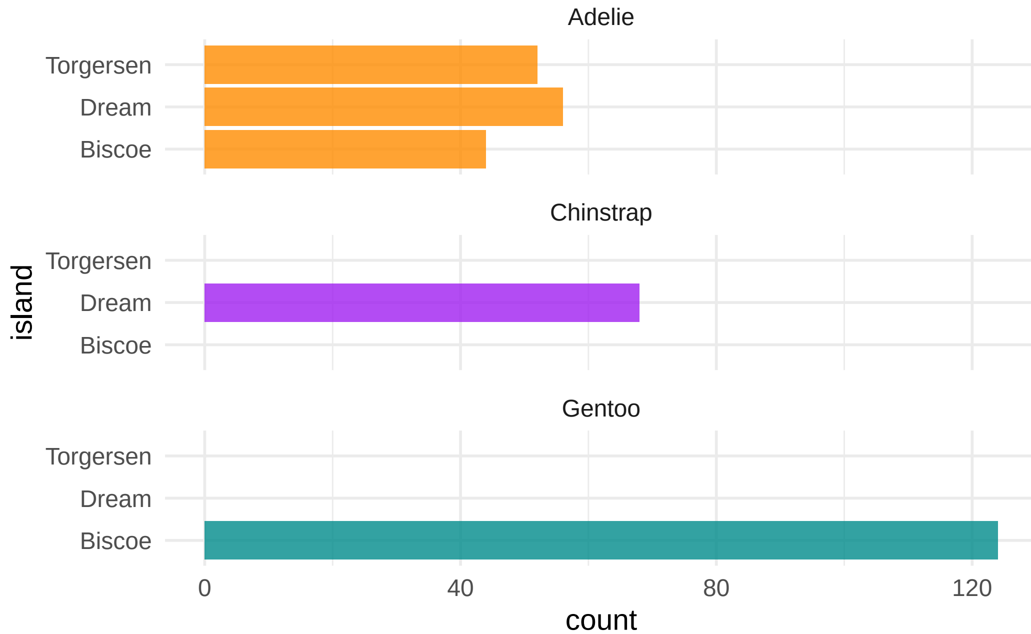


Figure 1: Bills of penguins

And also planes (Figure 2). (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

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.

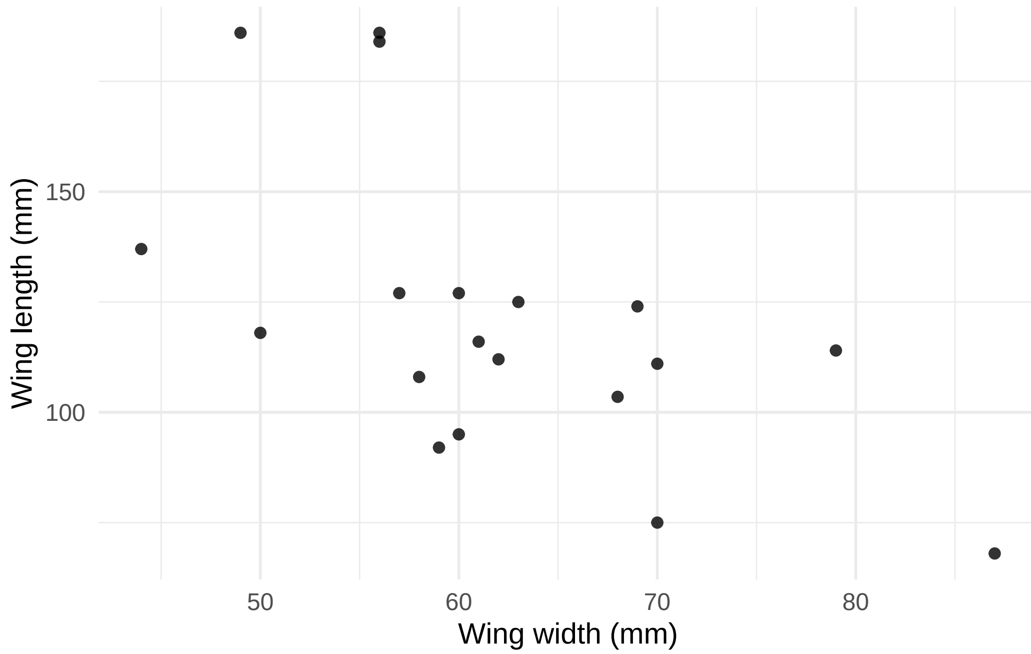


Figure 2: Relationship between wing length and width

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.

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

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…

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

	First model
(Intercept)	1.12
	(1.70)
length	0.01
	(0.01)
width	−0.01
	(0.02)
Num.Obs.	19
R2	0.320
R2 Adj.	0.019
Log.Lik.	−18.128
ELPD	−21.6
ELPD s.e.	2.1
LOOIC	43.2
LOOIC s.e.	4.3
WAIC	42.7
RMSE	0.60

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

4 Results

Our results are summarized in Table 1.

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

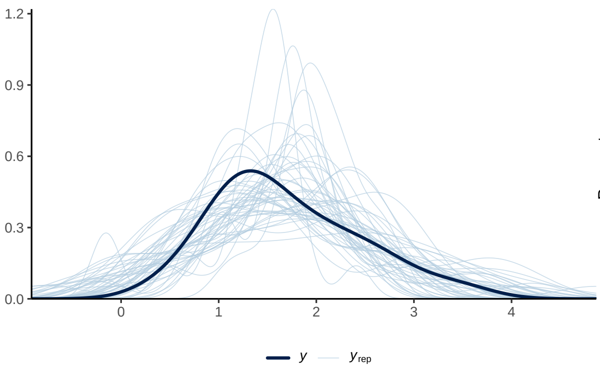
A Additional data details

B Model details

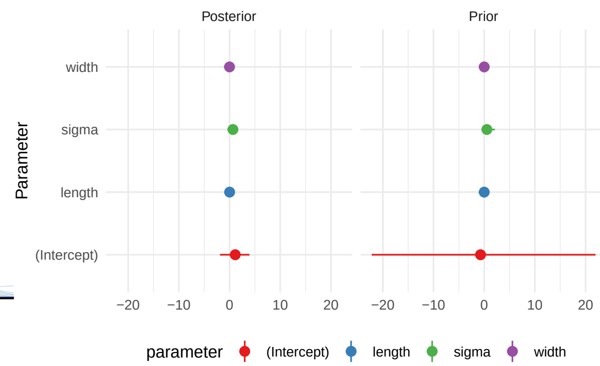
B.1 Posterior predictive check

In Figure 3a we implement a posterior predictive check. This shows...

In Figure 3b we compare the posterior with the prior. This shows...



(a) Posterior prediction check



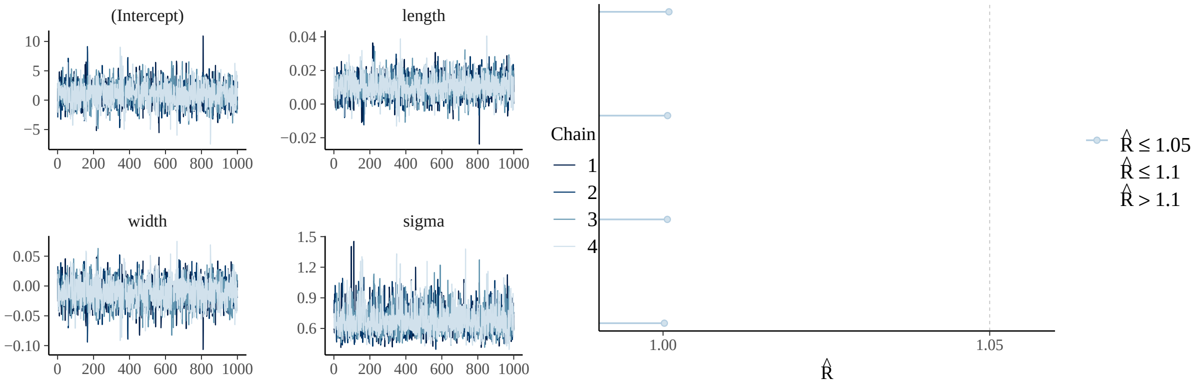
(b) Comparing the posterior with the prior

Figure 3: Examining how the model fits, and is affected by, the data

B.2 Diagnostics

Figure 4a is a trace plot. It shows... This suggests...

Figure 4b is a Rhat plot. It shows... This suggests...



(a) Trace plot

(b) Rhat plot

Figure 4: Checking the convergence of the MCMC algorithm

References

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

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

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*. https://doi.org/10.5281/zenodo.3960218.

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