

Joe Biden Predicted to Win 59% of the Popular Vote in the 2024 U.S. Presidential Election*

A 2024 U.S. Election Forecast Using Multi-Level Regression with Post-Stratification (MRP)

Talia Fabregas

Fatimah Yunusa

March 15, 2024

The 2024 U.S. Presidential Election will take place on November 5th and it will be a rematch between President Joe Biden and Former President Donald Trump. Using multi-level regression with post-stratification (MRP) we found that ... Based on our results, we can predict that Joe Biden will win 59% of the popular vote.

1 Introduction

The 2024 U.S. Presidential Election will take place on Tuesday November 5 2024. Incumbent President Joe Biden will seek a second term in a rematch against his 2020 opponent and, arguably, the most polarizing figure in American politics, former President Donald Trump.

Packages used for data cleaning, model, visualizations, and tables: `tidyverse` Wickham et al. (2019), `janitor` Firke (2023), `rstanarm`, `readr`, `ggplot2`, `knitr`, and `arrow`.

The remainder of this paper is structured as follows. Section 2 discusses the survey and post-stratification data used.

Talk about polarization in a discussion point instead of introduction

Clear gap that needs to be filled ... what is the research gap and why is this important?

- 2024 US Presidential Election
- how consequential this election is

*Code and data are available at: <https://github.com/taliafabs/US-Election-Forecast-2024.git>

- what is on the ballot
- women's rights
- trump's project 2025 thing

If someone's done it before not on this data set not in this context

Previous groups have looked at ...

R Core Team (2023) was used

ESTIMAND: SUPPORT FOR BIDEN

We will use ... to estimate support for President Joe Biden in the upcoming election.

2 Data

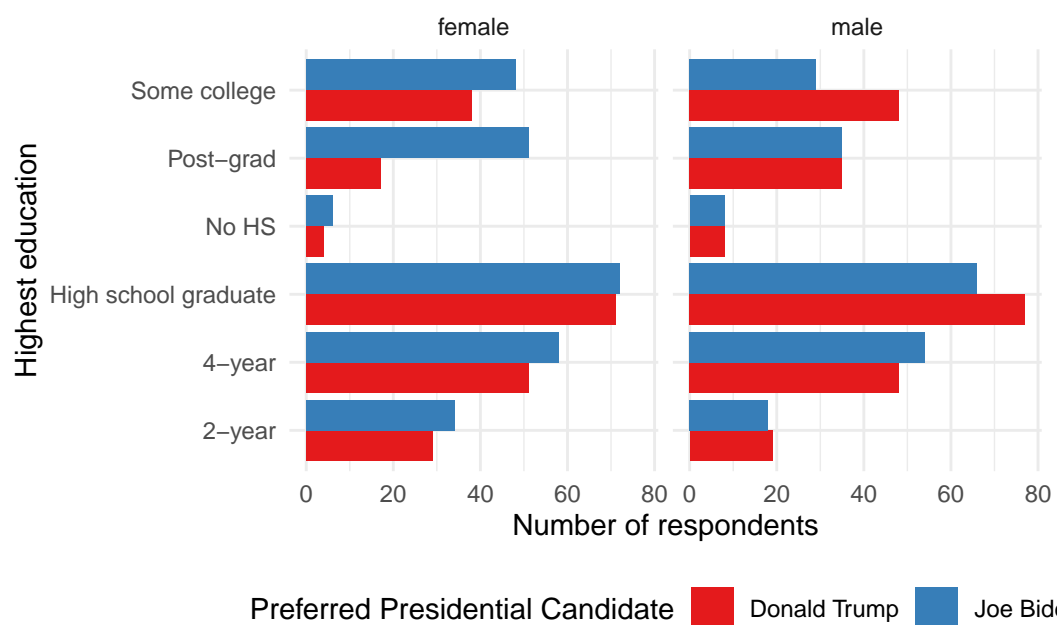


Figure 1: The Distribution of Presidential Preferences, by gender and highest level of education

2.1 Individual-Level Survey Data

Our survey data is from the America's Political Pulse Survey conducted by the Polarization Research Lab. The Polarization Research Lab is a research group founded by top political science research scholars at Dartmouth College, Stanford University, and the University of

Pennsylvania, dedicated to applying scientific research methods to the study of democracy and political polarization Iyengar, Lelkes, and Westwood (2024). The Polarization Research Lab studies political polarization across the United States by conducting the Political Pulse Survey each week and making the results available to anyone, in real time, via their website. The America’s Political Pulse Survey interviews 1000 American adults each week and asks them questions that aim to track affective polarization, support for violations of democratic norms, and support for the use of political violence in America Iyengar, Lelkes, and Westwood (2024). The Polarization Research Lab makes high-level results of their survey easily accessible and visible via their website’s dashboard.

The specific survey data set used is the America’s Political Pulse Survey Data Week 3 2024 January 12-19. It contains the responses of the 1000 American adults interviewed between January 12-19 2024. Demographic information about respondents, including sex, age, race, home state, employment status, marital status, and the size and type of city that they live in is included. Participants were also asked who they voted for in the 2016 and 2020 U.S. presidential characteristics, what their party affiliation is, and what their political ideology is.

However, this data set has a few major limitations. It contains very few respondents from Hawaii, Wyoming, and Kansas. It contains zero respondents from Vermont.

The survey does not contain a question about preferred 2024 Presidential Candidate. In fact, it was conducted in January 2024, before Super Tuesday took place and Donald Trump had secured almost enough delegates to win the Republican nomination. At the time when this survey was conducted, former President Donald Trump, former South Carolina Governor Nikki Haley, and Florida Governor Ron DeSantis were all GOP Presidential primary candidates. Media personality Vivek Ramaswamy ended his primary campaign on January 15th, however this was in no way a clear indication that Trump had secured the nomination. As of March 11, 2024, former President Trump has secured 1078 out of the 1215 delegates needed to win the GOP nomination.

Survey Data figures and tables go [here](#)

Within our survey data, the vote distribution between Trump and Biden varies by sex and race. As seen in Figure 2, Black females were more likely to vote for President Joe Biden (D-Delaware) than former President Donald Trump (R-Florida) in 2020, whereas white males were more likely to vote for former President Trump. This indicates that race might be a predictor of who an individual’s preferred presidential candidate is.

Table 1: Races of Survey Respondents

| Race | Count | Proportion |
|-------|-------|------------|
| asian | 18 | 0.0194805 |
| black | 93 | 0.1006494 |
| mixed | 25 | 0.0270563 |

| Race | Count | Proportion |
|-----------------|-------|------------|
| native american | 6 | 0.0064935 |
| other | 141 | 0.1525974 |
| white | 641 | 0.6937229 |

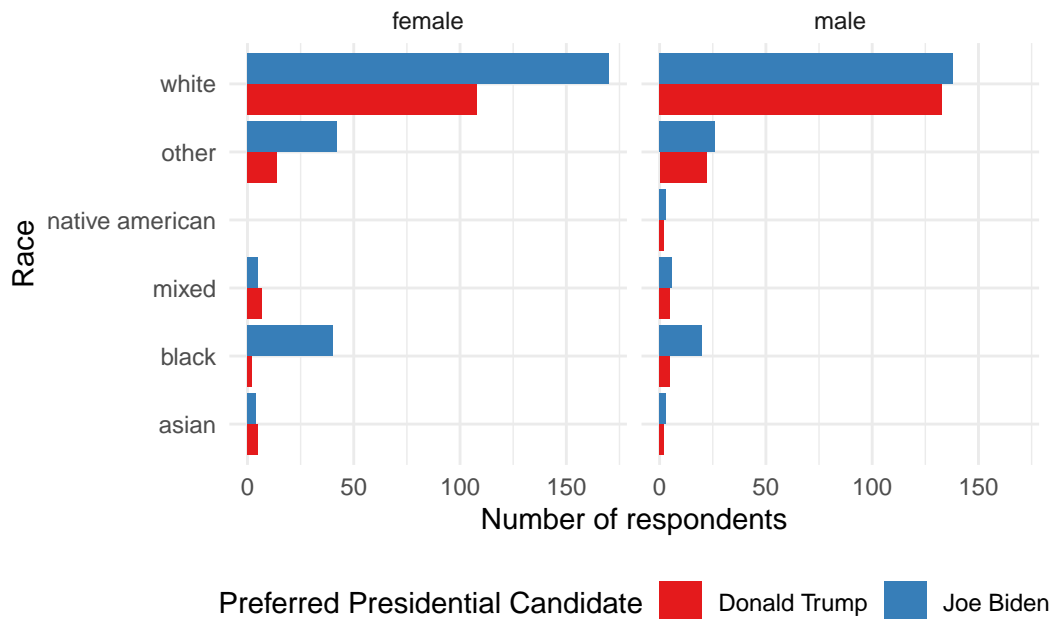
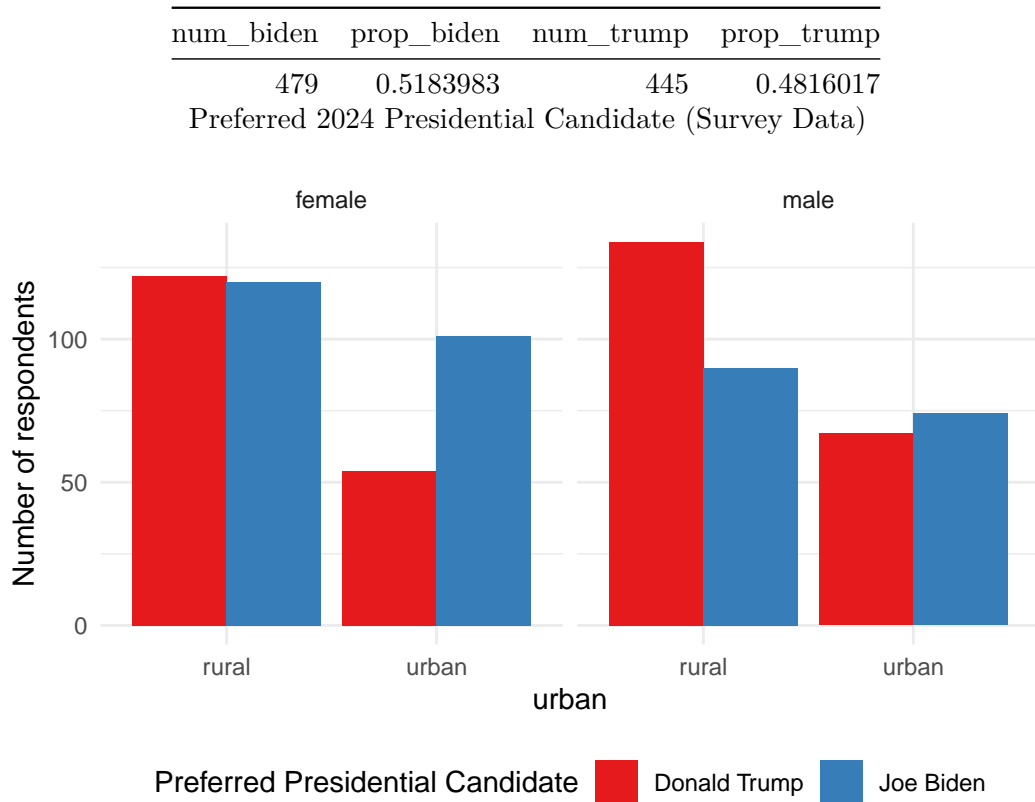


Figure 2: The Distribution of Presidential Preferences, by gender and race

```
# write about America's urban-rural divide
#| echo: false
#| warning: false
#| label: fig-survey_urban_rural
#| fig-cap: The Distribution of Presidential Preferences of Respondents Living in Urban vs
vote2020 |>
  ggplot(aes(x = urban, fill = vote24)) +
  stat_count(position = "dodge") +
  facet_wrap(facets = vars(sex)) +
  theme_minimal() +
  labs(
    y = "Number of respondents",
    fill = "Preferred Presidential Candidate"
  ) +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position = "bottom")
```

Table 2: ?(caption)



```
# do a table of the proportion of respondents for trump vs biden in each state
# overall is like 51% biden
```

Figure 3 illustrates the proportion of voters in each U.S. state who plan to support President Biden in the 2024 Presidential Election. We can see that there is strong support for President Biden in Washington, Oregon, New York, New Jersey, Maryland, Massachusetts, and Connecticut. We also see that Oklahoma, Idaho, and Utah heavily favor former President Trump. These results are expected and reflective of the 2020 election. These states have also been Dem and GOP strongholds since 1992 (or whatever year).

| Biden | Trump |
|-------|-------|
| 324 | 211 |

Electoral College based on survey data

Map of the US States that plan to support Biden or Trump

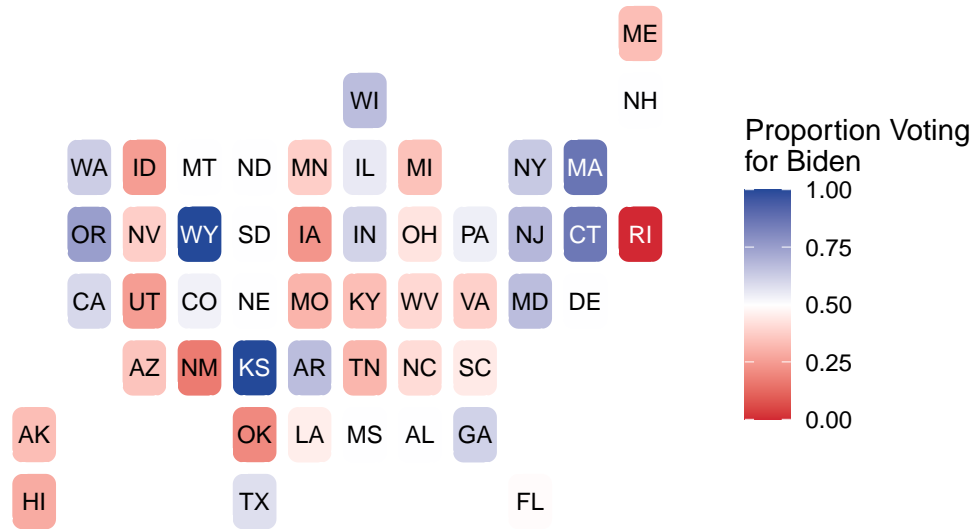


Figure 3: Electoral map based solely on the survey data

However, this electoral map also highlights some of the limitations for the survey data. Wyoming and Kansas are very blue in the electoral map based solely on the survey data, however that is very far from recent election results in those states. In fact, Wyoming was the most Pro-Trump state in both the 2016 and 2020 presidential elections, with Trump receiving 2016% and 2020% of the vote in each of those elections, respectively. The survey data set includes very few examples from small states.

There are no respondents from Vermont in the survey data, thus Vermont is missing from the electoral college map and `?@tbl-electoralcollegesurvey` is missing the 3 electoral votes allocated to Vermont.

2.2 Post Stratification Data

The post-stratification data used is comes from the American Community Survey (ACS) 2022 provided by IPUMS (Steven Ruggles and Schouweiler (2024)).

Talk more about it.

Table 5: Races of Population Data

| Race | Count | Proportion |
|-------|-------|------------|
| asian | 23855 | 0.0630671 |
| black | 31195 | 0.0824723 |

| Race | Count | Proportion |
|-----------------|--------|------------|
| mixed | 36430 | 0.0963125 |
| native american | 4440 | 0.0117383 |
| other | 19366 | 0.0511992 |
| white | 262962 | 0.6952105 |

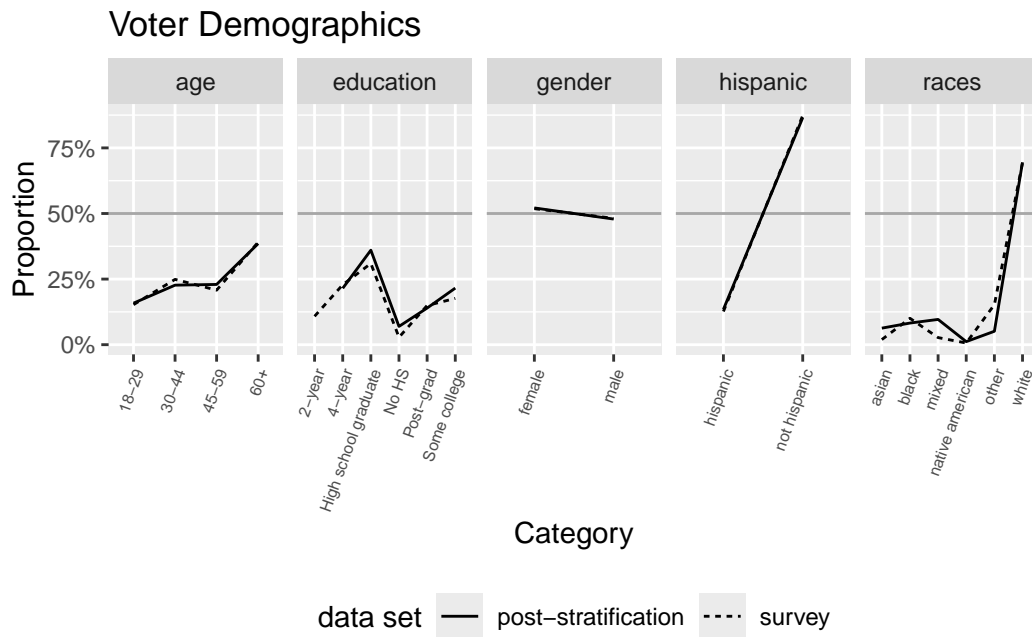


Figure 4: Voter Demographics in Survey and Post-Stratification Data Sets

Talk way more about it.

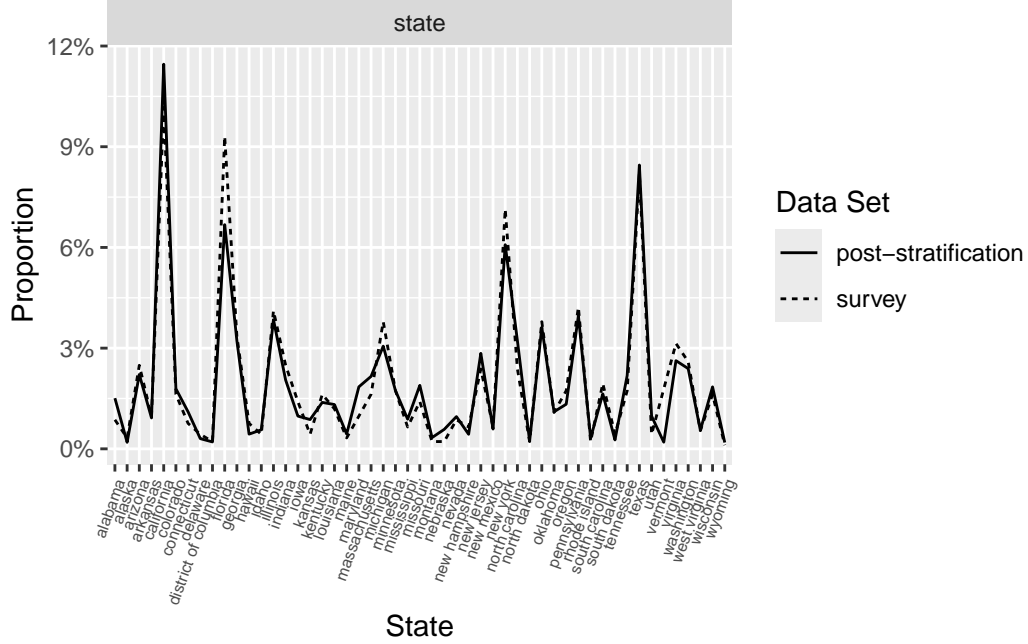


Figure 5: ?(caption)

3 Model

We will use multilevel regression with post-stratification (MRP) for our analysis. MRP is ...
Steps for MRP are:

3.1 Model set-up

To predict whether a person will voter for Joe Biden or Donald Trump, we built a logistic regression model using data from the America’s Political Pulse Survey (Iyengar, Lelkes, and Westwood 2024) and post-stratified it using ACS census data (Steven Ruggles and Schouweiler 2024). Logistic regression only works for binary classification, therefore we created a `vote_biden` variable during our data cleaning process which is equal to 1 if the person will vote for Joe Biden and 0 if the person will vote for Donald Trump.

The logistic regression model used to predict support for Biden using sex, age, race, highest level of education, state, and whether the individual lives in an urban or rural area is as follows:

$$y_i = \beta_0 + \beta_1 \times \text{sex}_i + \beta_2 \times \text{age_bracket}_i + \beta_3 \times \text{race}_i + \beta_4 \times \text{education_level}_i + \beta_5 \times \text{state}_i + \beta_6 \times \text{urban}_i$$

logistic activation

$$Pr(\text{vote_biden} = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \times \text{sex}_i + \beta_2 \times \text{age_bracket}_i + \beta_3 \times \text{race}_i + \beta_4 \times \text{education_level}_i + \beta_5 \times \text{state}_i + \beta_6$$

another try

$$Pr(\text{vote_biden} = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age_bracket}_i + \beta_3 \text{race}_i + \beta_4 \text{education_level}_i + \beta_5 \text{state}_i + \beta_6 \text{urban}_i)$$

We run our model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

We use the `predict()` function in R (R Core Team 2023) to apply our logistic regression model to the ACS Census data (Steven Ruggles and Schouweiler 2024).

Model weaknesses: logistic regression output is binary so it does not include the options of not voting or voting for a third-party candidate.

Concerns about model fit

3.1.1 Model justification

We expect to see increased support for President Biden amongst Black college-educated women in urban areas.

Table 6: Coefficient estimates

| | (1) |
|-------------------------------------|----------|
| (Intercept) | 0.400 |
| sexmale | −0.384 |
| age_bracket30-44 | −0.410 |
| age_bracket45-59 | −0.367 |
| age_bracket60+ | −0.417 |
| racesblack | 0.719 |
| racesmixed | 0.072 |
| racesnative american | 0.665 |
| racesother | −0.041 |
| raceswhite | −0.184 |
| education_level4-year | 0.154 |
| education_levelHigh school graduate | −0.110 |
| education_levelNo HS | 0.175 |
| education_levelPost-grad | 0.612 |
| education_levelSome college | −0.120 |
| urbanurban | 0.489 |
| Num.Obs. | 924 |
| R2 | 0.071 |
| Log.Lik. | −612.787 |
| ELPD | −629.8 |
| ELPD s.e. | 7.6 |
| LOOIC | 1259.5 |
| LOOIC s.e. | 15.3 |
| WAIC | 1259.3 |
| RMSE | 0.49 |

4 Results

| state | mean | lower | upper |
|----------------|-----------|-------------------------|-----------|
| connecticut | 0.5754088 | 0.5248635 | 0.6222304 |
| maine | 0.4860971 | 0.4489121 | 0.5233534 |
| massachusetts | 0.5852886 | 0.5330406 | 0.6340056 |
| new hampshire | 0.4912788 | 0.4543522 | 0.5272668 |
| rhode island | 0.5848102 | 0.5296204 | 0.6381118 |
| vermont | 0.4667625 | 0.4276716 | 0.5059922 |
| delaware | 0.5473492 | 0.5085107 | 0.5841437 |
| new jersey | 0.5970693 | 0.5408734 | 0.6492624 |
| new york | 0.5728537 | 0.5267325 | 0.6169554 |
| pennsylvania | 0.5329271 | 0.4905312 | 0.5737605 |
| illinois | 0.5469410 | 0.5058230 | 0.5866994 |
| indiana | 0.5210156 | 0.4799103 | 0.5611735 |
| michigan | 0.5277864 | 0.4877190 | 0.5672444 |
| ohio | 0.5393198 | 0.4963414 | 0.5809379 |
| wisconsin | 0.4889599 | 0.4511689 | 0.5271849 |
| iowa | 0.4726799 | 0.4362736 | 0.5105143 |
| kansas | 0.5135237 | 0.4757692 | 0.5506278 |
| minnesota | 0.4978665 | 0.4607639 | 0.5350637 |
| missouri | 0.5190974 | 0.4803314 | 0.5578227 |
| nebraska | 0.4877335 | 0.4515470 | 0.5244076 |
| north dakota | 0.4672411 | 0.4270508 | 0.5065665 |
| south dakota | 0.4586366 | 0.4124970 | 0.5028220 |
| virginia | 0.5694773 | 0.5281672 | 0.6084366 |
| alabama | 0.5517981 | 0.5118414 | 0.5903551 |
| arkansas | 0.5043985 | 0.4647812 | 0.5423465 |
| florida | 0.5810211 | 0.5258436 | 0.6331681 |
| georgia | 0.5715763 | 0.5307858 | 0.6086468 |
| louisiana | 0.5590368 | 0.5176668 | 0.5978165 |
| mississippi | 0.5338959 | 0.4921527 | 0.5746857 |
| north carolina | 0.5593944 | 0.5202705 | 0.5960924 |
| south carolina | 0.5674811 | 0.5247226 | 0.6066085 |
| texas | 0.5688479 | 0.5177381 | 0.6179507 |
| kentucky | 0.4957231 | 0.4566412 | 0.5337063 |
| maryland | 0.6087211 | 0.5611360 | 0.6528187 |
| oklahoma | 0.5140486 | 0.4608834 | 0.5632331 |
| tennessee | 0.5407975 | 0.5001774 | 0.5798433 |
| west virginia | 0.4746619 | 0.4338729 | 0.5147159 |
| arizona | 0.5714357 | 0.5163296 | 0.6230348 |
| colorado | 0.5607805 | 0.5138103 | 0.6055714 |
| idaho | 0.4978886 | 0.4596462 | 0.5361135 |
| montana | 0.4764823 | 0.4328175 | 0.5173764 |
| nevada | 0.5644374 | 0.5122927 ₁₂ | 0.6146024 |
| new mexico | 0.5605125 | 0.4878460 | 0.6256928 |
| utah | 0.5421761 | 0.4974419 | 0.5866311 |
| wyoming | 0.4696439 | 0.4275987 | 0.5105773 |
| california | 0.5925536 | 0.5268058 | 0.6537928 |
| oregon | 0.5403438 | 0.4953945 | 0.5835541 |
| washington | 0.5590623 | 0.5086821 | 0.6065638 |
| alaska | 0.5273360 | 0.4115645 | 0.6267830 |

Table 7: Estimated Popular Vote

| Lower Estimate | Mean Estimate | Upper Estimate |
|----------------|---------------|----------------|
| 0.3393468 | 0.5939406 | 0.8462525 |

electoral college

```
# googled the electoral college and made a csv
electoral_college <- read_csv("electoral_college_votes.csv", show_col_types = FALSE)
# downloaded and cleaned it

electoral_college$electoral_votes <- as.numeric(electoral_college$electoral_votes)

merged_ec <- merge(biden_support_by_state, electoral_college, by = "state")

biden_states_mean <- merged_ec |>
  filter(mean >= 0.5)

trump_states_mean <- merged_ec |>
  filter(mean < 0.5)

biden_ec_mean <- sum(biden_states_mean$electoral_votes)
trump_ec_mean <- sum(trump_states_mean$electoral_votes)

biden_states_mean <- merged_ec |>
  filter(lower >= 0.5)

trump_states_mean <- merged_ec |>
  filter(lower < 0.5)

biden_ec_lower<- sum(biden_states_mean$electoral_votes)
trump_ec_lower <- sum(trump_states_mean$electoral_votes)
```

The estimates for the support for Biden in each state are shown in `?@tbl-supporttbl`

| state | mean | lower | upper |
|----------------------|-----------|-----------|-----------|
| connecticut | 0.5754088 | 0.5248635 | 0.6222304 |
| maine | 0.4860971 | 0.4489121 | 0.5233534 |
| massachusetts | 0.5852886 | 0.5330406 | 0.6340056 |
| new hampshire | 0.4912788 | 0.4543522 | 0.5272668 |
| rhode island | 0.5848102 | 0.5296204 | 0.6381118 |
| vermont | 0.4667625 | 0.4276716 | 0.5059922 |
| delaware | 0.5473492 | 0.5085107 | 0.5841437 |
| new jersey | 0.5970693 | 0.5408734 | 0.6492624 |
| new york | 0.5728537 | 0.5267325 | 0.6169554 |
| pennsylvania | 0.5329271 | 0.4905312 | 0.5737605 |
| illinois | 0.5469410 | 0.5058230 | 0.5866994 |
| indiana | 0.5210156 | 0.4799103 | 0.5611735 |
| michigan | 0.5277864 | 0.4877190 | 0.5672444 |
| ohio | 0.5393198 | 0.4963414 | 0.5809379 |
| wisconsin | 0.4889599 | 0.4511689 | 0.5271849 |
| iowa | 0.4726799 | 0.4362736 | 0.5105143 |
| kansas | 0.5135237 | 0.4757692 | 0.5506278 |
| minnesota | 0.4978665 | 0.4607639 | 0.5350637 |
| missouri | 0.5190974 | 0.4803314 | 0.5578227 |
| nebraska | 0.4877335 | 0.4515470 | 0.5244076 |
| north dakota | 0.4672411 | 0.4270508 | 0.5065665 |
| south dakota | 0.4586366 | 0.4124970 | 0.5028220 |
| virginia | 0.5694773 | 0.5281672 | 0.6084366 |
| alabama | 0.5517981 | 0.5118414 | 0.5903551 |
| arkansas | 0.5043985 | 0.4647812 | 0.5423465 |
| florida | 0.5810211 | 0.5258436 | 0.6331681 |
| georgia | 0.5715763 | 0.5307858 | 0.6086468 |
| louisiana | 0.5590368 | 0.5176668 | 0.5978165 |
| mississippi | 0.5338959 | 0.4921527 | 0.5746857 |
| north carolina | 0.5593944 | 0.5202705 | 0.5960924 |
| south carolina | 0.5674811 | 0.5247226 | 0.6066085 |
| texas | 0.5688479 | 0.5177381 | 0.6179507 |
| kentucky | 0.4957231 | 0.4566412 | 0.5337063 |
| maryland | 0.6087211 | 0.5611360 | 0.6528187 |
| oklahoma | 0.5140486 | 0.4608834 | 0.5632331 |
| tennessee | 0.5407975 | 0.5001774 | 0.5798433 |
| west virginia | 0.4746619 | 0.4338729 | 0.5147159 |
| arizona | 0.5714357 | 0.5163296 | 0.6230348 |
| colorado | 0.5607805 | 0.5138103 | 0.6055714 |
| idaho | 0.4978886 | 0.4596462 | 0.5361135 |
| montana | 0.4764823 | 0.4328175 | 0.5173764 |
| nevada | 0.5644374 | 0.5122927 | 0.6146024 |
| new mexico | 0.5605125 | 0.4878460 | 0.6256928 |
| utah | 0.5421761 | 0.4974419 | 0.5866311 |
| wyoming | 0.4696439 | 0.4275987 | 0.5105773 |
| california | 0.5925536 | 0.5268058 | 0.6537928 |
| oregon | 0.5403438 | 0.4953945 | 0.5835541 |
| washington | 0.5590623 | 0.5086821 | 0.6065638 |
| alaska | 0.5273360 | 0.4115645 | 0.6267830 |
| hawaii | 0.5555602 | 0.4436609 | 0.6643134 |
| district of columbia | 0.6745628 | 0.6191542 | 0.7272233 |

visualization here

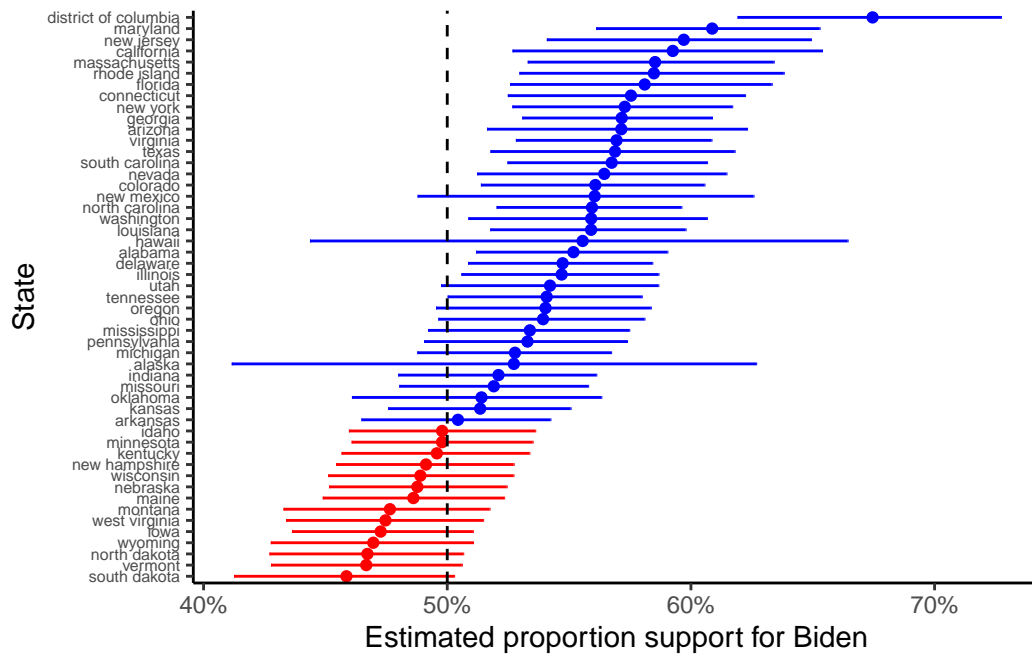


Figure 6: Estimated proportion of each state voting for Biden in 2024 (Post-Stratification)

compared to survey data shown in Figure 7

electoral map

type here

5 Discussion

5.1 Popular Vote Projection

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

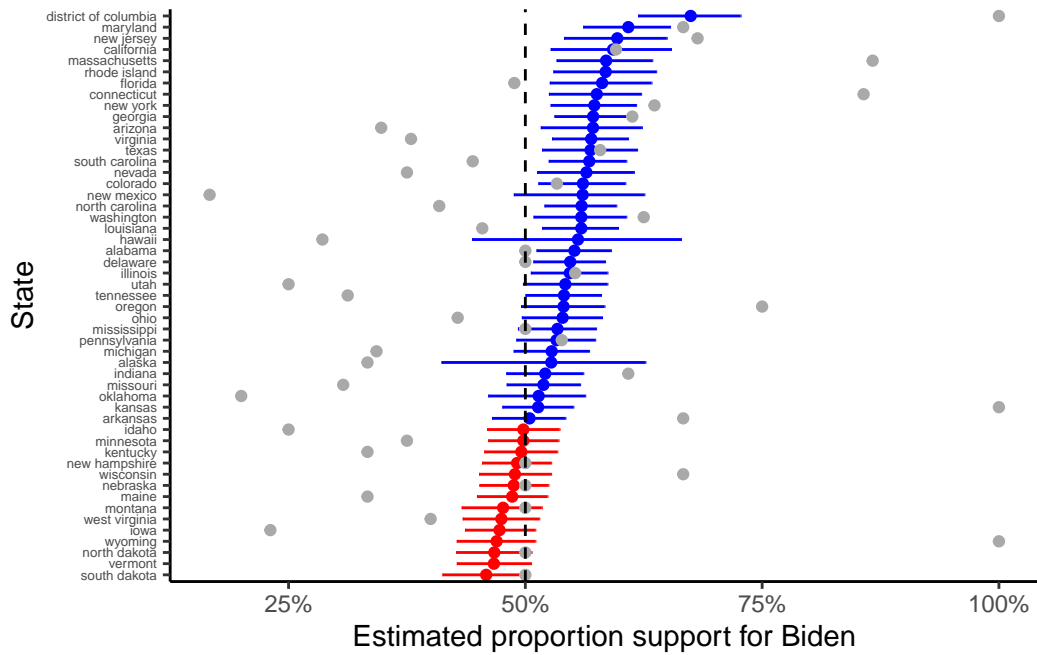


Figure 7: Estimated proportion of each state voting for Biden in 2024 Post-Stratification vs Survey Data

Map of the US States that plan to support Biden or Trump (Post-Str

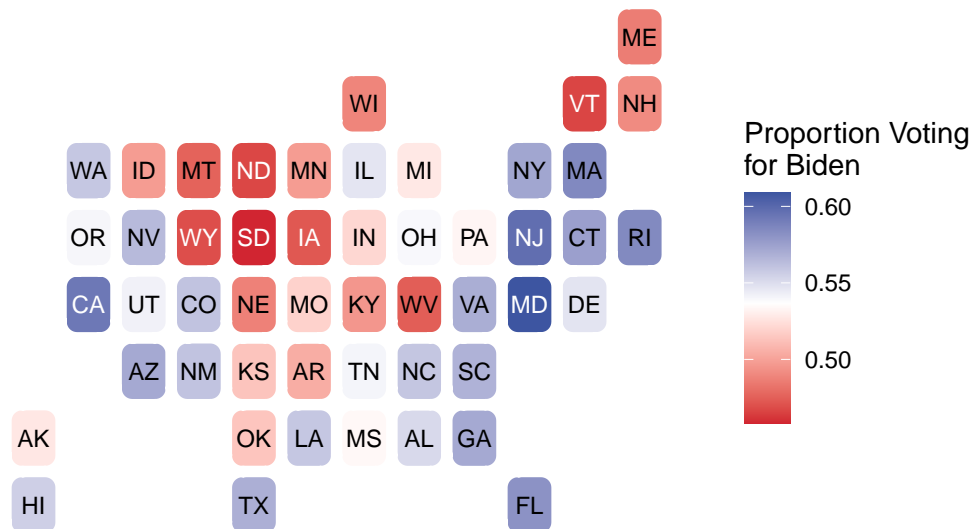


Figure 8: Electoral map based on post-stratification data

5.2 Electoral College Projection

5.3 Polarization and America's Urban-Rural Divide

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

A.1 Data Cleaning

B Model details

type something here

B.1 Correlation Map

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

Figure 9: `?(caption)`

Model plot (still need to fix this)

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

Figure 10: ?(caption)

References

- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://github.com/sfirke/janitor>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Iyengar, Shanto, Yphtach Lelkes, and Sean Westwood. 2024. *America’s Political Pulse*. <https://polarizationresearchlab.org/americas-political-pulse/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Steven Ruggles, Matthew Sobek, Sarah Flood, and Megan Schouweiler. 2024. *IPUMS USA: Version 15.0 [ACS 2022]*. Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D010.V15.0>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.