2024 US Presidential Election Model*

My subtitle if needed

Jimin Lee Sarah Ding Xiyan Chen

October 19, 2024

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

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

```
# Load the cleaned poll data
clean_poll_data <- read_csv("/Users/jamielee/2024_US_Elections/data/02-analysis_data/analysis
```

```
Rows: 629 Columns: 52
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (23): pollster, sponsors, display_name, pollster_rating_name, methodolog...
dbl (14): poll_id, pollster_id, sponsor_ids, pollster_rating_id, numeric_gra...
lgl (15): sponsor_candidate_id, sponsor_candidate, sponsor_candidate_party, ...
```

*Code and data are available at: https://github.com/jamiejiminlee/2024_US_Elections.git.

```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Group by state and candidate, and calculate the average percentage (pct)
state_average <- clean_poll_data |>
  group_by(state, candidate_name) |>
  summarise(avg_pct = mean(pct, na.rm = TRUE)) |>
  pivot_wider(names_from = candidate_name, values_from = avg_pct) |>
  rename(Trump_pct = "Donald Trump", Harris_pct = "Kamala Harris")
```

```
`summarise()` has grouped output by 'state'. You can override using the
`.groups` argument.
```

```r
# Replace missing values only in the numeric columns (Trump_pct, Harris_pct)
state_average <- state_average |>
  mutate(across(c(Trump_pct, Harris_pct), ~replace_na(., 0)))
```

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data (Toronto Shelter & Support Services 2024)…. Following Alexander (2023), we consider…

Overview text

2.2 Measurement

To predict the election outcome, we will assume that the higher pct in a state between Trump and Harris determines the winner of the election. The difference in pct is used as the predictor for which candidate wins each state.

2.3 Outcome variables

?@fig-averagetable provides an overview of the average support for Trump and Harris across all states, along with the predicted winner for each state based on our model.

```
# A tibble: 16 x 5
# Groups:   state [16]
   state          Trump_pct Harris_pct winner predicted_winner
   <chr>              <dbl>      <dbl>  <dbl> <chr>
 1 Arizona             46.5       45.3      1 Trump
 2 Colorado            40          0        1 Trump
 3 Florida             53         40.5      1 Trump
 4 Georgia             46.6       44.4      1 Trump
 5 Michigan            44.5       46.3      0 Harris
 6 Missouri            54         41        1 Trump
 7 Montana             56.5       39        1 Trump
 8 Nebraska CD-2       42         50.8      0 Harris
 9 Nevada              46.6       43.8      1 Trump
10 North Carolina      46.8       46.8      0 Harris
11 Ohio                49.5       44.2      1 Trump
12 Pennsylvania        45.1       47        0 Harris
13 Texas               47.1       43.6      1 Trump
14 Virginia            41         45.2      0 Harris
15 Wisconsin           45.2       48.8      0 Harris
16 <NA>                44.5       46.6      0 Harris
```

Figure 1 illustrates the average polling support for Donald Trump and Kamala Harris in each state. The side-by-side comparison represents the percentage of support each candidate has received, based on aggregated poll data from pollsters with high-quality scores. Trump's support is shown in red, while Harris's support is depicted in red.

Figure 2 visualizes the predicted winner of the 2024 US Presidential Election across all states, based on aggregated polling data. Each state is color-coded according to the candidate predicted to win - blue represents Donald Trump, and red represents Kamala Harris.

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,...

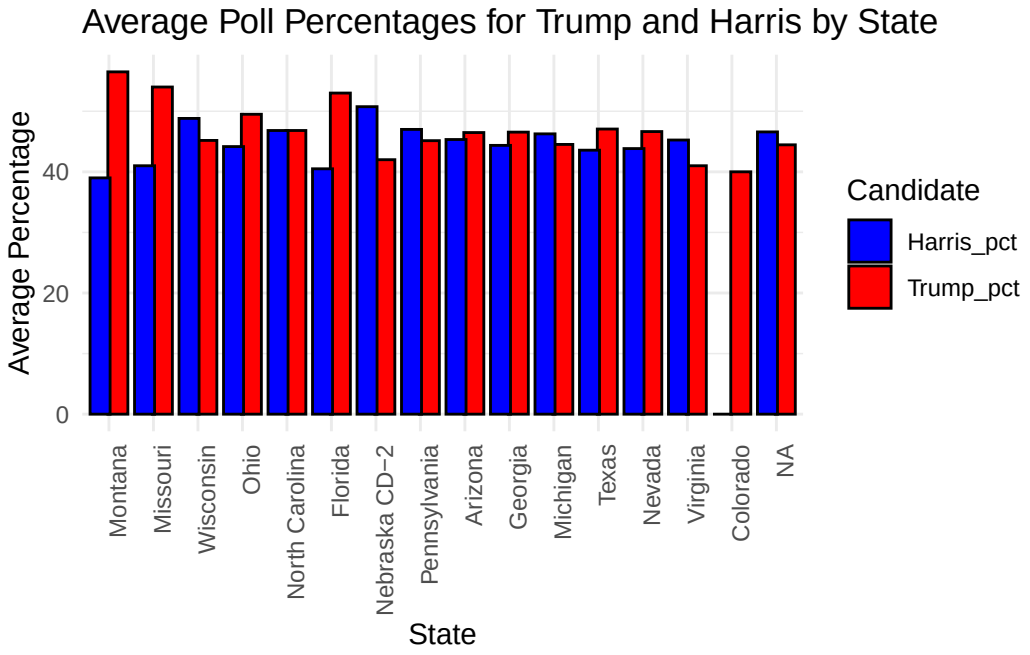

Figure 1: Average Poll Percentage by State

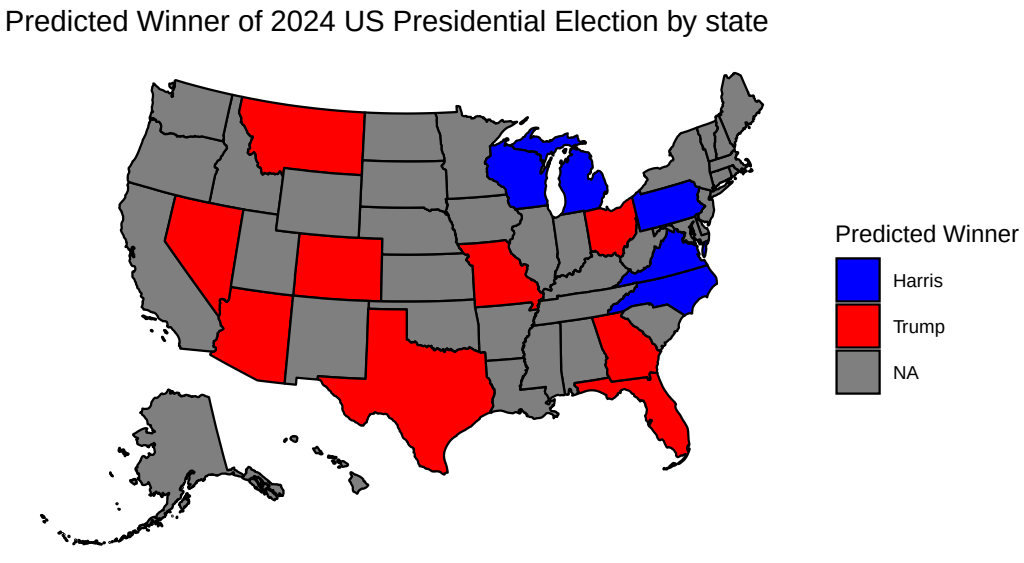

Figure 2: Predicted Election Winner by State

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 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 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 algo-
rithm

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/.

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

Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.