

My title*

My subtitle if needed

Talia Fabregas Fatimah Yunusa Aamishi Sundeep Avarsekar

March 12, 2024

The 2024 U.S. Presidential Election ...

1 Introduction

The 2024 U.S. Presidential Election will take place on Tuesday November 5 2024. Incumbent President Joseph R. Biden Jr. will seek a second term. Former President Donald J. Trump GOP nominee.

Delegates to secure the nomination, Nikki Haley has dropped out of GOP You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and

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.

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

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

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

2 Data

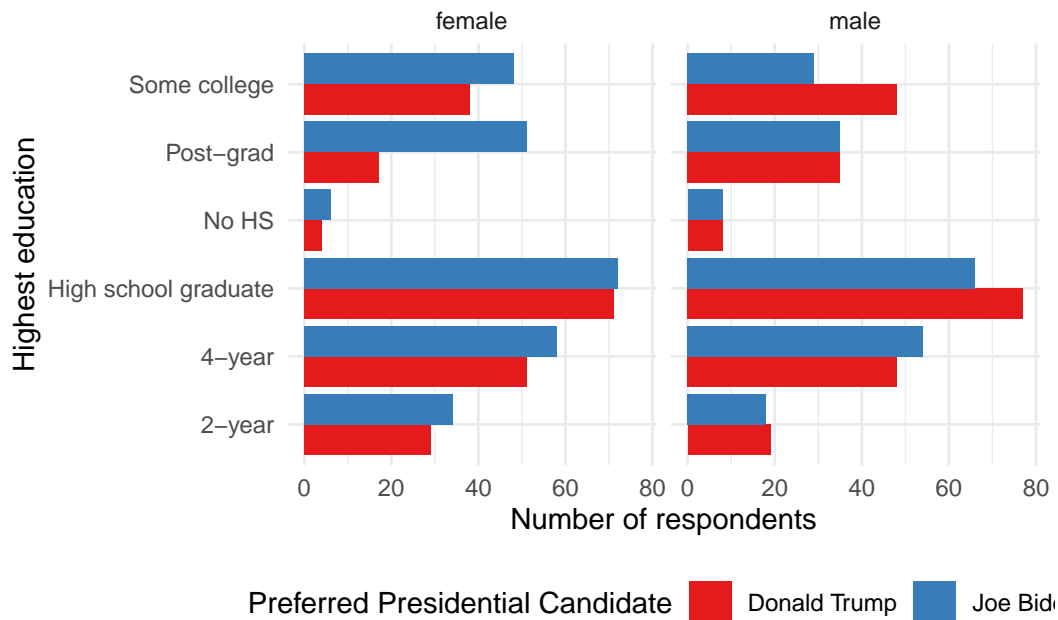


Figure 1: The Distribution of Presidential Preferences, by gender and highest level of education

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, ... 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.

2.1.1 Survey Data figures and tables go here

In order to gain a better understanding of our survey data, we conducted exploratory data analysis.

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.

```
# 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
survey_analysis_data |>
  ggplot(aes(x = urban, fill = vote24)) +
  stat_count(position = "dodge") +
  # facet_wrap(facets = vars(sex)) +
```

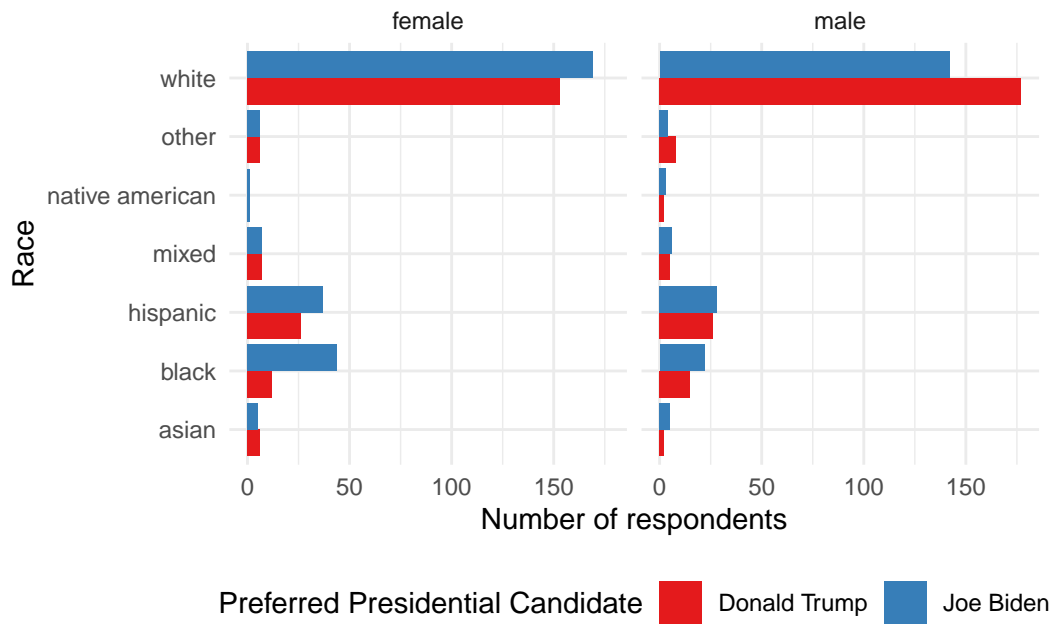
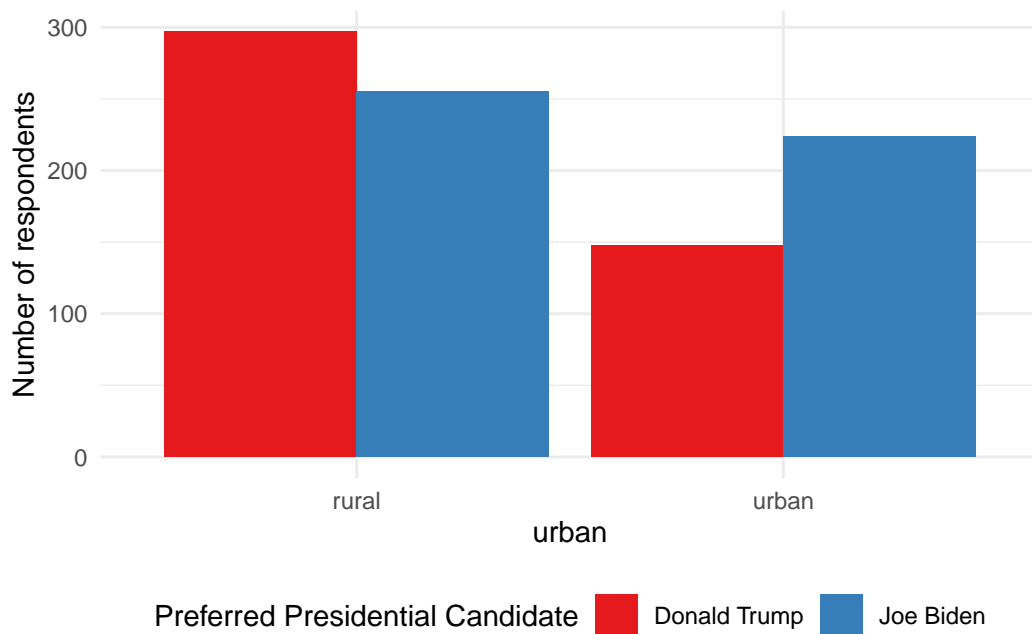


Figure 2: The Distribution of Presidential Preferences, by gender and race

```
theme_minimal() +
labs(
  y = "Number of respondents",
  fill = "Preferred Presidential Candidate"
) +
scale_fill_brewer(palette = "Set1") +
theme(legend.position = "bottom")
```



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

Map of the US States that plan to support Joe Biden or Donald Trun

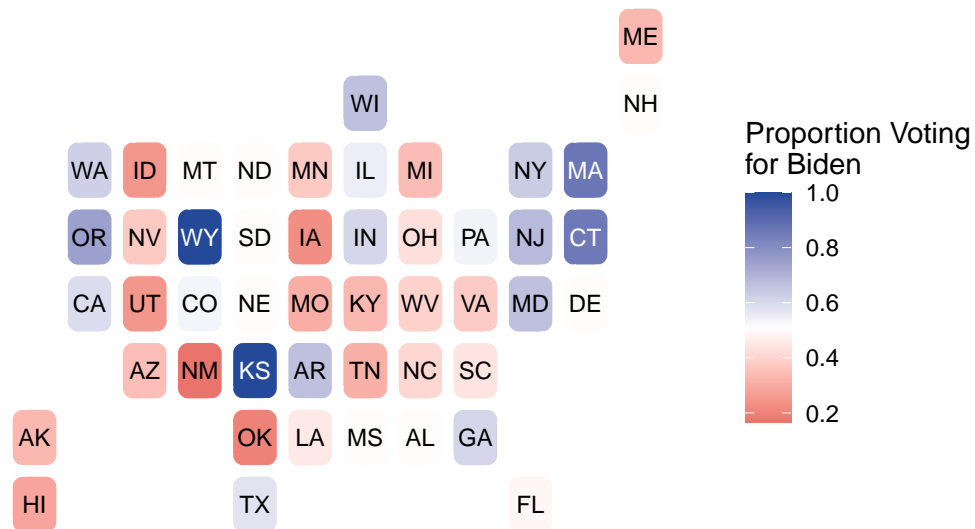


Figure 3: Electoral map based solely on the survey data

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 there were similar margins in the 2020 election. These states have also been Dem and GOP strongholds since 1992 (or whatever year). However, this electoral map also highlights some of the limitations of 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.

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.

Talk way more about it.

3 Model

logistic regression

binary

predicts support for trump or biden

3.1 Model set-up

Define y_i whether an individual's preferred 2024 presidential candidate is Joe Biden (1) or Donald Trump (0). 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:

$$Pr(\text{vote_biden}_i = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \times x_{\text{sex}_i} + \beta_2 \times x_{\text{age_bracket}_i} + \beta_3 \times x_{\text{race}_i} + \quad (1)$$

$$\beta_3 \times x_{\text{education_level}_i} + \beta_4 \times x_{\text{state}_i} + \beta_5 \times x_{\text{urbani}_i} + \beta_5 \times x_{\text{race_black}_i}) \quad (2)$$

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.

3.1.1 Model justification

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

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 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.

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.

Table 1: Explanatory models of presidential vote based on ...

	Presidential Vote Model
(Intercept)	1.75 (3.75)
sexmale	−0.61 (0.16)
age_bracket30-44	−0.52 (0.29)
age_bracket45-59	−0.69 (0.30)
age_bracket60+	−0.68 (0.26)
raceshispanic	0.00 (3.61)
racemixed	−0.30 (3.58)
racenative american	0.17 (3.76)
racesoother	−0.37 (3.66)
racesswhite	−0.36 (3.57)
race_black	1.14 (3.62)
education_level4-year	0.54 (0.29)
education_levelHigh school graduate	−0.16 (0.27)
education_levelNo HS	0.61 (0.52)
education_levelPost-grad	1.21 (0.33)
education_levelSome college	−0.01 (0.30)
stateAlaska	−0.85 (1.71)
stateArizona	−1.48 (0.98)
stateArkansas	−0.25 (1.15)
stateCalifornia	−0.26 (0.89)
stateColorado	−0.79 (1.04)
stateConnecticut	1.29 (1.55)
stateDelaware	−1.50 (1.45)
stateDistrict of Columbia	34.35 (31.62)
stateFlorida	−0.91 (0.91)

Appendix

A Additional data details

B Model details

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 4: `?(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 5: `?(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 Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.