

# Forecasting the 2024 U.S. Presidential Election\*

President Joe Biden Projected to win the Popular Vote Based on MRP Analysis

Talia Fabregas

April 17, 2024

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

## 1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and (`rohan?`).

The 2024 U.S. Presidential election will take place on Tuesday, November 5th. Amidst unprecedented levels of political polarization, American voters are set to see a remat

## 2 Data

### 2.1 Survey Data

Figure 13 illustrates the proportion of subsetting survey respondents in each state who plan to support President Biden in the 2024 election. Overall, the survey data set appears to have stronger overall support for President Joe Biden than the general U.S. electorate.

---

\*Code and data are available at: <https://github.com/taliafabs/us-election-analysis.git>

Table 1: Popular vote and electoral college based on subset survey data

Survey Estimate:	Biden	Trump
Num Votes	6033.00	3967.00
% Votes	60.33	39.67
Electoral College	460.00	78.00

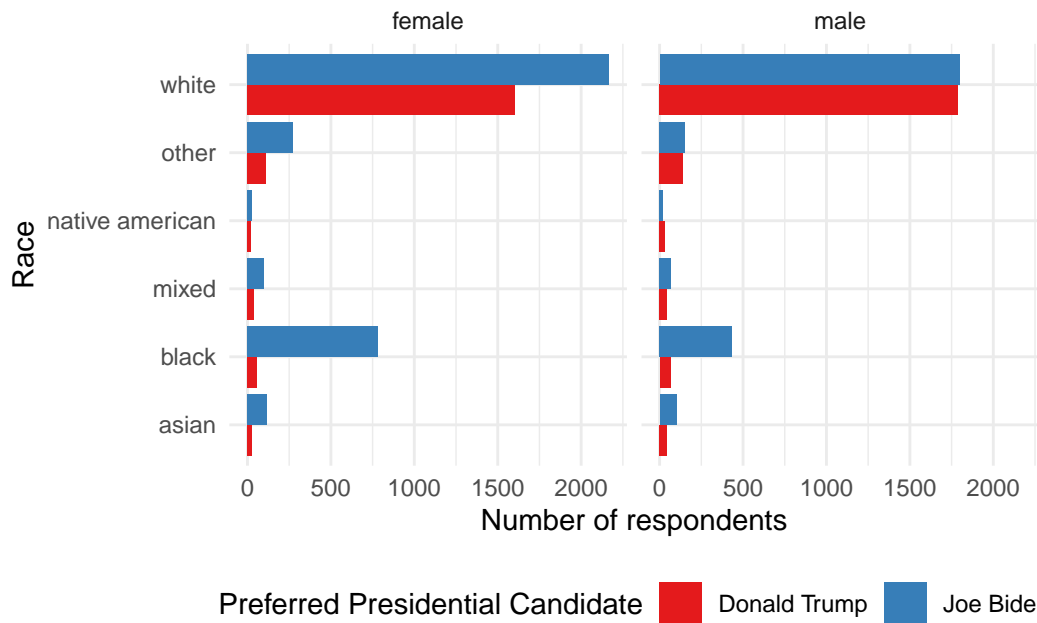


Figure 1: Preferred presidential candidates of survey subset respondents, by gender and race

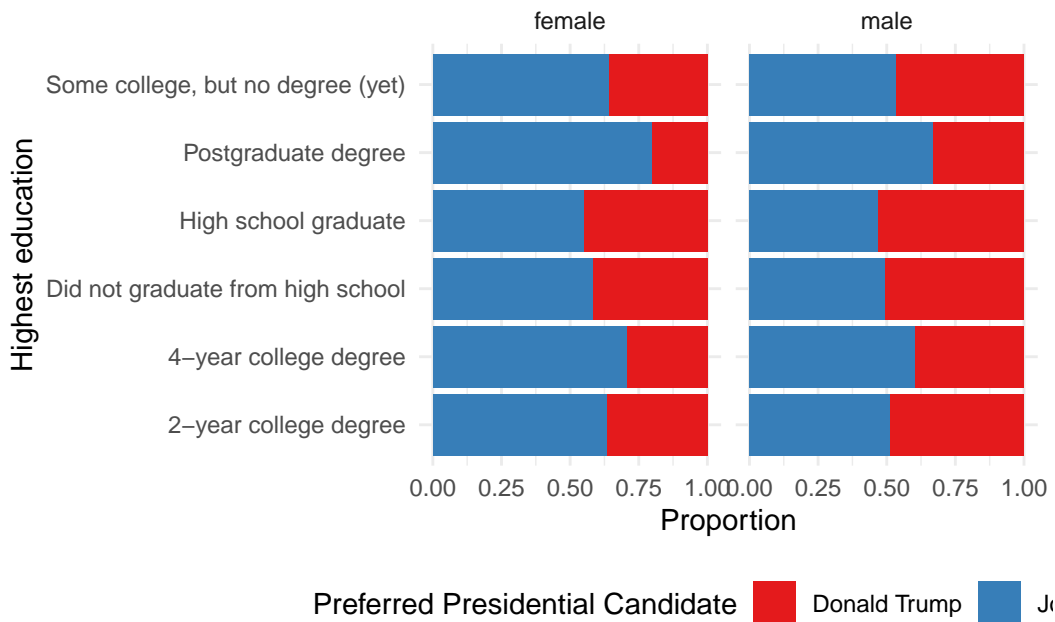


Figure 2: Preferred presidential candidates of survey subset respondents, by highest level of education

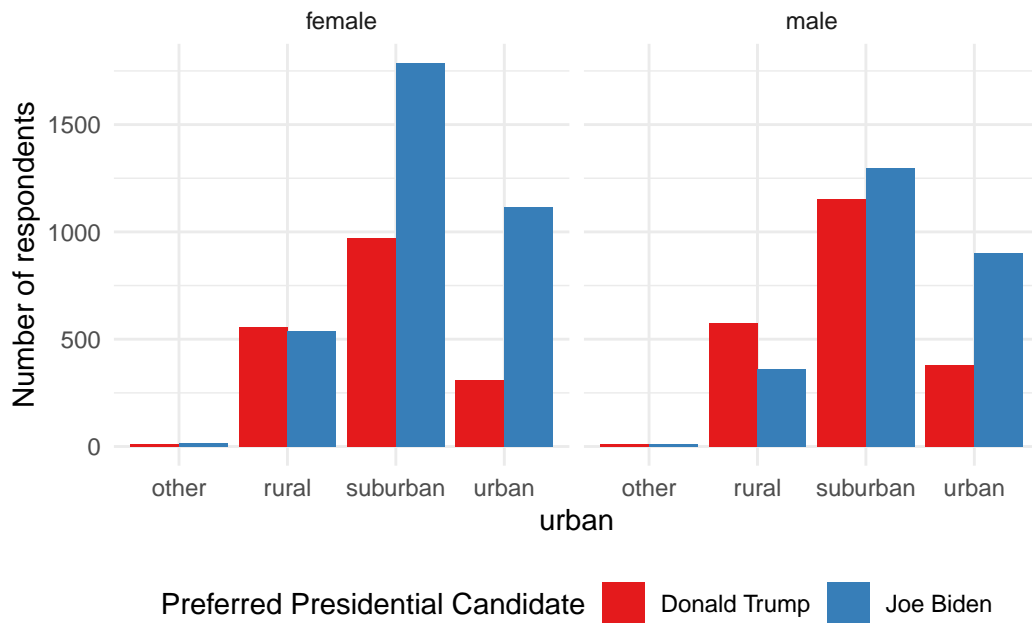


Figure 3: Preferred presidential candidate of subset respondents living in urban vs rural areas

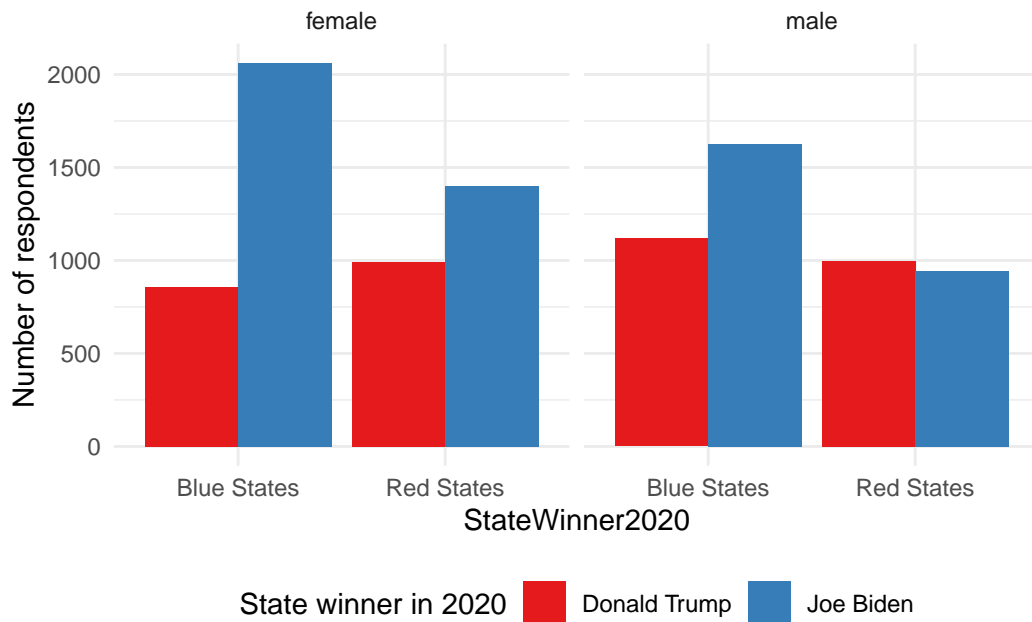


Figure 4: Preferred presidential candidate of subset respondents in states won by Trump vs Biden in 2020

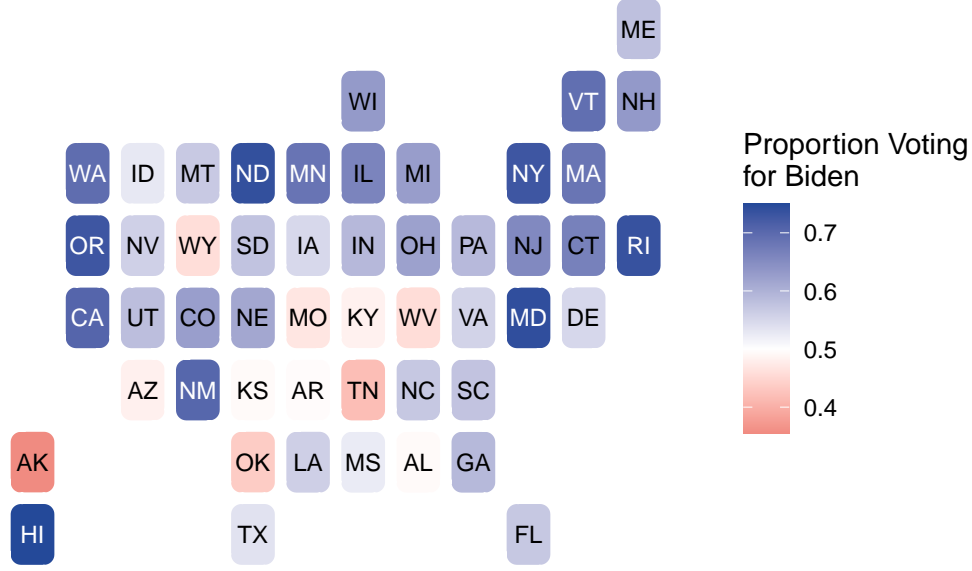


Figure 5: Electoral college map based on the subsetting survey data

### 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 C.

#### 3.1 Model set-up

$$vote\_biden_i | \pi_i \sim \text{Bern}(\pi_i) \quad (1)$$

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age\_bracket}_i + \beta_3 \text{race}_i + \beta_4 \text{race\_hispanic}_i + \beta_5 \text{education\_level}_i + \beta_6 \text{urban}_i \quad (2)$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\beta_3 \sim \text{Normal}(0, 2.5) \quad (6)$$

$$\beta_4 \sim \text{Normal}(0, 2.5) \quad (7)$$

$$\beta_5 \sim \text{Normal}(0, 2.5) \quad (8)$$

$$\beta_6 \sim \text{Normal}(0, 2.5) \quad (9)$$

\end{align}

Table 2: 2024 U.S. election result estimates for Joe Biden based on post-stratification

(a) Popular Vote		
Estimate:	Biden %	Trump %
Lower Estimate	48.55	51.45
Mean Estimate	55.47	44.53
Upper Estimate	62.56	37.44
(b) Electoral College		
Electoral College Estimate:	Biden	Trump
Lower Estimate	220	318
Mean Estimate	413	125
Upper Estimate	517	21

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

## 4 Results

### 4.1 Popular Vote Prediction

### 4.2 Electoral College Prediction

## 5 Discussion

### 5.1 First discussion point

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. The survey dataset appears to favor President Joe Biden more than the general U.S. electorate.

### 5.2 Second discussion point

Swing states and error ranges in the MRP analysis electoral college prediction

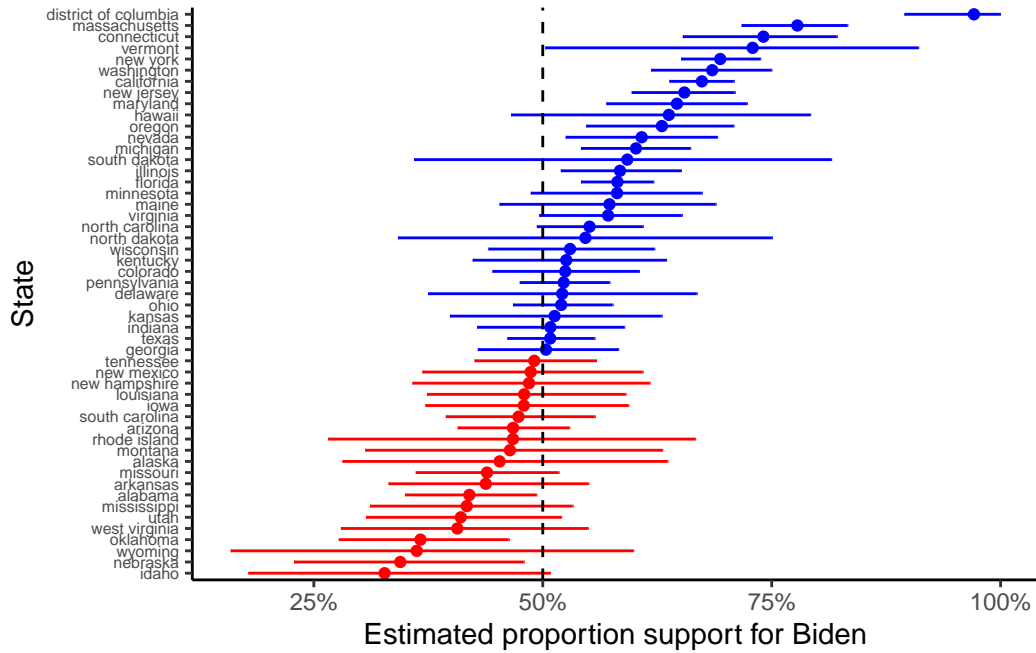


Figure 6: Estimated proportion of each state voting for Biden in 2024 based on MRP analysis

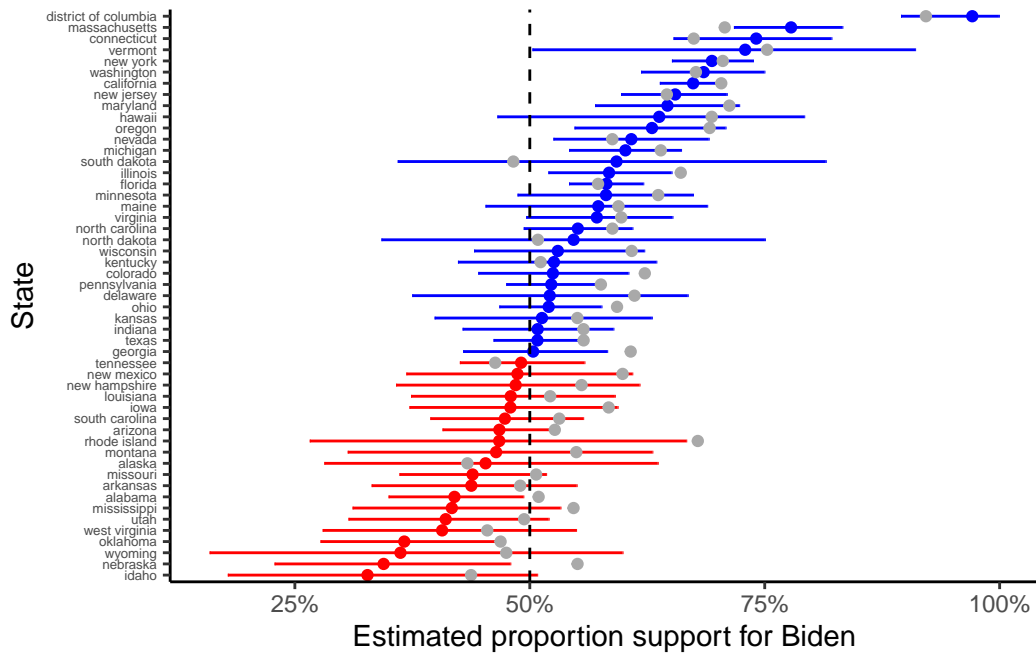


Figure 7: Estimated proportion of each state voting for Biden in 2024 Post-Stratification vs Subsampled Survey Data

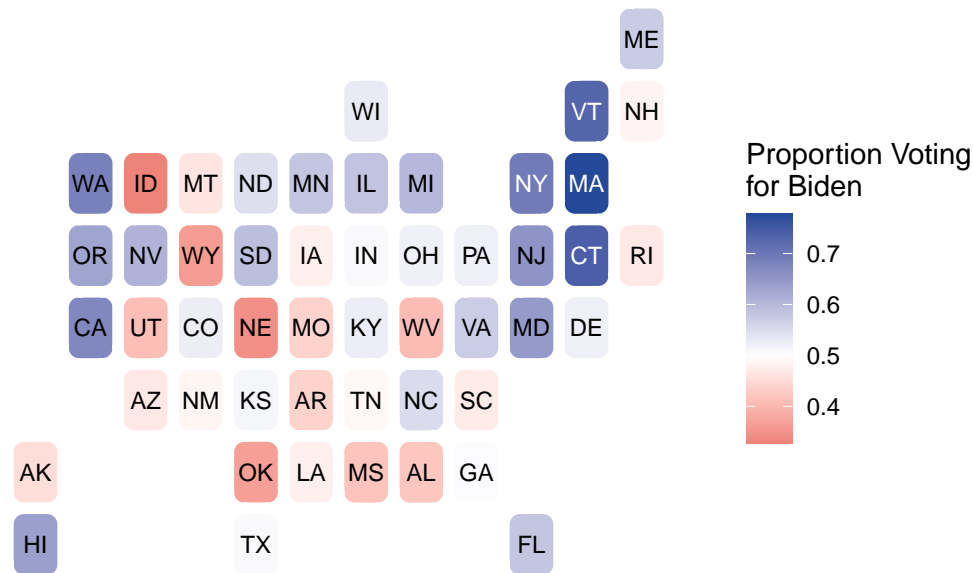


Figure 8: Electoral map based on MRP analysis

### 5.3 Third discussion point

### 5.4 Weaknesses and Limitations

### 5.5 Next Steps

Python, more recent data set, softmax regression, gradient descent Split the survey data into training, validation, and test Use gradient descent to find the optimal weights to maximize validation accuracy Apply the model to the post-stratification data Softmax regression does risk overfitting

## Appendix

### A Additional data cleaning details

### B Additional survey data details

The original survey data set contained 60,000 responses, but it was subsetting to 10,000 so that R and `rstanarm` could handle it (R Core Team 2023). The `glm` function of the `rstanarm` package was used to fit the logistic regression model to predict 2024 presidential vote choice based on state, whether Biden or Trump won that state in 2020, sex, age bracket, race, and highest level of education completed. However, when I tried to fit the model using my original survey data set, it took hours to run and I was not able to post-stratify due to the following error: `Error: vector memory exhausted (limit reached?)`. 10,000 of the 60,000 responses were randomly selected using the `sample` function of base R. The figures below show the results of the exploratory data analysis conducted on the original survey data set.

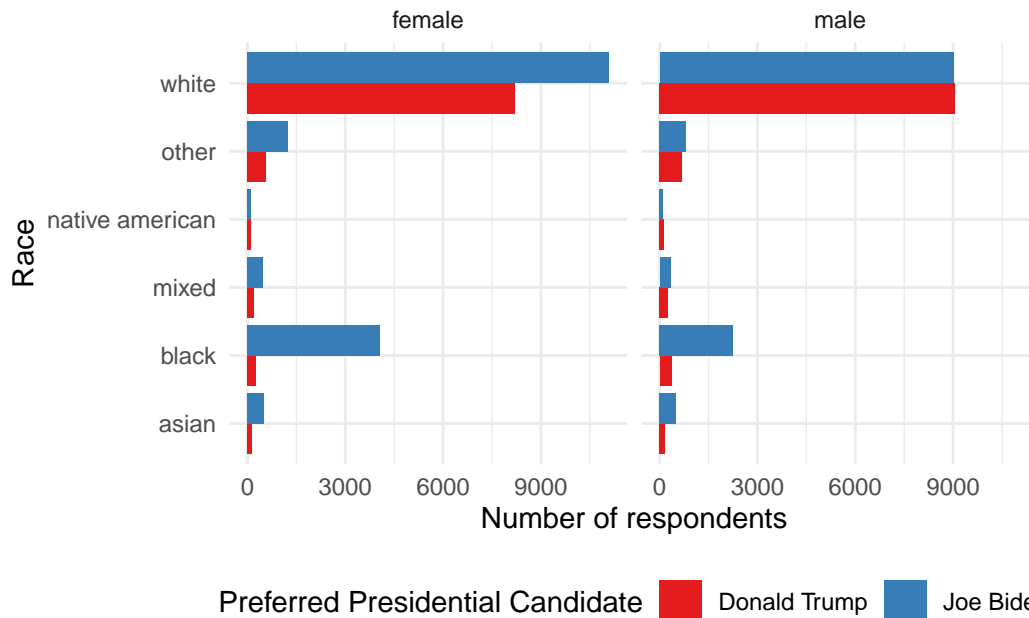


Figure 9: Preferred presidential candidates of survey respondents, by gender and race



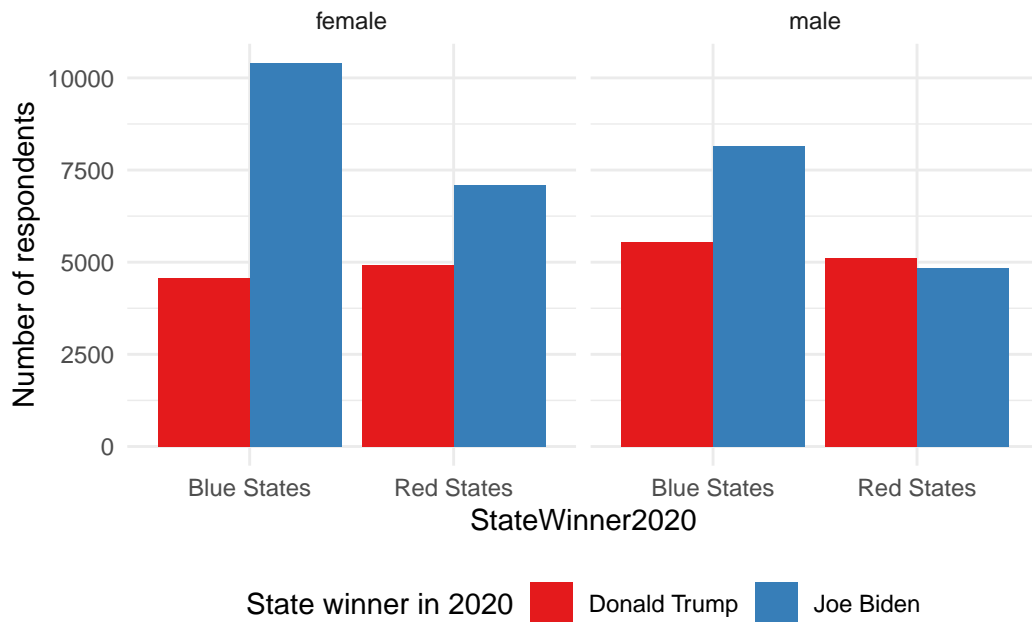


Figure 10: Preferred presidential candidate of survey respondents in states carried by Trump vs Biden in 2020

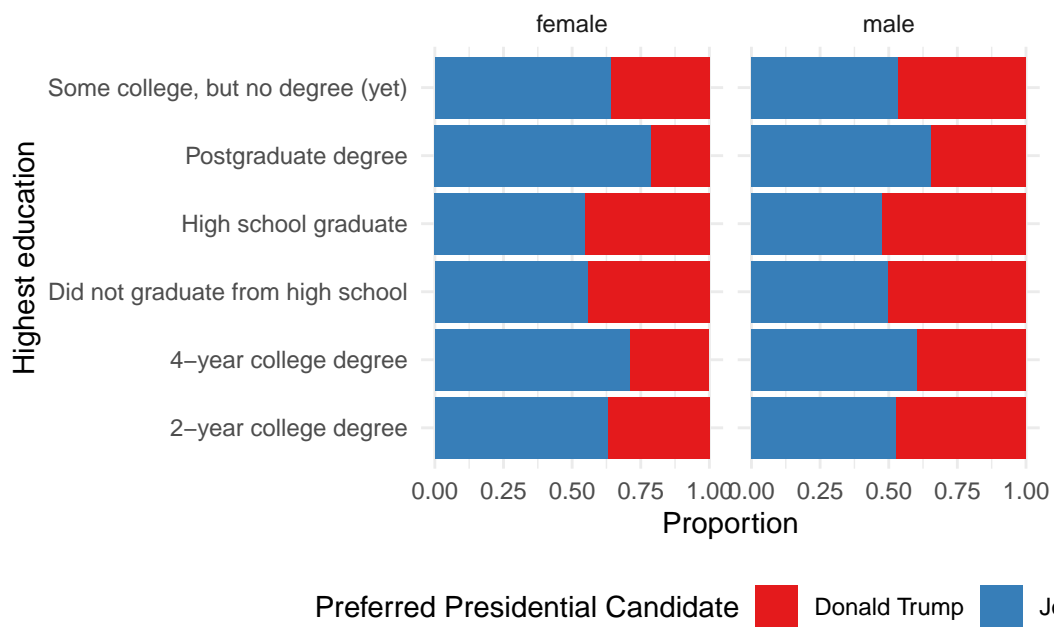


Figure 11: Preferred presidential candidates of survey respondents, by highest level of education

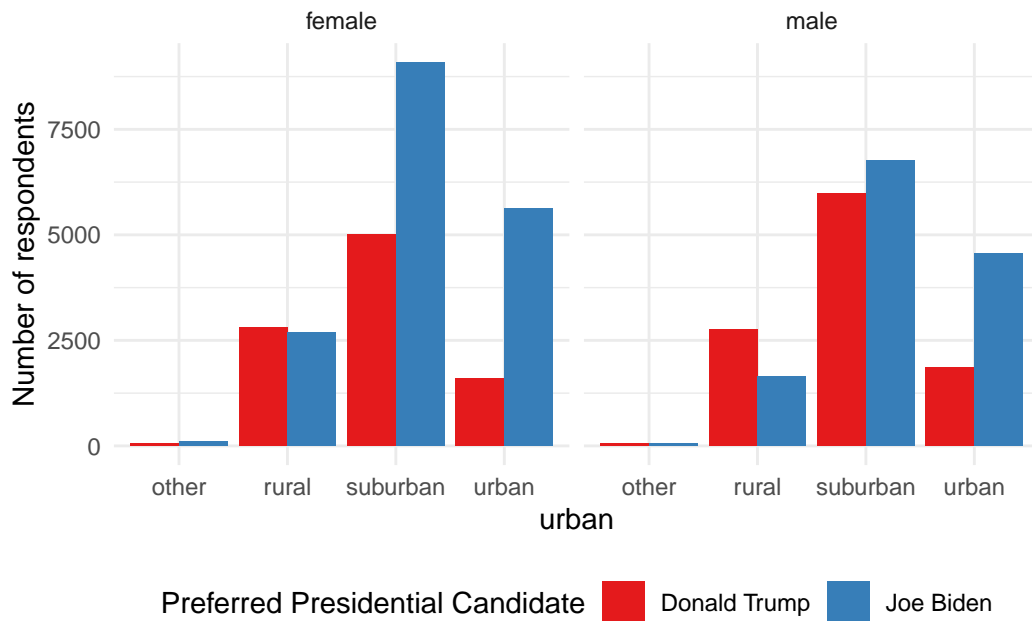


Figure 12: Preferred presidential candidate of survey respondents living in urban vs rural areas

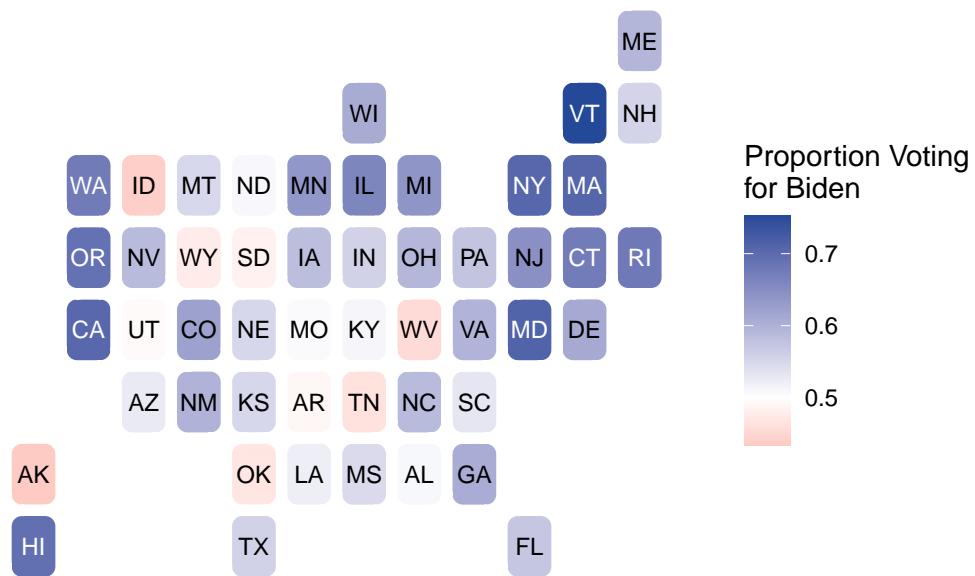
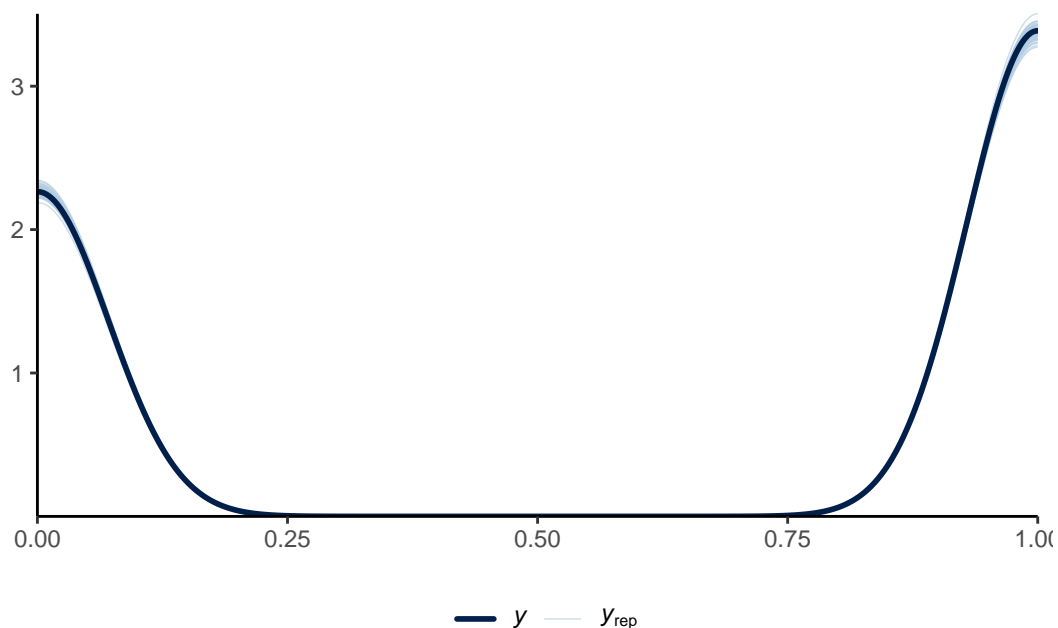


Figure 13: Electoral college map based on the survey dataset



(a) Posterior prediction check

sexmale	statealaska	stateindiana	s
sexfemale	statearizona	stateiowa	s
bracket30-44	statearkansas	statekentucky	s
bracket45-59	statecalifornia	statelouisiana	s
bracket60+	statecolorado	statemaine	s
4-year college degree	stateconnecticut	statemaryland	s
Did not graduate from high school	statedelaware	statemassachusetts	s
High school graduate	statedistrict of columbia	statemichigan	s
Postgraduate degree	stateflorida	stateminnesota	s
Some college, but no degree (yet)	stategeorgia	statemississippi	s
white			
black			
hispanic			
other			

(b) Comparing the posterior with the prior

Figure 14: Examining how the model fits, and is affected by, the data

## **C Model details**

**C.1 Posterior predictive check**

**C.2 Markov Chain Monte Carlo**

**C.3 Credibility intervals**

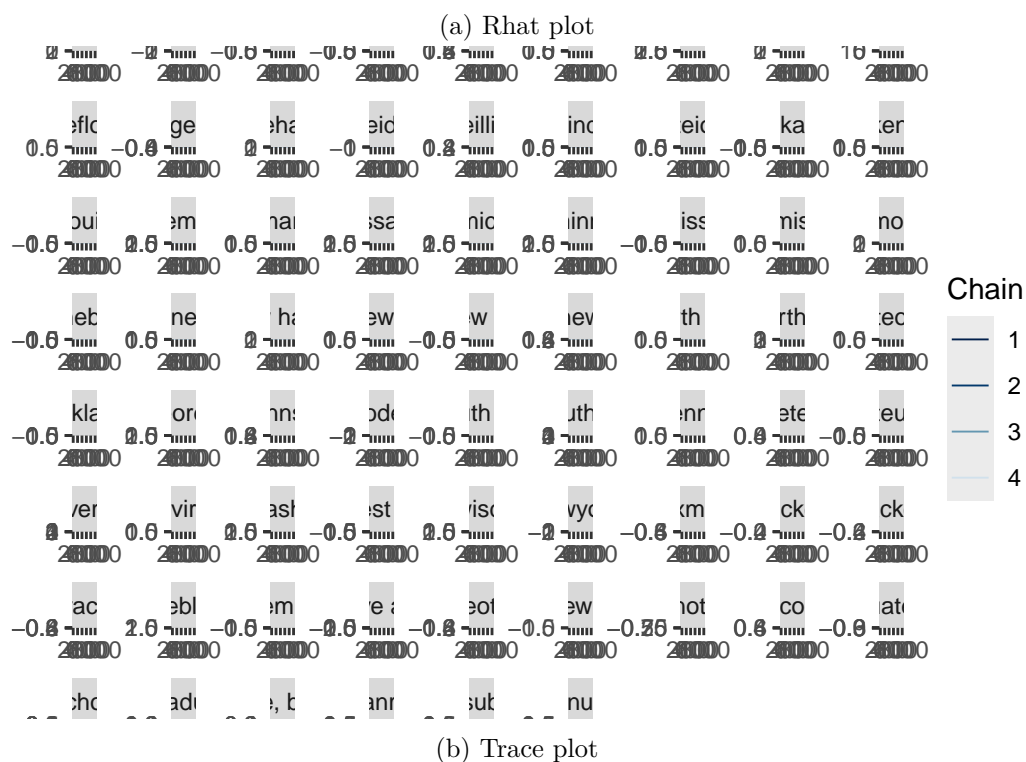
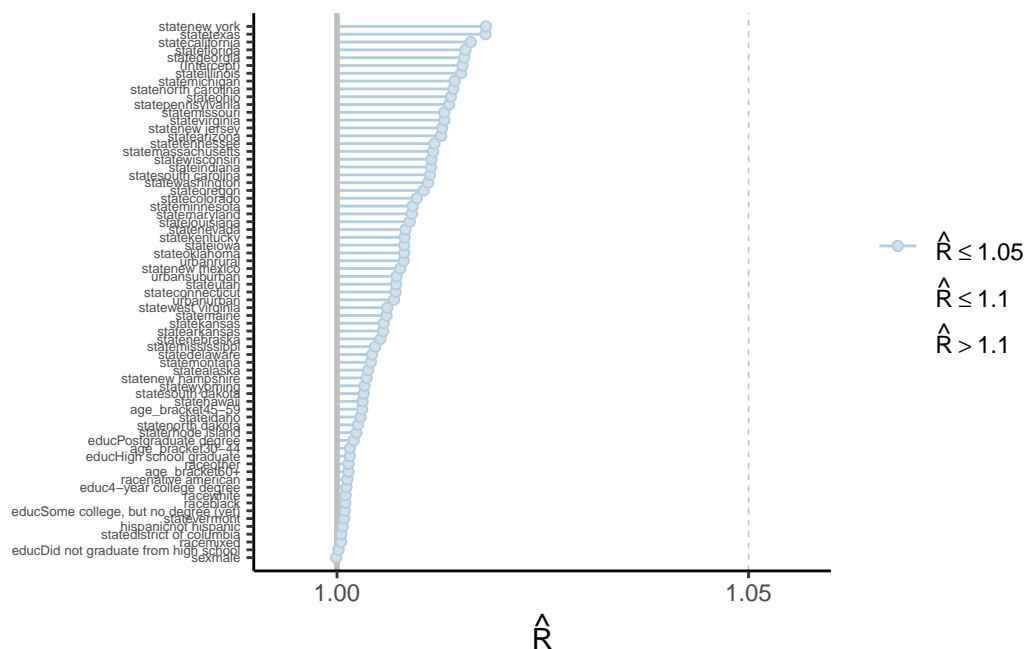


Figure 15: Checking the convergence of the Markov Chain Monte Carlo (MCMC) algorithm

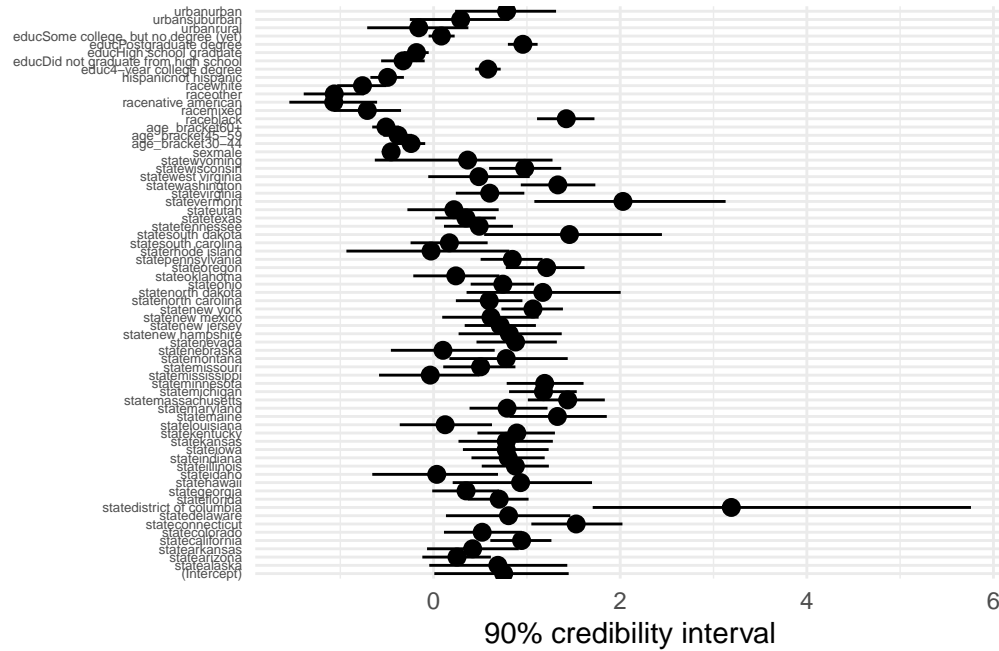


Figure 16: 90% Credibility intervals for the predictors of `vote_biden`

## References

- 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/>.