Education, Hig*

Logistic Regression Analysis of Voter Behavior

Talia Fabregas

December 2, 2024

First sentence. Second sentence. Third sentence. Fourth sentence. Key Trump 2024 voting blocs, including young men without a college education were less likely to vote in the 2022 midterms when he was not on the ballot compared to 2020

1 Introduction

Overview paragraph

This study aims to estimate how race, gender, education, trust in the federal government, civic engagement, and interest in politics impact voting behavior. The first thing that this study will examine is the effects of age, gender, race, education, trust in the federal government, civic knowledge, and civic engagement on the probabiliry that an individual supports Trump. The second thing that this study will examine is the effects of past presidential vote choice, age, highest level of education completed, trust in the federal government, knowledge of which party holds a majority in Congress, and interest in politics on the probability that an individual who voted for Donald Trump or Joe Biden in the 2020 U.S. presidential election would also vote in the 2022 U.S. midterm election. The estimands, which can never be known with complete certainty, include:

- The true effects of age, gender, race, education, trust in the federal government, knowledge of which party has a majority in the U.S. House, and interest in politics on the likelihood that an individual supports Trump.
- The true effects of presidential vote preference, age, education, trust, civic knowledge, and interest in politics on the likelihood that an individual who voted in a presidential election will also vote in the subsequent midterm election (Alexander 2023) .

*Code and data are available at: https://github.com/taliafabs/US-Midterms-2022.git.

Results paragraph

Based on the results of applying logistic regression models to predict voter turnout and vote preference, I estimate that

Why it matters paragraph In the 2024 election, Democratic Senate candidates (both incumbent and non-incumbent) were elected in states that President-Elect Trump carried: Michigan, Wisconsin, Nevada, and Arizona. In Nevada and Wisconsin, Senators Jackie Rosen and Tammy Baldwin

The remainder of this paper is structured as follows. Section 2 contains an overview of the survey data set from Schaffner, Ansolabehere, and Shih (2023) that was used, visualizations of different variables, and summary statistics. Section 3 contains the logistic regression models used to predict the probability that an individual who voted for either Trump or Biden in the 2020 presidential election would also vote in the 2022 midterm election and the probability that an individual would support Trump over Biden in 2020. **?@sec-results** contains tables and data visualizations that present results about what characteristics made individuals who voted in the 2020 presidential election more likely to also vote in the 2022 midterm election and what characteristics make individuals more likely to support Trump. **?@sec-discussion** contains detailed a detailed discussion about the results presented in **?@sec-results**, including demographics where support for Trump is strong, why some 2020 presidential election voters were more likely than others to vote in the 2022 midterm election, and how the 2016, 2020, and 2024 polling misses can be explained by low civic engagement and distrust in government among Trump voters. [Appendix - @sec-surveys] contains a deep-dive into CES survey methodology.

2 Data

2.1 Overview

The survey data set from Schaffner, Ansolabehere, and Shih (2023) contains …

I cleaned the data set to include only individuals who

We use the statistical programming language R (R Core Team 2023)…. Our data (Schaffner, Ansolabehere, and Shih 2023)…. Following Alexander (2023), we consider…

2.2 Measurement

The primary measurement task is to capture how Americans view their representatives, how they hold the different levels of government to account during elections, how they voted, their electoral experiences, and how voting behavior and experiences vary across different regions, demographics, and social contexts (Schaffner, Ansolabehere, and Shih 2023). Surveys are a

widely-used instrument for measuring public opinion during election cycles (Alexander 2023). During U.S. presidential and midterm election cycles, pollsters and researchers conduct surveys that measure candidate preferences, public opinion, and how Americans' diverse geography, demographics, and experiences impact them. Researchers often use these measurements to predict election outcomes, analyze election outcomes, and analyze which factors predict voting behavior and vote preference.

The Cooperative Election Study (CES) survey data set that I used is a collection of 60,000 responses from a nationally-representative sample of American adults. Each entry represents the political preferences, voting intentions, ideological leanings, demographics, issue evaluations, and past voting decisions of one respondent. The CES survey has been conducted every year since 2006 (Schaffner, Ansolabehere, and Shih 2023). In presidential and midterm election years, it consists of a pre-election wave and a post-election wave. The pre-election wave aims to measure the opinions, vote preferences, vote intentions of the American public, and demographics (Schaffner, Ansolabehere, and Shih 2023). The post-election wave aims to measure how different factors, including geography, demographics, issue-evaluations, and the state of the economy influenced Americans' decisions about who to vote for or whether to vote at all in the recent election (Schaffner, Ansolabehere, and Shih 2023).

The transformation of an individual American adult's opinion to an entry in the CES 2022 data set follows three steps, as outlined by Schaffner, Ansolabehere, and Shih (2023):

1. **Survey:** selected voters from a nationally representative sample respond to a Common Content survey.
2. **Adjustment:** survey responses are weighted to adjust for any imbalances that exist in the sample.
3. **Reporting:** the weighted survey results are recorded as entries in the data set. The data set then serves as a snapshot of American public opinion in the weeks leading up to and shortly after the midterm election.

2.3 Outcome variables

Firstly, I will use age, gender, race, highest level of education completed, trust in the government, and knowledge of which party controls the U.S. House of Representatives to predict whether an individual supports Donald Trump.

Then, I will use 2020 presidential vote, age, highest level of education completed, trust in the government, knowledge of which party holds a majority in the U.S. House of Representatives, and interest in politics to predict whether someone who did vote for either Donald Trump or Joe Biden in the 2020 presidential election would also vote in the 2022 midterm election, or more generally, in an election where neither Trump nor Biden was on the ballot.

It is possible that variables that predict support for Trump can also be used to predict turnout in the 2022 midterm election. Support for Trump is high among voters who are male, do not

have a college education, do not trust the government, and have low civic engagement. These factors make them less likely to respond to a survey, and it is possible that these factors also make them less likely to vote in a midterm election where Trump is not on the ballot.

Table 1: Biden voters are overrepresented in the CES 2022 survey data set.

2020 Vote	Num respondents	%
Donald Trump	17442	41.56
Joe Biden	24526	58.44

The 2020 presidential election had record-setting turnout (Grant 2023). Hartig et al. (2023) found that midterm elections typically have lower turnout than the preceding presidential election and the responses to the 2022 CES survey are consistent with that. Just under 30% of 2022 CES respondents who did vote in the 2020 presidential election did not vote in the 2022 midterm election.

Table 2: 70.64% of respondents who voted in the 2020 presidential election also voted in the 2022 midterm election

Voting Status	Num respondents	%
Did Not Vote in 2022	12323	29.36
Voted in 2022	29645	70.64

Table 3: Biden 2020 Voters Voted in 2022 midterms slightly more than Trump 2020 voters. 73.26% of respondents who voted for Trump in 2020 voted in 2022, compared to 74.67% of respondents who voted for Biden.

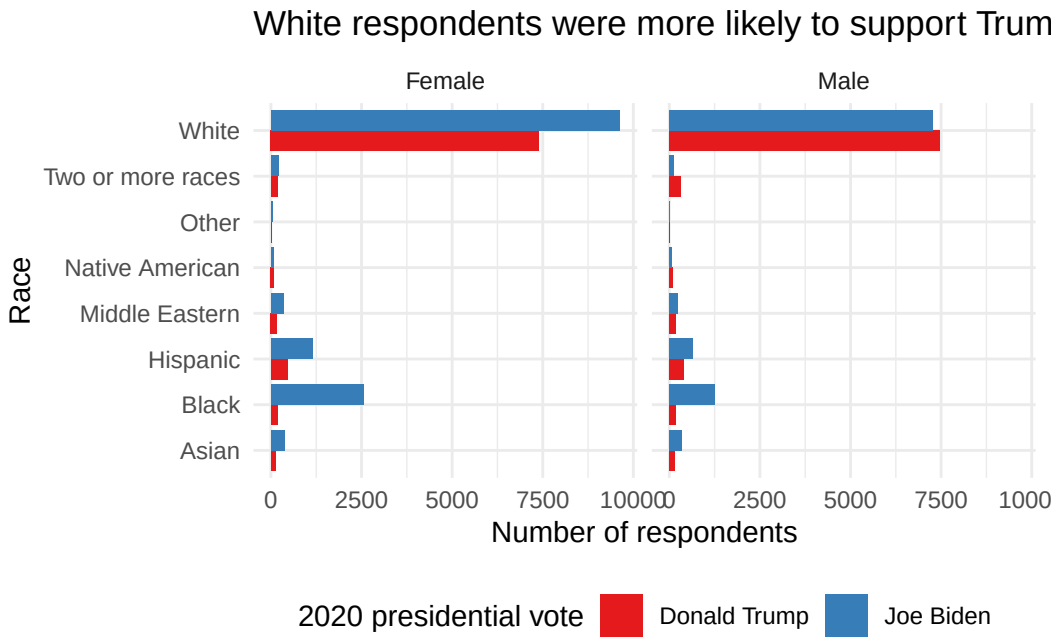
presvote2020	voted_status	n	percentage
Donald Trump	Did Not Vote	5310	30.44376
Donald Trump	Voted	12132	69.55624
Joe Biden	Did Not Vote	7013	28.59415
Joe Biden	Voted	17513	71.40586

2.4 Predictor variables

The tables and visualizations below present possible relationships between predictor variables and either support for Trump or voting in the 2022 midterm election.

2.4.1 Race and gender

In the 2024 U.S. presidential election, there were gender and racial gaps in support for Trump, with more women supporting Vice President Harris and more men supporting Trump. Since 2016, a higher percentage of white voters than voters of color has supported Trump, but in the 2024 presidential election, Trump increased his vote share among voters of color, especially Hispanic voters and black men. As shown in Table 3, white male survey respondents were more likely to support Trump, while black female respondents favored Biden in 2020. The gender gap in support for Trump is also shown, with white men more likely to support Trump than white women in the 2022 CES survey (Schaffner, Ansolabehere, and Shih 2023). Support for Trump varies by race and gender, but these are not the only variables that predict vote preference.



2.4.2 Age

2.4.3 Education

Education gap in American politics.

Race, gender, highest level of education, income, and religion are widely used when predicting political preference. In 2016, 2020, and 2024, pre-election polls and predictive models underestimated support for Trump.

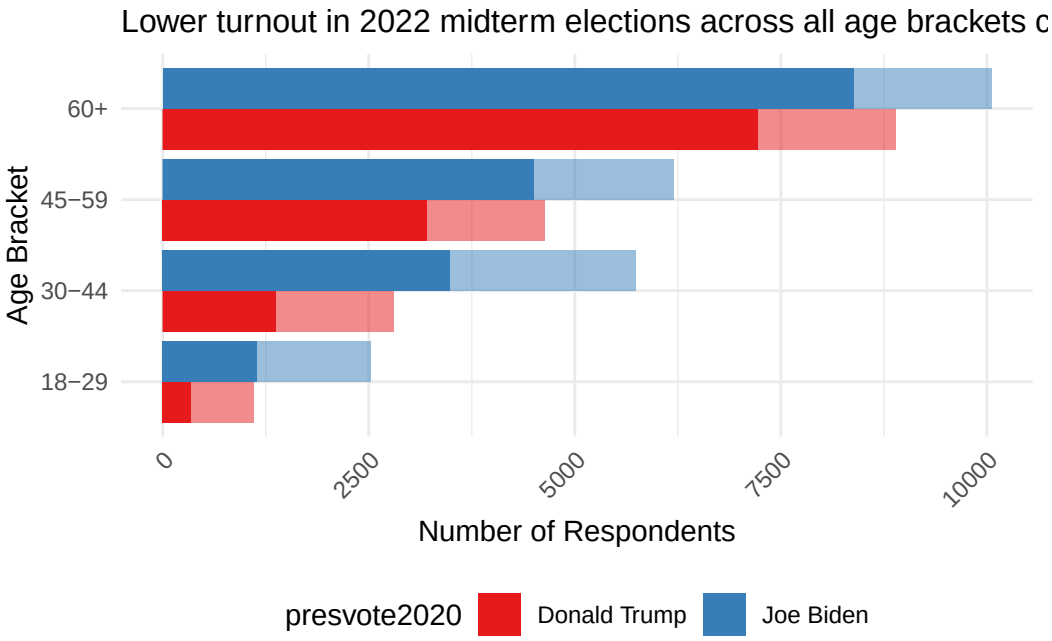


Figure 1: More than half of 18-29 year old respondents who voted for Trump in 2020 presidential election did not vote in the 2022 midterm election.

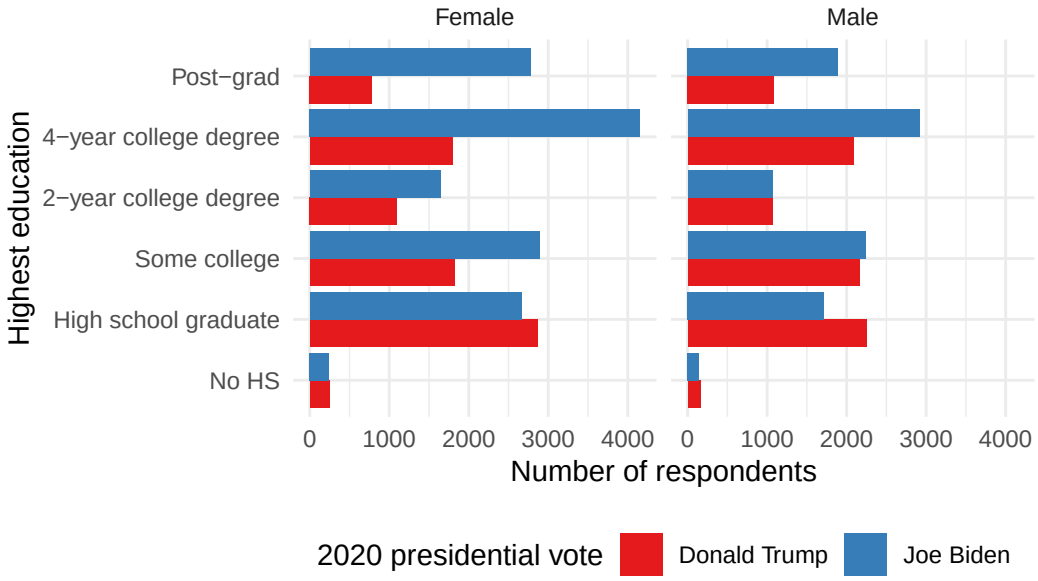


Figure 2: College-educated respondents were less likely to support Trump in 2020

2.4.4 Civic engagement and trust in government

Midterm elections typically have lower turnout than presidential elections. But the drop in turnout between the 2020 presidential election and the 2022 midterm election was not distributed equally across all demographics.

The CES 2022 survey does not contain an explicit question about civic engagement and there is no direct way to determine who is a low-propensity or low-information voter. I have used the following variables to measure civic engagement and trust in government:

- `trustfed`: response to the question about how much they trust the federal government)
- `newsint`: response to survey question about interest in politics
- `CC22_310b`: response to survey question asking which party has a majority in the U.S. House of Representatives.

As shown in Figure 3, survey respondents with less trust in the government favored Trump in 2020.

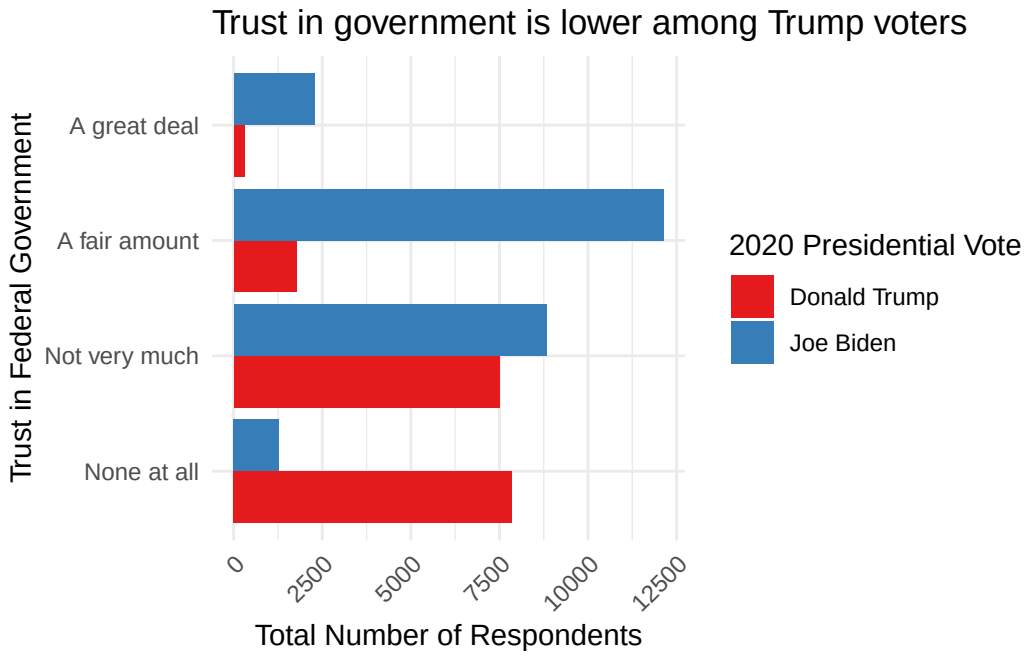


Figure 3: Voters with less trust in the government favored Trump in 2020, while voters with more trust favored Biden in 2020. The bars show the number of respodents with each level of trust in government who voted for Trump and Biden in 2020.

As shown in Figure 4 and Figure 5, respondents who voted in 2020, have at least some interest in politics, and knew which party held a majority in the U.S. House of Representatives turned

out to vote at a higher rate in the 2022 midterm election compared to respondents who do not.

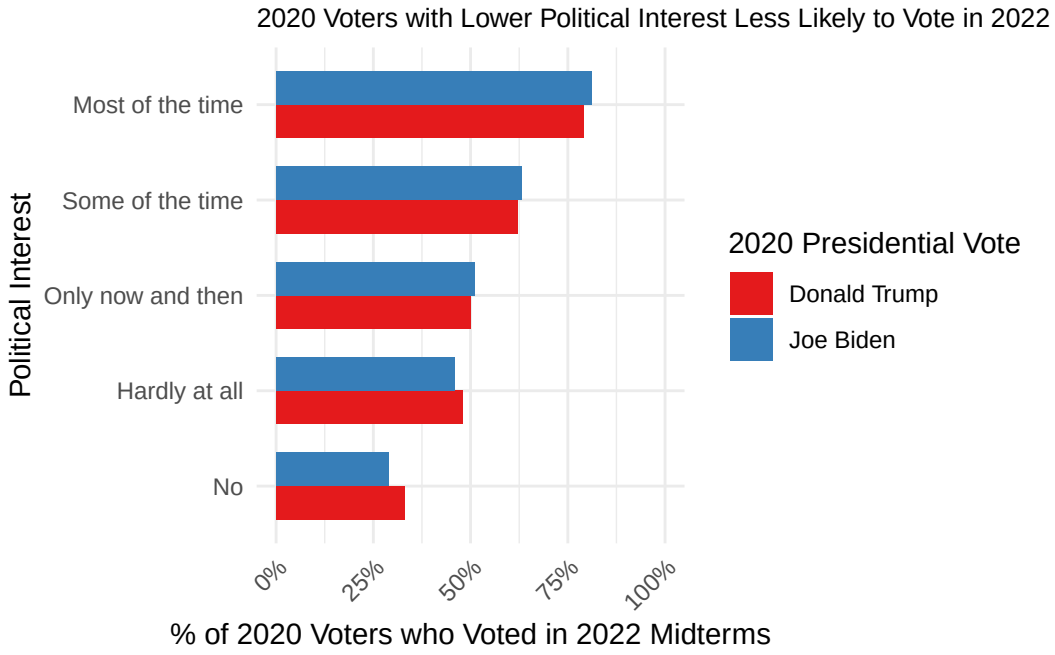


Figure 4: Respondents with low interest in politics who voted in the 2020 presidential election were less likely to vote in the 2022 midterm election.

3 Model

The goal of my modeling strategy is to use logistic regression models to investigate … I use two logistic regression models: one to model the probability that an individual who voted in the 2020 presidential election would also vote in the 2022 midterm election, and one to model the probability that an individual voter supports Trump.

The goal of my modeling strategy is twofold. Firstly, I want to The goal of our modelling strategy is twofold. Firstly,…

The model to predict $voted\_in\_2022$ uses the following predictors:

The model to predict $voted\_for\_trump$ uses the following predictors:

Here, I provide a brief description of how I used logistic regression to predict voter turnout and vote preference, and how my models are set up. Model details and diagnostics can be found in Appendix C.

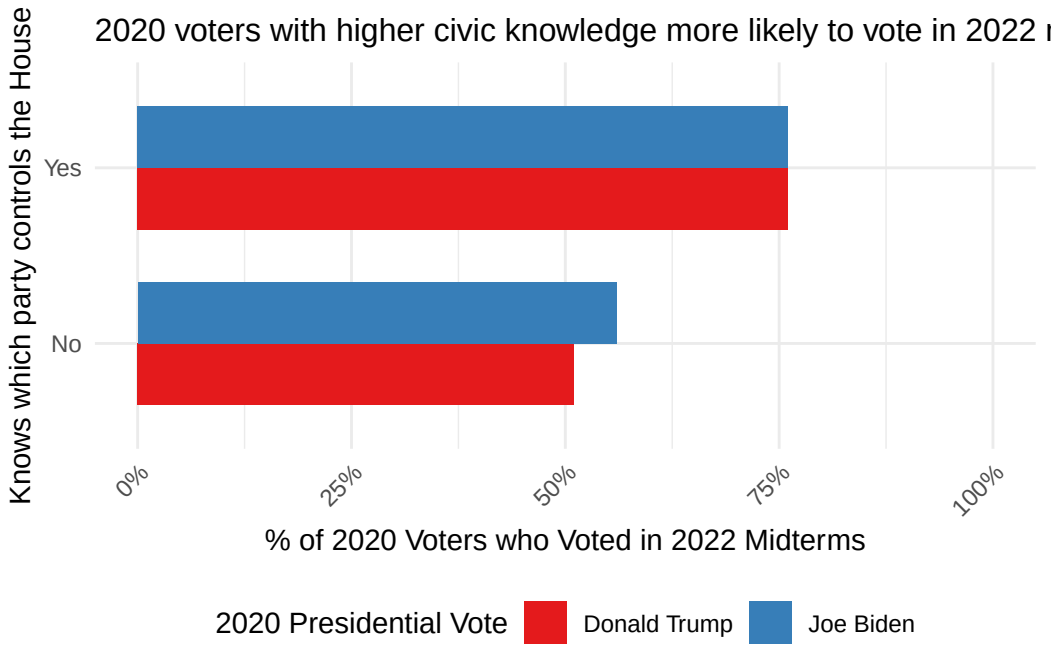


Figure 5: There is a nearly 15-point gap in 2022 turnout amongst respondents who voted for Trump or Biden in 2020 and know which party has a majority in the U.S. House of Representatives versus those who do not.

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.

Models the probability that someone who voted for either Trump or Biden in 2020 also voted in the 2022 mid-term elections.

Define π_i as the probability that survey respondent i voted in the 2022 midterm election

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times presvote2020_i + \beta_2 \times educ_i + \beta_3 \times trust_i + \beta_4 \times know\_us\_house_i \\
&\quad + \beta_5 \times know\_us\_senate_i + \beta_6 \times political\_interest_i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5) \\
\beta_4 &\sim \text{Normal}(0, 2.5) \\
\beta_5 &\sim \text{Normal}(0, 2.5) \\
\beta_6 &\sim \text{Normal}(0, 2.5)
\end{aligned}
$$

Models the probability that a respondent voted for Trump in 2020:

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times age\_bracket_i + \beta_2 \times gender_i + \beta_3 \times race_i + \beta_4 \times educ\_i \\
&\quad + \beta_5 \times trust_i + \beta_6 \times know\_us\_house - i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5) \\
\beta_4 &\sim \text{Normal}(0, 2.5) \\
\beta_5 &\sim \text{Normal}(0, 2.5) \\
\beta_6 &\sim \text{Normal}(0, 2.5)
\end{aligned}
$$

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…

We can use maths by including latex between dollar signs, for instance θ. Confounders have not been eliminated -

I found that trust in the federal government is related to both support for Trump and likelihood that someone who voted in 2020 also voted in 2022. education same thing.

3.2 Model Weaknesses and Limitations

The models were trained using a random subset of the 2022 CES survey data set. My analysis data set (after data cleaning) contained 41,968 observations. I took a random subset of 5000 observations to train my models. This was a trade-off because

Schaffner, Ansolabehere, and Shih (2023) warn that using a small sub sample of the CES survey data to train a model because measurement error can lead to flawed inferences. The CES is a large survey and as a whole, it provides a sufficient number of observations to analyze small sub-populations (Schaffner, Ansolabehere, and Shih 2023).

Trained using a random subset of the data. Refer to codebook about the risks of that. I randomly sampled 5,000 observations from the 2022 CES survey data set. This was a tradeoff; training a model on more observations would have had a significantly longer runtime. However, randomly subsetting the 2022 CES survey data set is risky because …

4 Results

The results are summarized in Appendix C.1.

4.1 Voting blocs that favored Trump

4.1.1 Trump support among young men

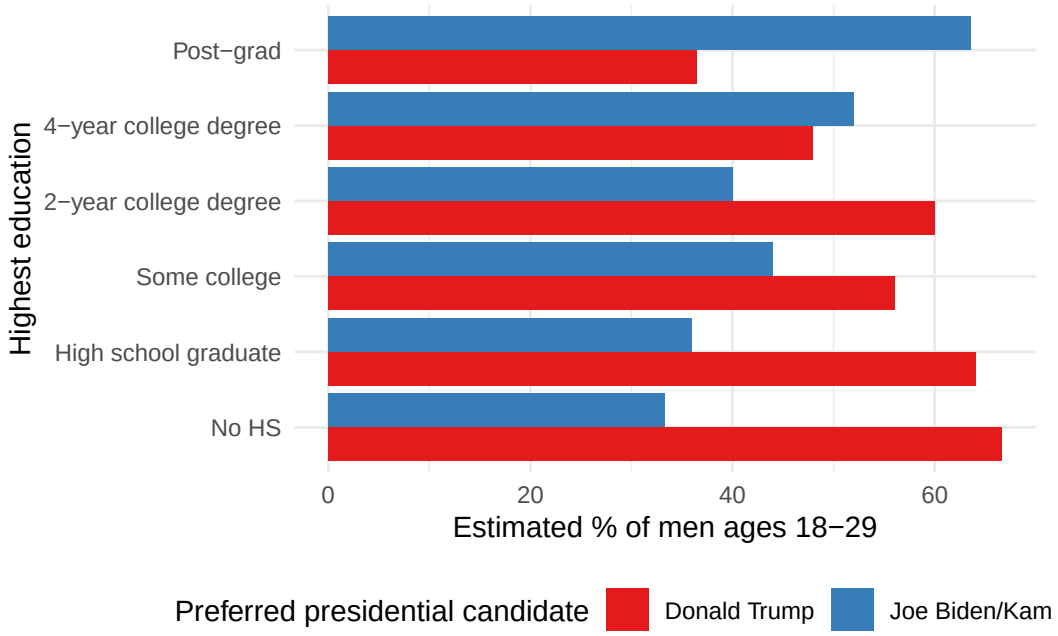


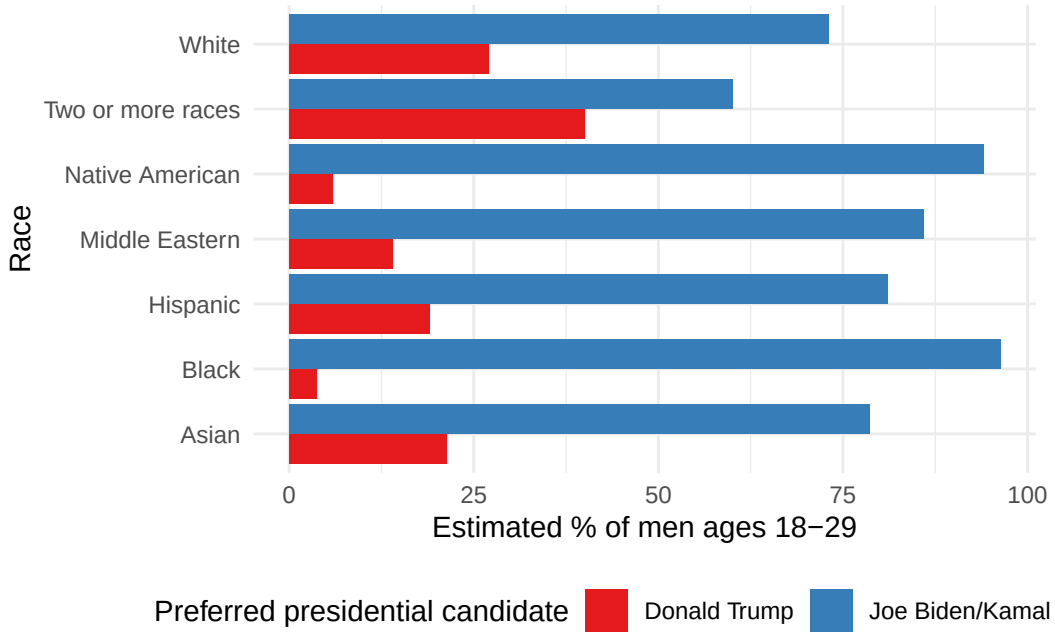
Figure 6: The majority of men ages 18-29 without a college education support Trump. The blue bars show the estimated % of young men with at each level of education that are Biden/Harris/Democratic supporters and the red bars show the % of young men at each level of education that support Trump.

4.2 2022 midterm election turnout

Table 5: Overall, Trump 2020 voters were more likely than Biden 2020 voters to vote in the 2022 midterm election based on model estimates.

Biden 2020 voters	Trump 2020 voters
74.66	75.48

Table 4: The majority of white, hispanic, asian, and multiracial men ages 18-29 support Trump based on model estimates. The red bars



4.2.1 Voters with low interest in politics and low civic engagement are less likely to vote in midterm elections

4.2.2 Low propensity voters that favored Trump in 2020 were less likely to vote in 2022

This result foreshadowed the fact that young men, especially those without a college degree who are infrequent voters, would become a key part of Trump's winning coalition in 2024.

Table 6: Based on model estimates, 2020 presidential election voters with no college education, low trust in government, and low civic engagement/knowledge were likely to support Trump but less likely than the average 2020 voter to also vote in the 2022 midterm election.

Estimated Trump %	Estimated 2022 turnout % among low propensity 2020 voters
69.54	62.56

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 Overlapping predictors for Trump support and not voting in an election where he is not on the ballot

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Polling misses and Trump's unique appeal to infrequent, low-propensity voters

Young men who are not college-educated, do not know which party controls the U.S. House of Representatives,

5.4 2024 Polling Misses and Non-Response Bias

5.5 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Surveys, Sampling, and Observational Data

A.1 Deep Dive into CES Methodology

A.1.1 Sampling Method

A.1.2 Weighting

Table 7: Respondents without a college education were given more weight because they are less likely to respond to a survey.

Highest Education Completed	Average weight
No HS	2.3837276
High school graduate	0.8909107
Some college	0.6702328
2-year college degree	0.6637045
4-year college degree	0.7888546
Post-grad	0.7993014

Table 8: Respondents who voted for Trump in 2020 were given more weight.

2020 presidential vote	Average weight
Donald Trump	0.9091925
Joe Biden	0.7284850

Table 9: Respondents had less trust in government were given more weight because they are less likely to respond to a survey.

Trust in government	Average weight
None at all	0.8871197
Not very much	0.8097035
A fair amount	0.7397012
A great deal	0.8132750

Table 10: Explaining whether someone who voted in the 2020 presidential election also voted in the 2022 midterm election, based on who they voted for, age, education, trust in government, civic knowledge, and political interest

	Voter Turnout Model
(Intercept)	−0.79
	(0.21)
presvote2020Joe Biden	0.24
	(0.08)
age_bracket30-44	0.55
	(0.10)
age_bracket45-59	1.16
	(0.11)
age_bracket60+	1.79
	(0.11)
educ4-year college degree	0.36
	(0.11)
educHigh school graduate	−0.27
	(0.12)
educNo HS	−1.01
	(0.17)
educPost-grad	0.42
	(0.13)
educSome college	0.10
	(0.12)
trustfedA great deal	−0.86
	(0.13)
trustfedNone at all	0.35
	(0.11)
trustfedNot very much	0.39
	(0.08)
truststateA great deal	−0.07
	(0.11)
truststateNone at all	0.15
	(0.10)
truststateNot very much	−0.07
	(0.08)
know_us_houseYes	0.36
	(0.07)
know_us_senateYes	0.02
	(0.06)
political_interestMost of the time	0.28
	(0.15)
political_interestNo	−1.07
	(1.12)
political_interestOnly now and then	−0.32
	(0.17)
political_interestSome of the time	0.00
	(0.16)
Num.Obs.	7500
R2	0.134
Log.Lik.	−3270.706
WAIC	6606.0
RMSE	0.42

C.3 Credibility Intervals

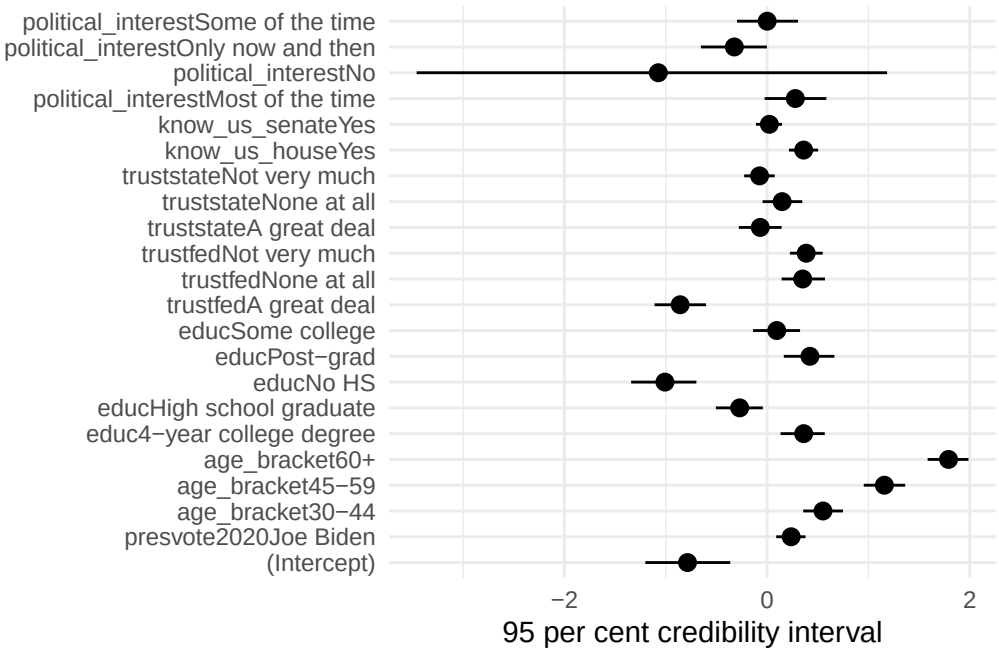
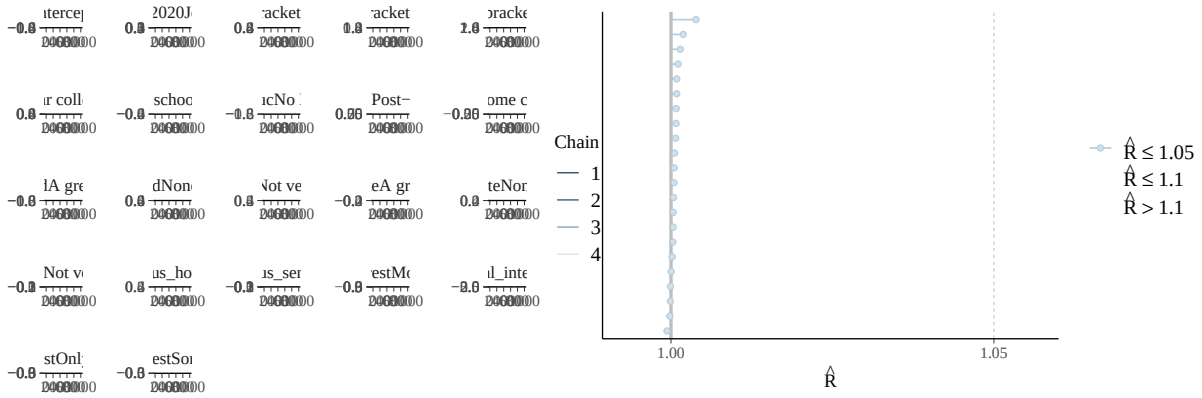


Figure 7: Credible intervals for predictors of voter turnout in the 2022 U.S. midterm elections

C.4 Diagnostics

Figure 8a is a trace plot. It shows… This suggests…

Figure 8b is a Rhat plot. It shows… This suggests…



(a) Trace plot

(b) Rhat plot

Figure 8: 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/.

Grant, Tracy. 2023. *United States Midterm Elections*. Britannica. https://www.britannica.com/topic/midterm-election.

Hartig, Hannah, Andrew Daniller, Scott Keeter, and Ted Van Green. 2023. *Republican Gains in 2022 Midterms Driven Mostly by Turnout Advantage*. Pew Research Center. https://www.pewresearch.org/politics/2023/07/12/voter-turnout-2018-2022/.

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

Schaffner, Brian, Stephen Ansolabehere, and Marissa Shih. 2023. "Cooperative Election Study Common Content, 2022." Harvard Dataverse. https://doi.org/10.7910/DVN/PR4L8P.