Voter Preference and Turnout Analysis*

Voters with no college education, low civic engagement, and low trust in government were among the most likely to support Trump in 2020 and least likely to vote in the 2022 midterm election based on logistic regression modeling

Talia Fabregas

December 12, 2024

The most recent U.S. midterm election took place on November 8, 2022. The 2020 U.S. presidential election set voter turnout records, and as expected, turnout in the subsequent midterm election was lower. In this project, I used logistic regression to estimate what characteristics make a voter more likely to support Trump and which individuals who voted for either Trump or Biden in the 2020 election were more likely to vote in the 2022 midterm election. Based on the results of applying my models, I estimate that not having a college education, low trust in government, and low civic engagement make an individual more likely to support Trump, but less likely to vote in a midterm election.

Table of contents

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

1 Introduction

During even years, elections take place in all 438 seats in the United States House of Representatives and 33 or 34 of the 100 seats in the United States Senate (USAgov 2024). Presidential elections take place every fourth year and midterm elections take place in non-presidential even years (Grant 2023). The most recent U.S. midterm election took place on November 8, 2022 and the most recent U.S. presidential election took place on November 5, 2024. The 2020 presidential election set turnout records, but as expected, turnout decreased in the 2022 midterm election (Hartig et al. 2023).

This project 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 probability 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) .

Based on the results of applying first logistic regression model to predict vote preference, I estimate that young men, voters without a college education, and voters with low trust in government are likely to support Trump. Based on the results of applying second logistic regression model to predict voter turnout, I estimate that presidential election voters with higher civic engagement and civic knowledge, higher trust in government, and higher interest

in politics are more likely to vote in the subsequent midterm election. Based on the results of both my models, I estimate that voters with no college education, low civic engagement, and low trust in government are more are to support Trump and less likely to vote in non-presidential elections. This means that low-propensity Trump voters may be less likely to vote in elections where he is not on the ballot.

In the 2024 election, four Democratic Senate candidates (both incumbent and non-incumbent) were elected in states that President-Elect Trump won (CNN 2024). Three of those four Democratic Senate candidates, Elisa Slotkin (Michigan), Tammy Baldwin (Wisconsin), and Jacky Rosen (Nevada) received fewer votes statewide than Vice President Harris (CNN 2024). It is typical for senate races to have slightly lower turnout than presidential races (Hartig et al. 2023), but there were approximately 117,000 Trump voters in Michigan, 54,000 Trump voters in Wisconsin, and 70,000 Trump voters in Nevada who did not vote for the Republican Senate candidates (CNN 2024). This 2024 election data suggests that there could be a decisive number of low-propensity Trump only voters who have no interest in voting for anyone else. Extensive 2024 pre and post-election data is not available, so this project instead uses 2022 CES data to try to understand who is more likely to vote for Trump and which Trump voters are actually Trump only voters. These patterns highlight the importance of understanding voter behavior, especially among low-propensity Trump voters who may vote in presidential elections but not in midterms, as their future engagement (or lack thereof) could be decisive in future elections.

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. Section 4 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. Section 5 contains detailed a detailed discussion about the results presented in Section 4, 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 A contains an in-depth discussion evaluation of CES 2022 survey methodologies.

2 Data

2.1 Overview

The 2022 Congressional Election Study Common Content (CES) data set from Schaffner, Ansolabehere, and Shih (2023) was used for this project. The data was obtained from Harvard Dataverse on November 27, 2024. It is a nationally-representative survey that aims to study voting behavior in the United States. It contains 60,000 observations and over 700 variables.

I cleaned the 2022 CES data set to only include respondents who voted for either Donald Trump or Joe Biden in the 2020 U.S. presidential election. I used the *pres20votepost* variable, which is each participant's response to a post-election survey wave question about who they voted for in 2020, to determine who each respondent voted for. This is not a perfect way to find out who each respondent voted for because it is subject to social desirability bias (Silver 2024). I am confident that most Trump and Biden 2020 voters are correctly labeled because most CES 2022 respondents' 2020 presidential votes are consistent with their party affiliation or ideological leanings (Schaffner, Ansolabehere, and Shih 2023). However, there is still is a chance respondents who actually voted for Trump may have said that they voted for Biden in 2020 because of Trump's controversial nature or to say that they voted for the winning candidate (Silver 2024). I discuss this in depth in Section 5.4. After sub-setting the data to only include identified Trump and Biden 2020 voters, I used TargetSmart voter file match status to determine which respondents voted in the 2022 midterm election and which respondents did not. Additional data details can be found in Appendix B.1.

I used the statistical programming language R (R Core Team 2023) and the `dplyr`, `janitor`, `ggplot2`, and `kable` packages to clean the data, prepare the data, and create tables and data visualizations.

2.2 Measurement

The 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 data collection:** selected voters from a nationally representative sample respond to a Common Content survey.
2. **Weighting:** survey responses are weighted to adjust for any imbalances that exist in the sample. Respondents who are less likely to answer a survey are given higher weights.
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

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

2020 Vote	Num respondents	%
Joe Biden	24526	58.44

As seen in Table 2, the majority of 2022 CES respondents who voted for either Trump or Biden in the 2020 presidential election were also identified as 2022 midterm election voters. However, it appears that around 30% of 2022 CES respondents who did vote for one of the major candidates in the 2020 presidential election abstained from the 2022 midterm election when neither Trump nor Biden was on the ballot. 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. The remainder of this project aims to study what affects vote preference and which presidential election voters are most likely to vote in the subsequent midterm election and which presidential election voters are most likely to abstain from the subsequent 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

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 **?@tbl-table1**, 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.

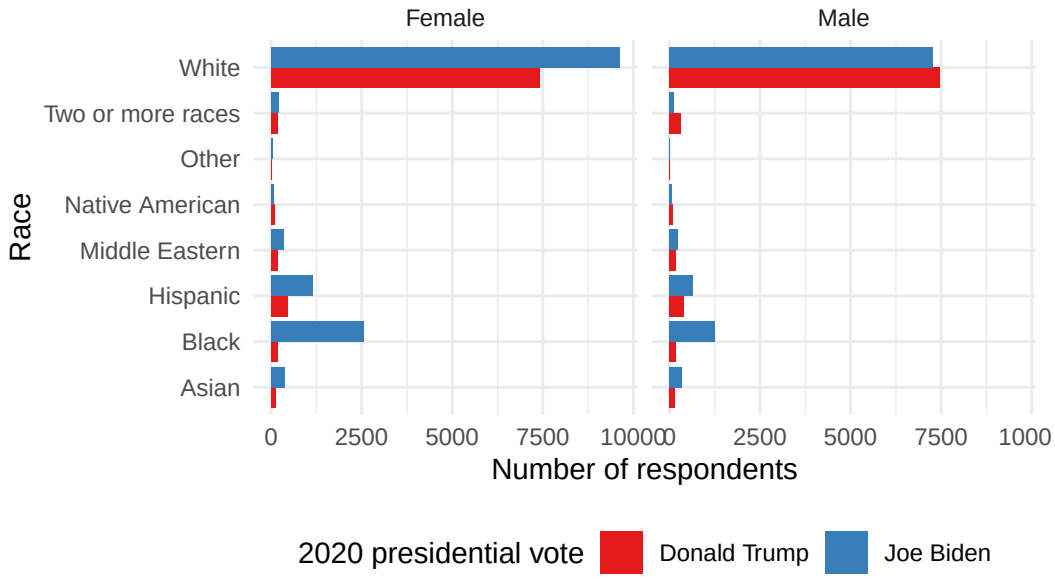


Figure 1: Support for Truo varied among CES respondents, with white men most likely to support Trump

2.4.2 Age

Figure 2 shows that there was lower turnout in 2022 midterm election than the 2020 presidential election across all age brackets. Voters in the 18-29 age bracket, especially those who voted for Trump in 2020, are under-represented in the 2022 CES survey.

2.4.3 Education

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. College-educated Biden voters are overrepresented in the 2022 CES dataset. Voters with no college education are more likely to support Trump and less likely to respond to surveys (Picciotto 2024)

2.4.4 Civic engagement and trust in government

The 2022 CES 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)

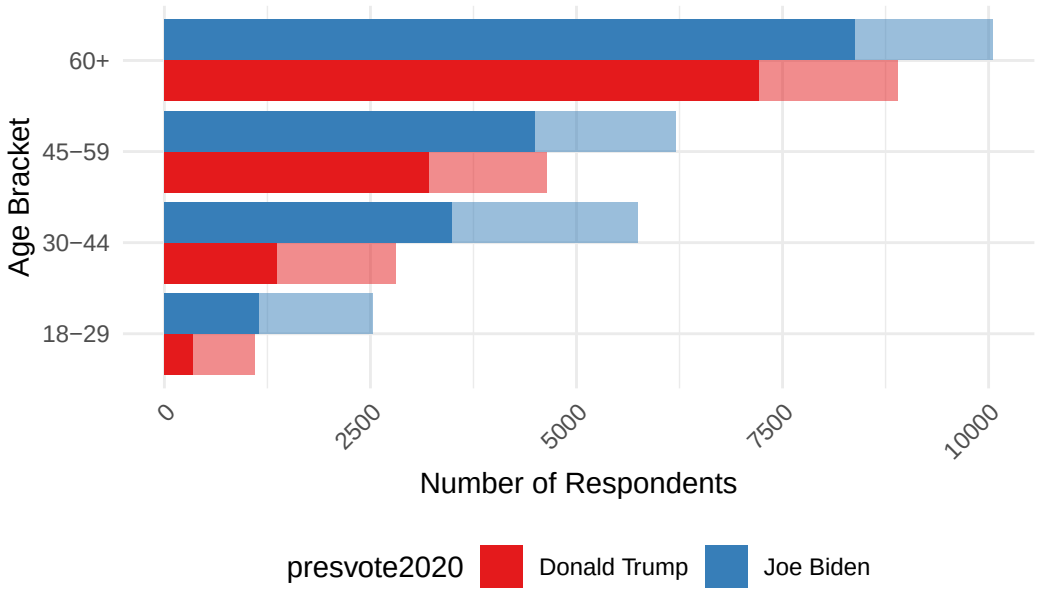


Figure 2: More than half of 18-29 year old respondents who voted for Trump in 2020 presidential election did not vote in the 2022 midterm election. The solid portion of the red and blue bars shows the number of respondents who voted in the 2020 election that also voted in the 2022 midterm.

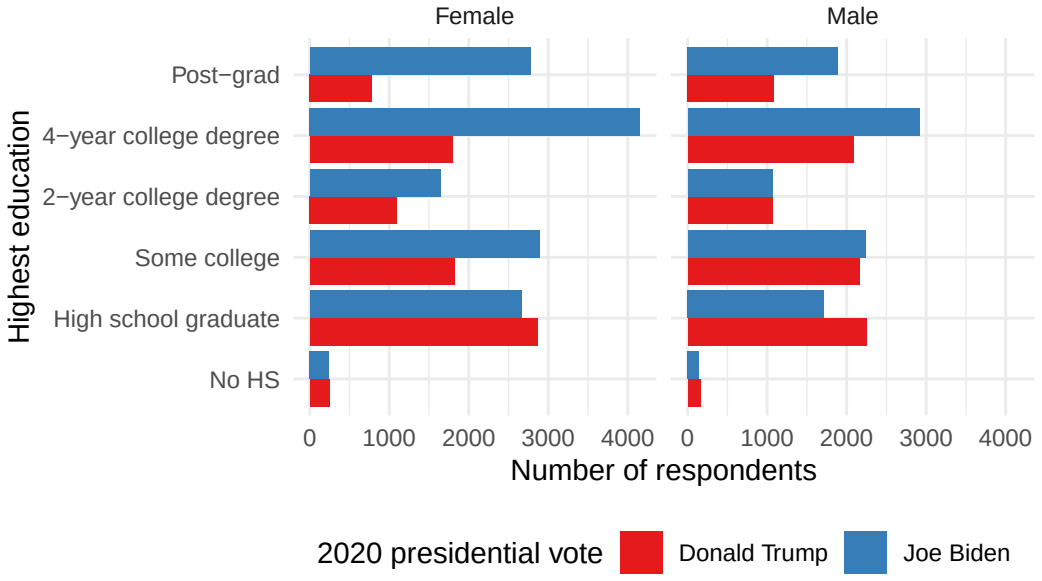


Figure 3: College-educated respondents were less likely to support Trump in 2020.

- `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.

In my data preparation process, I labeled respondents with little to no trust in the government, little to no interest in politics, and who do not know which party controlled the U.S. House as low-propensity voters.

As shown in Figure 4, survey respondents with less trust in the government favored Trump in 2020.

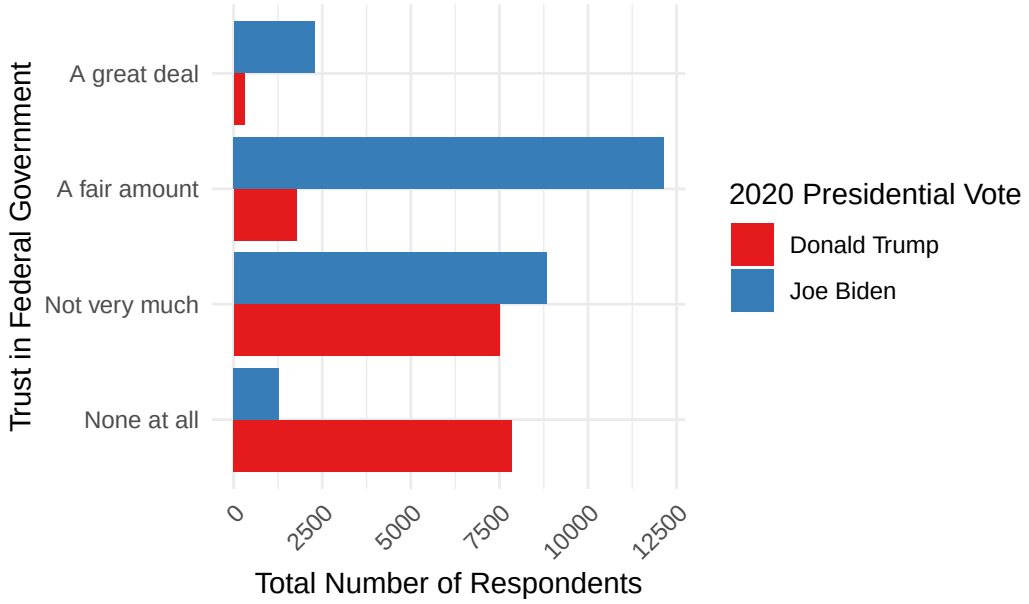


Figure 4: Voters with less trust in the government favored Trump in 2020, while voters with more trust favored Biden in 2020. The bars show the number of respodents with each level of trust in government who voted for Trump and Biden in 2020.

As shown in Figure 5 and Figure 6, 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.

3 Model

The goal of my modeling strategy is twofold. Firstly, I will use logistic regression to investigate the relationship between support for Trump and age, gender, race, education, trust in

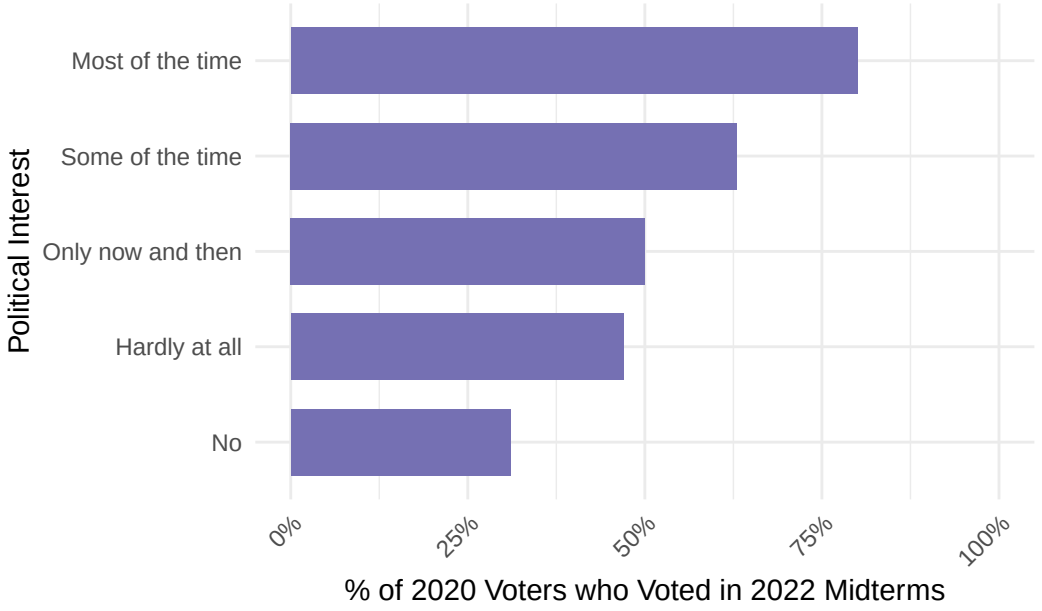


Figure 5: Respondents with low interest in politics who voted in the 2020 presidential election were less likely to vote in the 2022 midterm election.

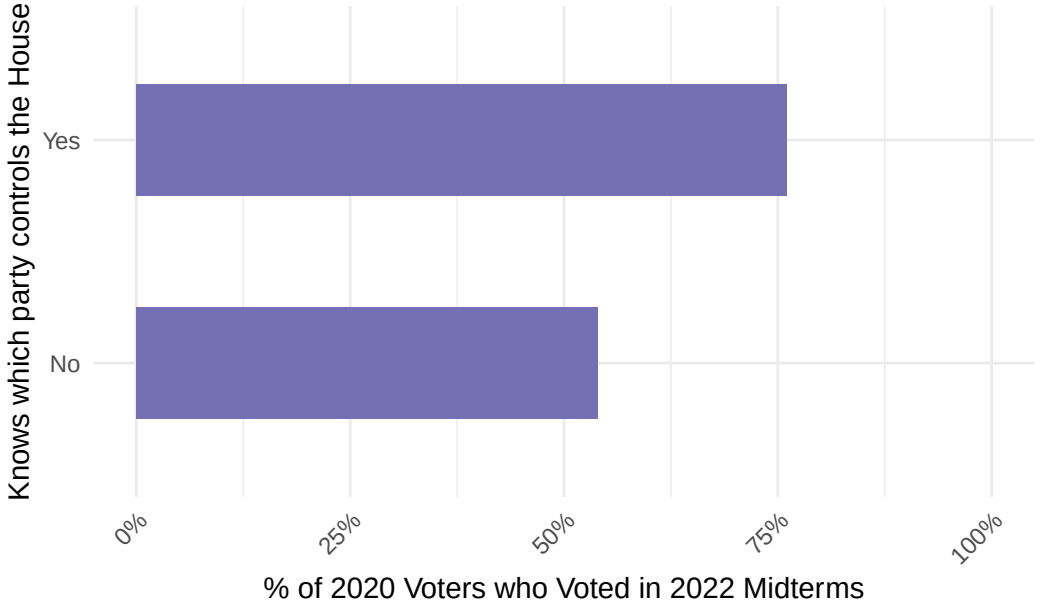


Figure 6: There is a nearly 15-point gap in 2022 turnout amongst respondents who voted for Trump or Biden in 2020 and know which party has a majority in the U.S. House of Representatives versus those who do not.

government, and civic knowledge. Secondly, I will use logistic regression to investigate the relationship between whether someone who voted for Trump or Biden in the 2020 presidential election also voted in the 2022 midterm election and vote choice, age, education, trust in government, civic knowledge, and interest in politics. Model details, validation, checks, and diagnostics can be found in Appendix C.

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 model to predict *voted_for_trump* uses the following predictors:

- `age_bracket`: the age bracket of the respondent (18-29, 30-39, 40-49, 50-59, or over 60). This variable was constructed using the `birthyr` variable in the raw dataset from Schaffner, Ansolabehere, and Shih (2023).
- `gender`: the respondent's gender identity.
- `race`: the respondent's race.
- `educ`: the respondent's highest level of education completed.
- `trust`: the respondent's self-reported level of trust in the federal government.
- `know_us_house`: whether respondent knows which party held a majority in the United States House of Representatives as of 2022.

The model to predict *voted_in_*2022 uses the following predictors:

- `presvote2020`: who the respondent voted for in the 2020 presidential election (only Trump and Biden were considered in this project)
- `age_bracket`: the respondent's age bracket
- `educ`: the respondent's highest level of education completed.
- `trust`: the respondent's self-reported level of trust in the federal government.
- `know_us_house`: whether the respondent knows which party held a majority in the United States House of Representatives as of 2022.
- `political_interest`: the respondent's level of interest in politics.

3.1 Model set-up

The models are run in R (R Core Team 2023) using the `rstanarm` package and default priors from `rstanarm` (Goodrich et al. 2022). The `commonpostweight` weights were used to train both models because they aim to measure the voting behavior of American adults who responded to the pre and post-election waves of the 2022 CES survey (Schaffner, Ansolabehere, and Shih 2023).

3.1.1 Vote choice model

Define π_i as the probability that respondent i supports Trump, $age\_bracket_i$ as their age bracket, $gender_i$ as their gender, $race_i$ as their race, $educ_i$ as the highest level of education that they completed, $trust_i$ as their self-reported level of trust in the government, and $know\_us\_house_i$ as whether or not they know which party had a majority in the U.S. House of Representatives as of 2022:

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

3.1.2 Turnout model

Define π_i as the probability that survey respondent i who voted in 2020 also voted in the 2022, $presvote2020_i$ as who they voted for in the 2020 election (Trump or Biden), $educ_i$ as the highest level of education they completed, $trust_i$ as their level of trust in the government,

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

3.2 Model justification

I used two different logistic regression models to estimate support for Trump and voter turnout. This allows me to investigate how characteristic such as highest educational attainment, trust in government, civic knowledge, and interest in politics affect two different aspects of voting behavior: who someone votes for and whether someone who voted in a presidential election would also vote in a midterm election.

Logistic regression is appropriate for predicting whether an individual supports Trump and whether someone who voted for Trump or Biden in 2020 would also vote in 2022 because both variables have binary outcomes: yes or no (Alexander 2023).

3.3 Model weaknesses and limitations

The use of two separate logistic regression models with overlapping predictors to estimate support for Trump and 2022 midterm election turnout among those who voted for either Trump or Biden in 2020 has weaknesses and limitations. This modeling approach can identify which variables are predictors for both support for Trump and abstaining from a midterm election, but it does not connect the two or address confounders perfectly. To improve my modeling approach in a future project, I would use a Bayesian regression model. Bayesian models take longer to run and the laptop that this project was completed on kept crashing when I tried to use Bayesian models, so my decision to use logistic regression was a trade off to reduce the runtime and computational load (Alexander 2023).

Another potential weakness is the fact that my 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 decision was

a trade-off because training the models on more observations would have had a significantly longer runtime, but randomly subsetting the 2022 CES survey dataset is risky. Schaffner, Ansolabehere, and Shih (2023) warn that using a small sub sample of the CES survey data to train a model is risky because the CES is a large survey that provides just enough observations to analyze small sub-populations and even slight measurement errors can lead to flawed inferences. As a result, there is a chance that my vote preference model and my voter turnout model did not "learn" how to accurately estimate whether members of small subpopulations within the dataset support Trump or would have voted in the 2022 midterm election.

4 Results

The model results are summarized in Appendix C.2.

4.1 Voters that favor Trump

I applied the vote preference model defined in Section 3.1 to estimate which subpopulations were most likely to support Trump between 2020 and 2022.

4.1.1 Young men

As shown in Figure 7 and Figure 8, one of Trump's strongest voter demographics is men between the ages of 18 and 29.

4.1.2 Low propensity voters

Based on my vote preference model estimates, there is a large gap in government trust between Trump and Biden voters. As seen in Figure 9, close to 100% of voters with no trust at all in the government were classified as Trump voters by my logistic regression model, while most voters with a fair amount or a great deal of trust in the government were classified as Biden voters. This shows that low trust in government is a strong predictor of support for Trump.

4.2 2022 midterm election turnout

I applied the turnout model defined in Section 3.1 to estimate which voters who voted for either Trump or Biden in 2020 were most likely to also vote in the 2022 midterm election. My model estimated a very small difference between the percentage of Trump 2020 voters and Biden 2020 voters that would vote in the 2022 midterm election. Based on these estimates, there is no evidence that vote choice (Trump or Biden) has any significant impact on the likelihood of a presidential election voter participating in the subsequent midterm election.

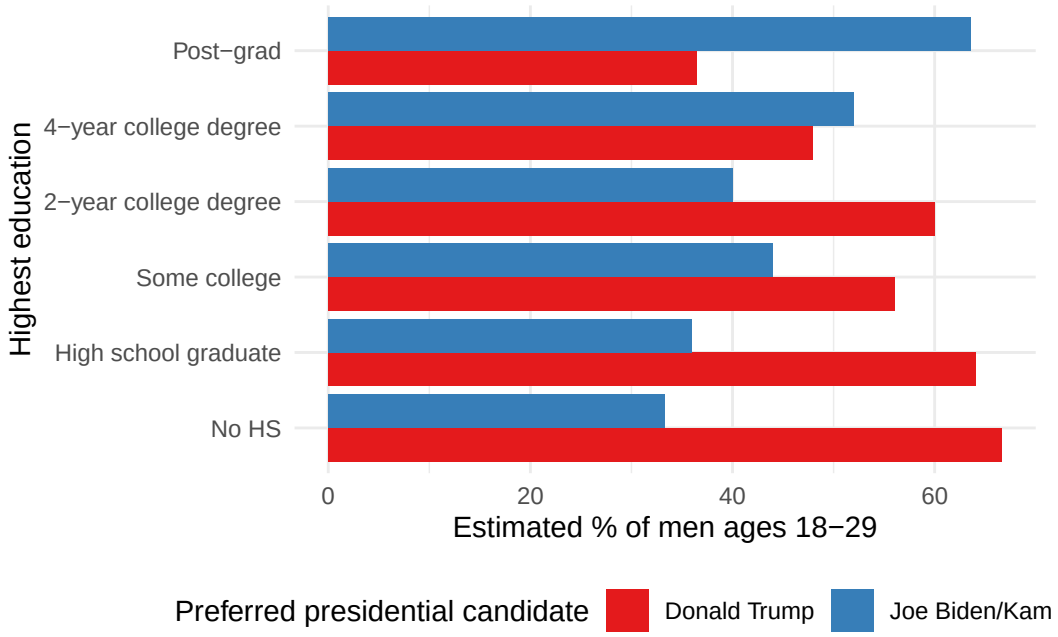


Figure 7: The majority of men ages 18-29 without a 4-year college degree support Trump, based on model estimates. The blue bars show the estimated % of young men with at each level of education that are Biden/Harris/Democratic supporters and the red bars show the % of young men at each level of education that support Trump.

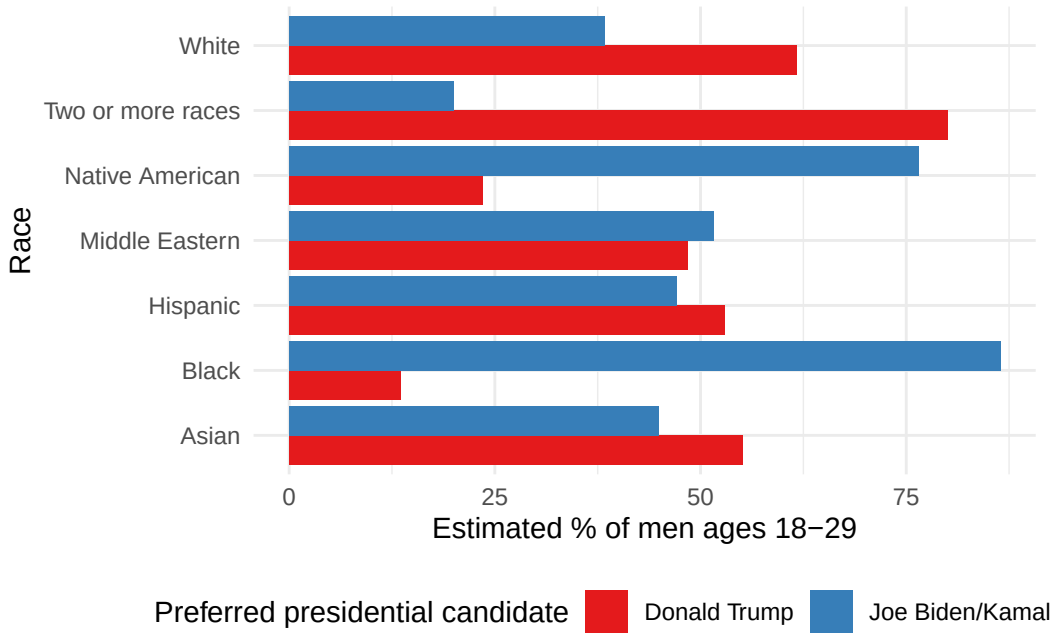


Figure 8: The majority of white, hispanic, asian, and multiracial men ages 18-29 support Trump based on model estimates. The red bars

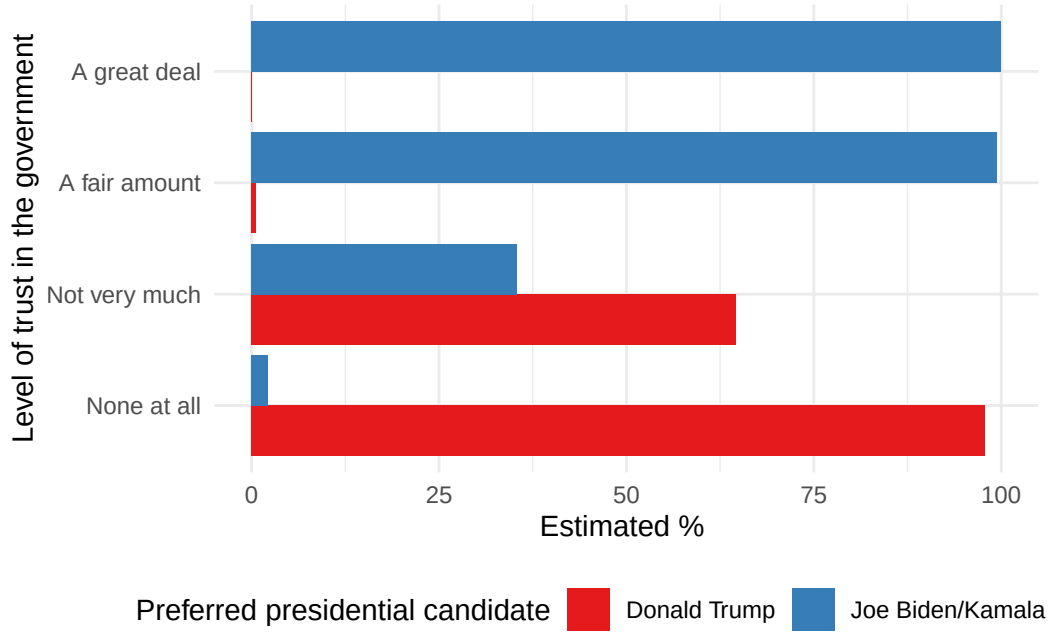


Figure 9: Nearly all voters with no trust in the government were classified as Trump voters by my logistic regression vote preference model

Table 3: Overall, Trump 2020 voters may have been a bit 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

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

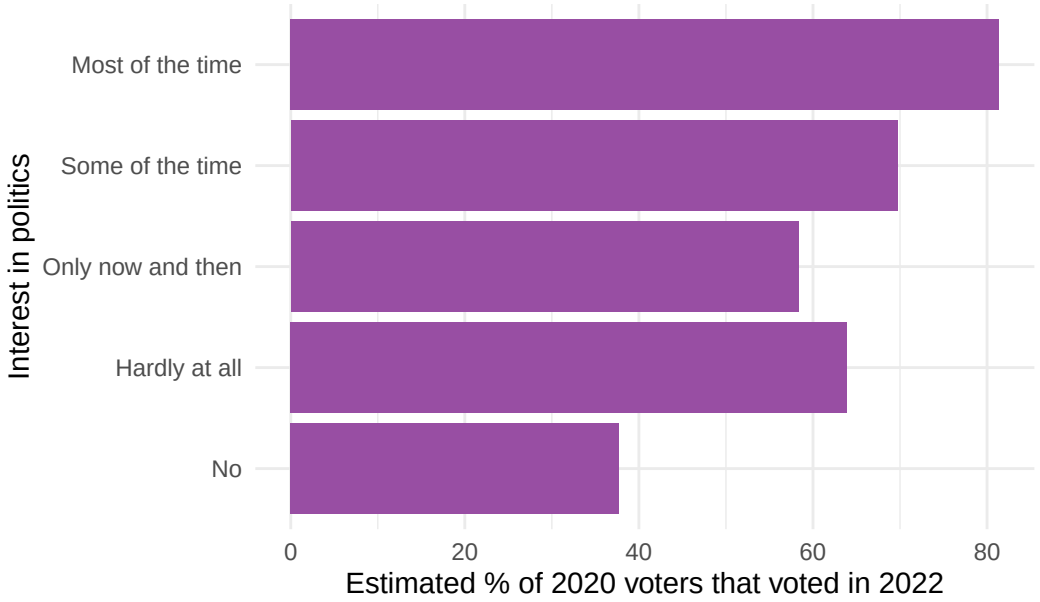


Figure 10: Based on model estimates, 80% of Trump and Biden 2020 voters who are informed about and interested in politics most of the time also voted in 2022, compared to just over 40% of 2020 voters with no interest in politics

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

As shown in Table 4, my voter preference model estimates that almost 70% of voters with no college education, low trust in government, and low civic engagement would support Trump. If we refer back to Table 3, an estimated 75% of all Trump 2020 voters also voted in the 2022 midterm election. This means that Trump 2020 voters with no college education, no trust in the government, and low civic engagement were less likely to vote in the 2022 midterm election when their preferred candidate, Trump was not on the ballot.

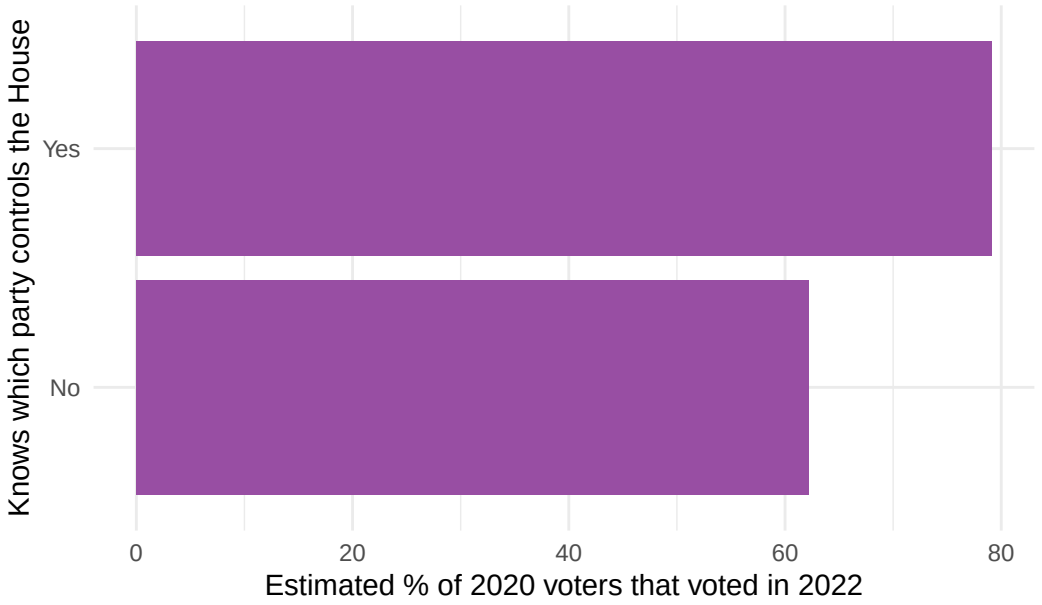


Figure 11: 2020 voters who did not know which party had a majority in the House were less likely to vote in 2022

Table 4: 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 % of 2020 voters that voted in 2022
69.54	62.56

5 Discussion

5.1 Measuring civic engagement and trust in government

5.2 Young men were an important part of Trump's 2024 winning coalition

In Section 4, I found across most races and education levels, a majority of men between the ages of 18 and 29 support Trump. This foreshadowed the 2024 election, where young men without a college degree became a major part of Trump's winning coalition (CNN 2024). Based on my model estimates presented in Section 4, nearly two-thirds of men between the ages of 18 and 29 without a college education support Trump.

My results were produced by logistic regression models trained on the 2022 CES survey dataset. In Section 2, I found that both Trump voters and voters between the ages of 18 and 29 were underrepresented in the dataset, likely due to non-response bias. The 2022 CES survey is a reliable dataset that was constructed by 60 research teams, supported by Harvard University, and used sample-matching and online respondent recruitment (Schaffner, Ansolabehere, and Shih 2023). Even though its data acquisition process was good, there are observations that we know it does not have, but must be measured (Alexander 2023). The dataset included weights that could be used to train a model to measure voting behavior, and respondents who were part of underrepresented sub-populations or less likely to respond to a survey were given higher weights (Schaffner, Ansolabehere, and Shih 2023). That means that even though men between the ages of 18 and 29 without a college education who support Trump might have been less likely to opt in to the YouGov survey platform and respond to the 2022 CES survey, those who did respond were given higher weights (Schaffner, Ansolabehere, and Shih 2023). Their responses were missing not at random (MNAR) because the characteristics that made them less likely to respond to surveys were also related to who they support (Alexander 2023). The use of weights allowed my logistic regression models that were trained on the 2022 CES survey dataset to identify characteristics that make voters more likely to support Trump.

5.3 Polling misses and the "Trump only" voter

Silver (2024) found that Trump voters were less likely to respond to pollsters because they are more likely to distrust the government and have low civic engagement. After applying my logistic regression models to estimate vote preference and voter turnout, I found that voters with no college education, low or no trust in the government, and low civic engagement heavily favor Trump. I also found that Trump 2020 voters with these same characteristics were close to 14 percentage points less likely to vote in the 2022 midterm election than the average 2020 presidential election voter. These are likely the same voters that pollsters failed to account for when they underestimated support for Trump in 2016 and 2020.

My results suggest that the same characteristics that made low-propensity Trump voters less likely to respond to a pollster also made them less likely to vote in the 2022 midterm election when Trump was not on the ballot. What this means for the success of Republican candidates in the 2026 midterm election and in the 2028 presidential election remains unclear. Low-propensity Trump voters are less likely to respond to surveys, but they might also be less likely to vote in elections where he is not on the ballot.

Pollsters will need to use survey incentives and new multi-channel respondent recruitment strategies to reach this important, but hard-to-reach group in 2026 and 2028 . There are two ways that pollsters could inaccurately estimate support for the 2028 Republican presidential nominee. If they do not find a way to survey enough Republican voters, especially ones with low government trust and civic engagement, they could repeat their 2016 and 2020 polling

misses. Pollsters could overestimate support for the 2028 Republican nominee if they over-correct their 2016, 2020, and 2024 polling misses and ignore the 2024 Michigan, Wisconsin, and Nevada senate results.

Assuming that low-propensity Trump voters will vote (or not vote) a certain way in 2026 and 2028 will lead to polling misses one way or another. They have been difficult to survey since 2016, so there is limited data about Trump voters who did not support down-ballot Republicans available. New polling strategies, such as shortening surveys so that respondents do not find them tedious and promoting surveys on X (formerly known as Twitter) and other social media platforms widely used by low-propensity voters, will be needed to reach enough of them.

5.4 Weaknesses and next steps

This project's greatest weakness is the fact that it uses non-response biased data from 2022 to try to understand and identify the Trump only voter. The 2022 CES survey dataset from Schaffner, Ansolabehere, and Shih (2023) contains nearly one-third more Biden 2020 voters than Trump 2020 voters. Trump voters are missing not at random (MNAR) from this dataset: their low survey response rates are correlated to their low trust in government and low civic engagement (Alexander 2023). Throughout this project, I have tried to connect my results, which do show that the same characteristics that made low-propensity Trump voters unlikely to respond to surveys also made them less likely to vote in the 2022 midterm election to the 2024 U.S. Senate results in Michigan, Wisconsin, and Nevada. In these three states that Trump won in 2024, close to 250,000 voters did not vote in the Senate races (CNN 2024). This fact and my results both suggest, but do not confirm, that the could be a significant share of low-propensity Trump voters who are Trump only voters, and are likely to abstain from elections that he does not run in.

My next step towards understanding the "Trump only" voter would be to compare the results of 2024 pre-election polls to the actual results, do a deep-dive into 2024 pollster methodology, and conduct my own survey that targets low-propensity voters. Comparing the 2024 pre-election polls to the actual results and analyzing 2024 pollster methodology in depth will allow me to understand how much the polls underestimated support for Trump, yet again, what they did to try to address the non-response bias from 2016 and 2020, and how well it worked.

Low-propensity Trump only voters are extremely difficult to survey, so one of my future steps could be to design and distribute a short, clear, and concise survey using non-traditional methodologies. Current survey methodologies, such as stratified sampling, assigning higher weights to subpopulations that are harder to survey, and weighting on recalled vote have not worked (Silver 2024). Non-response bias has plagued pollsters since 2016. Causes of non-response and vote preference are correlated. The use of stratified sampling to collect a represenative sample and post-stratification weighting to correct imbalances within that sample is based on the assumption that the political preferences of the respondents within each

stratnum are representative of the actual voters within that stratnum who were not surveyed (Alexander 2023). Non-traditional voter blocks require non-traditional survey and respondent recruitment methodologies. Although this recruitment method is not backed by Alexander (2023) or supporting literature, I would develop a survey that takes less than five minutes to complete, asks only a few straightforward questions about demographics, and simply asks respondents who they voted for in the presidential race, and who they voted for/if they voted at all in a U.S. house or senate race. I would then promote the survey it where many young, low-propensity Trump only voters are: X and Parler. I would advertise it as a way for them to affirm their support for Trump, instead of using traditional respondent recruitment methods. Stantcheva (2023) did find that short surveys and simple questions reduce attrition, so there is some literature to justify this.

Once I build or obtain a survey dataset that successfully measures voting behavior of all Americans, including the ones who are most difficult to survey, I will develop a very specific logistic regression model to estimate whether or not an individual who voted for Trump or Harris in 2024 also voted in the U.S. Senate race on the same ballot based on vote preference, age, education, civic engagement, trust in government, and interest in politics.

Appendix

A Surveys, Sampling, and Observational Data

A.1 Deep dive into 2022 Cooperative Election Study (CES) survey methodology

A.1.1 Overview and survey objectives

The 2022 CES survey is a collaboration between 60 research teams that aims to study public opinion and voting behavior across the United States (Schaffner, Ansolabehere, and Shih 2023). It is part of the ongoing Cooperative Election Study (CES), previously known as the Cooperative Congressional Election Study (CCES). It consists of 60,000 responses from a nationally-representative sample of American adults. Each research team purchased a 1000 person nationally representative survey from YouGov (Schaffner, Ansolabehere, and Shih 2023). The 2022 CES survey was conducted online by YouGov and 60,000 American adults were surveyed in two waves: the pre-election wave (September 29 -November 8, 2022) and the post-election wave (November-December 2022) (Schaffner, Ansolabehere, and Shih 2023). Sample matching was used to build a nationally representative sample and weighting was used to further ensure that the sample is representative (Schaffner, Ansolabehere, and Shih 2023).

A.1.2 Population and sampling frame

The target population of a survey is the population that it aims to represent and speak about (Alexander 2023). The target population of the 2022 CES survey is all American adults (Schaffner and Kuriwaki 2022). Researchers using the 2022 CES survey dataset can subset it to only include respondents who matched to a validated voter registration record if they wish to study registered voters (Schaffner, Ansolabehere, and Shih 2023). The sampling frame of a survey is all potential respondents from its target population (Alexander 2023). The 2022 CES survey's sampling frame is American adults who are YouGov panelists and receive notifications about surveys (Schaffner and Kuriwaki 2022).

A.1.3 Sampling approach

The 2022 CES survey uses sample matching instead of stratified sampling to build a representative sample (Schaffner, Ansolabehere, and Shih 2023). Stratified sampling is a probability-based sampling approach that allows researchers to examine every stratnum of a population and perform simple random sampling within the strata (Alexander 2023). Sample matching is ideal for online suveys because it allows researchers to match respondents who are as similar as possible to members of the target sample (Schaffner, Ansolabehere, and Shih 2023). Stratified

sampling and sample matching have the same goal: to ensure that every strata of the target population is sufficiently represented in the sample (Alexander 2023).

Instead of using a traditional stratified sampling approach, Schaffner, Ansolabehere, and Shih (2023) drew a target sample, defined as a random sample from their target population of all American adults. The target sample represents every strata of the target population, or what their ideal representative sample would look like if they had used a stratified sampling approach (Schaffner, Ansolabehere, and Shih 2023). Stratified sampling is not realistic for the 2022 CES because it recruits its respondents from the YouGov online survey platform and cannot contact them directly (Schaffner, Ansolabehere, and Shih 2023). Instead of randomly sampling from each strata, Schaffner, Ansolabehere, and Shih (2023) find one or more opt-in YouGov respondent who matches each member of their target population using a large set of demographic, geographic, and ideological variables and a similarity function. Schaffner, Ansolabehere, and Shih (2023) repeat this process until they have "matched" every member of their target sample with a similar enough opt-in survey respondent.

A.1.4 Data validation

Schaffner, Ansolabehere, and Shih (2023) used the TargetSmart database of registered voters in the United States to validate individual voter records. The TargetSmart database can only be used to validate which general and party primary elections a respondent has voted in, and what method of voting they used, not who they voted for.

Respondents are matched to a TargetSmart voter record if there is a high level of confidence that the information they provided matches the record (Schaffner, Ansolabehere, and Shih 2023).

A.1.5 Weighting and data adjustments

Schaffner, Ansolabehere, and Shih (2023) used weighting to address non-response bias in the 2022 CES survey dataset. The 2022 CES survey dataset is weighted to adjust for any imbalances that still exist after sample matching is complete (Schaffner, Ansolabehere, and Shih 2023). Respondents who were underrepresented or less likely to respond to a survey, such as respondents with no college education, respondents who supported Trump in 2020, or respondents with low trust in government, were assigned higher weights. Respondents who are more likely to respond to a survey, including those with a college-education or interest in politics were assigned lower weights.

Table 5: 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 6: 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 7: 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

A.1.6 Strengths and limitations

One strength of the 2022 CES survey is its design, specifically the concise questions and the length of the post-election wave. According to Schaffner and Kuriwaki (2022), the pre-election wave is designed to be completed by respondents in around 20 minutes and the post-election wave is designed to be completed by respondents in around 10 minutes. Stantcheva (2023) identifies survey length versus the need to collect data as a trade-off that needs to be made when designing surveys. Attrition, defined by Alexander (2023) is the rate at which respondents drop out before completing a survey.

Initially, I was concerned about the 20-minute length of the pre-election wave because longer surveys are at risk of having a high attrition rate (Stantcheva 2023). However, with the amount of information about demographics, income, positions on issues like abortion, guns,

and the economy, voter registration status, past voting behavior, and 2022 vote intentions it collects, the length is worth it. 85% of respondents who answered the pre-election wave also answered the post-election wave a few weeks later (Schaffner, Ansolabehere, and Shih 2023). The overall survey design also addresses concerns about the length of the pre-election survey. Most questions are not required, so if a respondent wants to leave a question blank and move on, they are able to do that. Incomplete responses are collected, and the dataset just includes an `NA` value if a respondent left a question blank. The 10-minute length of the post-election wave is a major strength because it manages to collect sufficient information about respondents' voter registration status, how they voted (or didn't) in the different 2022 midterm election races, and their positions on hot issues following the midterm election.

The 2022 CES survey has a lot of questions, but its questions are clear, and in plain language. This is a strength because short questions, especially in a longer survey, help reduce the cognitive load, the risk of voters finding the survey too tedious to answer, and the attrition rate (Stantcheva 2023). Examples of 2022 CES survey questions from Schaffner, Ansolabehere, and Shih (2023) include:

- Who did you vote for in the election for president in 2020? Joe Biden; Donald Trump; Jo Jorgensen; Howie Hawkins; Other; Did not vote for president.
- Do you intend to vote in the 2022 general election on November 8th? Yes definitely; Probably; I already voted (early or absentee); I plan to vote before November 8th; No; Undecided.

The 2022 CES survey used only an online survey conducted by YouGov to collect responses. It does not use multi-channel respondent recruitment. This is a weakness because the types of individuals who sign up for online survey platforms like YouGov and willingly respond to online surveys about politics are not random. Individuals with time, access to a reliable computer, high civic engagement, and trust in government are more likely to respond. That makes the 2022 CES survey susceptible to non-response bias. Non-response bias happens when observations, or in this case, survey responses, are missing not at random (MNAR) (Alexander 2023). Survey responses are MNAR when there is a correlation between the reason that they are missing and one of the variables being examined (Alexander 2023). Silver (2024) found that Trump voters are less likely to trust the government, and low trust in government makes an individual less likely to respond to a survey. Multi-channel recruitment and survey incentives such as a small financial compensation, which the 2022 CES survey does not use, are effective ways to increase survey participation within hard-to-reach populations (Alexander 2023).

A.2 Identifying non-voters and social desirability bias

Schaffner, Ansolabehere, and Shih (2023) matched 2022 CES survey respondents to the TargetSmart registered voter database in August 2023. They matched a respondent to a TargetSmart record when there was a very high level of confidence that the respondent was being

matched to the correct record (Schaffner, Ansolabehere, and Shih 2023). The most common reason why a respondent was not matched to a TargetSmart record was because they were not registered to vote, but incomplete or inaccurate information, such as a recent name or address change could also prevent a match (Schaffner, Ansolabehere, and Shih 2023). If a respondent matched to a TargetSmart voter record, then information about whether they voted in 2022, and the voting method they used if they voted is available (Schaffner, Ansolabehere, and Shih 2023). The 2022 CES survey dataset does not include a variable that explicitly indicates whether an individual respondent voted in the midterm election. Instead, it includes information about whether each respondent matched to a TargetSmart record and whether they reported voting or non-voting in the post-election wave (Schaffner, Ansolabehere, and Shih 2023).

Schaffner, Ansolabehere, and Shih (2023) outlined three approaches for identifying respondents who did not vote in the 2022 midterm election: labeling both un-matched respondents and matched non-voters as non-voters, labeling only matched non-voters as non-voters or labeling both matched non-voters and un-matched respondents, who are self-reported non-voters as non-voters. Each approach has strengths and limitations. Schaffner, Ansolabehere, and Shih (2023) recommend labeling both un-matched respondents and matched non-voters as 2022 midterm non-voters because it is the best way to ensure that no non-voter will be falsely labeled as a voter. However, the weakness of this approach is that it assumes that all unmatched respondents did not vote in 2022. Labeling only matched non-voters as 2022 midterm voters accounts for the possibility that some 2022 CES respondents who did not match to a TargetSmart voter record may have actually voted (Schaffner, Ansolabehere, and Shih 2023).

Using self-reported participation instead of a blanket assumption to determine whether a respondent who did not match to a TargetSmart voter record addresses the tradeoffs associated with assuming that all un-matched respondents are non-voters (or vice versa). However, this approach relies on the assumption that social desirability bias was not a factor (Schaffner, Ansolabehere, and Shih 2023). Social desirability bias happens when a survey respondent becomes inclined to answer a question in a way that they think would make them "look good" to others (Alexander 2023). There is a chance that unmatched respondents might be inclined to say that they did vote in 2022 because participating in the democratic process is seen as socially desirable behavior.

B Additional data details

B.1 Data cleaning

Data cleaning and preparation was done using R (R Core Team 2023) and the `dplyr` (Wickham et al. 2023) and `janitor` (Firke 2023) packages. The original dataset from Schaffner, Ansolabehere, and Shih (2023) contained 60,000 observations and over 700 variables. 50,981 of

the 60,000 respondents took the post-election survey and I required responses to post-election questions, so I subsetted to only include respondents who participated in the post-election survey. I then subsetted the data it to include only respondents who voted for either Trump or Biden in 2020 using the `filter` function from Wickham et al. (2023). My analysis did not require the use of over 700 variables, so I used the `select` function from Wickham et al. (2023) to select the ones I needed.

The variables from the raw data that I used in my analysis data include:

- *TS_g2022*: which method of voting the respondent used in the 2022 midterm election. NA if the respondent was not matched to a TargetSmart voter file (Schaffner, Ansolabehere, and Shih 2023).
- *presvote2020post:* who the respondent voted for in the 2020 presidential election
- *race:* the respondent's race
- *birthyr:* year of birth (this was used to construct the age, and eventually *age_bracket* variables).
- *gender4:* gender identity
- *educ:* highest level of education completed
- *CC22_423*: the respondent's self-reported level of trust in the Government in Washington, D.C.
- *CC22_310a*: whether or not the respondent knew which party had a majority in the U.S. House of Representatives
- *newsint:* how often the respondent is interested in following what's going on in U.S. government and politics.

B.1.1 Identifying 2022 non-voters

I constructed a new variable, *voted_in_*2022, which is equal to 1 if the respondent voted in the 2022 midterm election and 0 if they did not. I identified non-voters as respondents who did not match to a TargetSmart voter file (respondents where *TS_g2022* was `NA`). According to Schaffner, Ansolabehere, and Shih (2023), identifying un-matched voters as non-voters is a justified approach because the most common reason why a respondent will not have a TargetSmart voting record is because they are not registered to vote. Self-reported non-voting rates are very high among respondents who did not match to a TargetSmart record (Schaffner, Ansolabehere, and Shih 2023). There is a small chance that this approach could falsely label un-matched voters without a TargetSmart match as non-voters, but it is the best way to make sure that no non-voters are falsely labeled as voters.

C Model details

This section provides additional details about the models, including model summaries, posterior predictive checks, credibility intervals, and model diagnostics.

C.1 Model summary

C.2 Model results

C.2.1 Vote preference model

The model summary for the vote preference model is shown in Table 8.

C.2.2 Turnout model

The model summary for the voter turnout model is shown in Table 9.

C.3 Credibility Intervals

C.3.1 Vote preference model

Figure 12 shows the 95% credibility intervals for the coefficients of predictors used in the vote preference model.

C.3.2 Turnout model

Figure 13 shows the 95% credibility intervals for the coefficients of predictors used in the voter turnout model.

C.4 Diagnostics

C.4.1 Vote preference model

Figure 14a is a trace plot. It shows horizontal lines that bounce and overlap in between chains. This suggests that the vote preference model does not have any major issues (Alexander 2023).

Figure 14b is a Rhat plot. It shows values very close to 1.0. This suggests that the predictors used in the vote preference model do not have any major issues and the model does not need to be simplified (Alexander 2023).

Table 8: Explaining whether someone voted for Trump or Biden in 2020 based on race, gender, age, education, trust in government, civic engagement

	Voter Turnout Model
(Intercept)	−2.92
	(0.29)
age_bracket30-44	0.36
	(0.12)
age_bracket45-59	0.69
	(0.12)
age_bracket60+	0.89
	(0.11)
genderMale	0.46
	(0.07)
genderNon binary	−3.36
	(1.04)
genderOther	0.25
	(1.13)
raceBlack	−1.56
	(0.21)
raceHispanic	−0.42
	(0.21)
raceMiddle Eastern	−0.70
	(0.29)
raceNative American	−0.44
	(0.41)
raceOther	−1.54
	(1.16)
raceTwo or more races	0.65
	(0.37)
raceWhite	0.60
	(0.17)
educ4-year college degree	−0.32
	(0.12)
educHigh school graduate	0.45
	(0.13)
educNo HS	0.11
	(0.19)
educPost-grad	−0.63
	(0.13)
educSome college	−0.15
	(0.13)
trustfedA great deal	−0.13
	(0.17)
trustfedNone at all	3.71
	(0.10)
trustfedNot very much	1.85
	(0.08)
know_us_houseYes	−0.31
	(0.08)
political_interestMost of the time	0.29
	(0.17)
political_interestNo	−76.69
	(64.18)
political_interestOnly now and then	0.79
	(0.19)
political_interestSome of the time	0.57
	(0.18)
Num.Obs.	7500

Table 9: 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

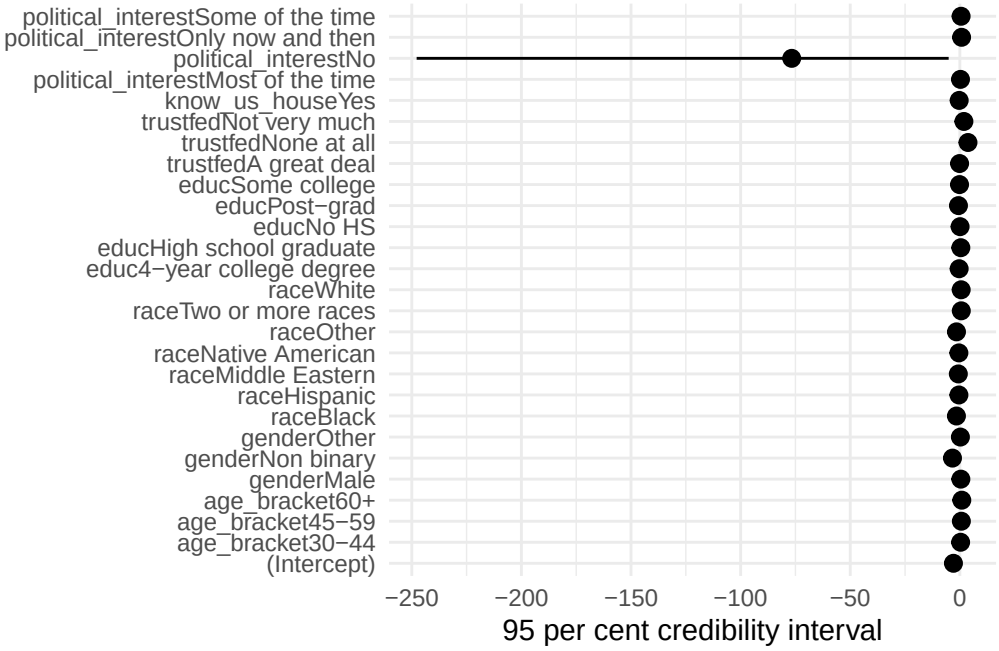


Figure 12: Credible intervals for predictors of support for Donald Trump

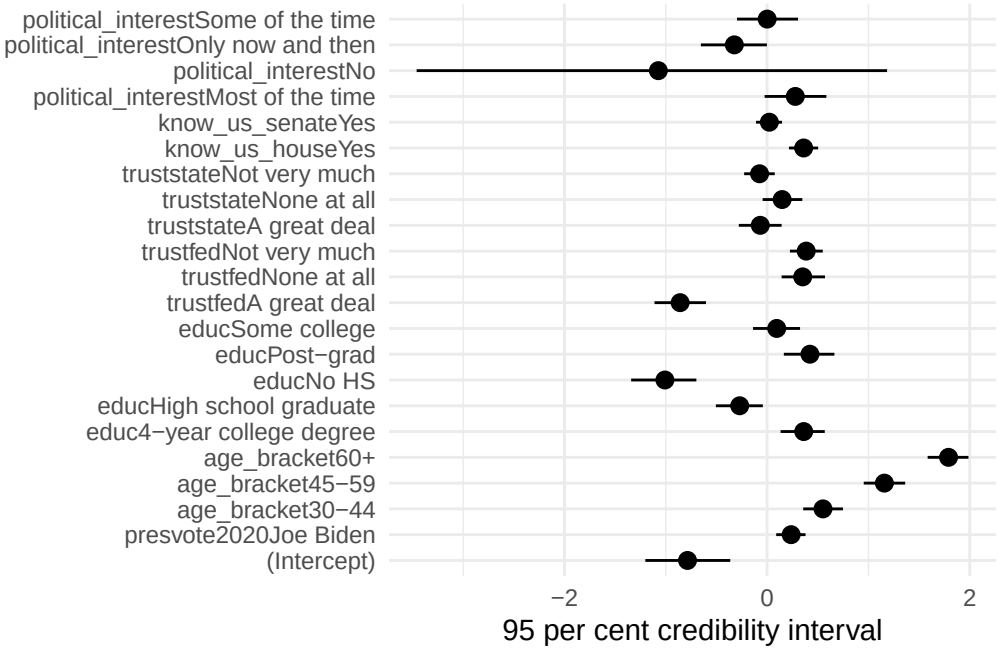
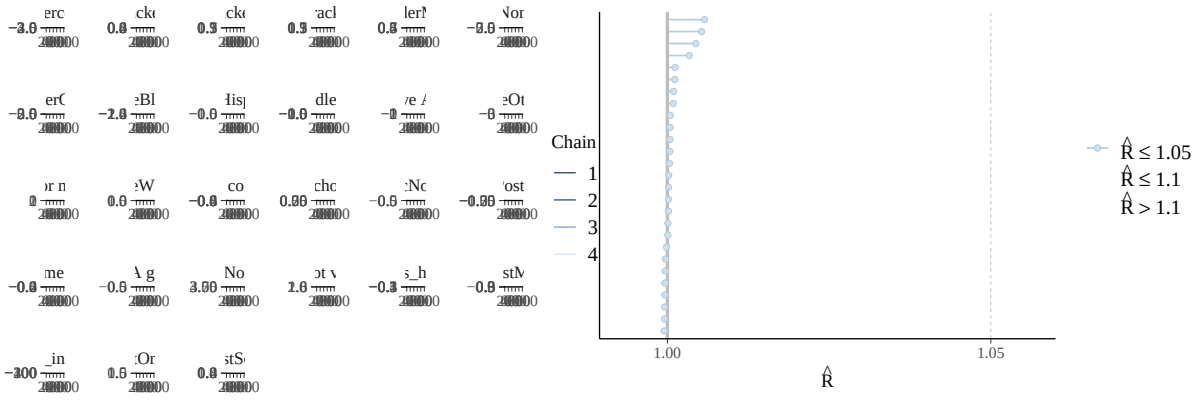


Figure 13: Credible intervals for predictors of voter turnout in the 2022 U.S. midterm elections



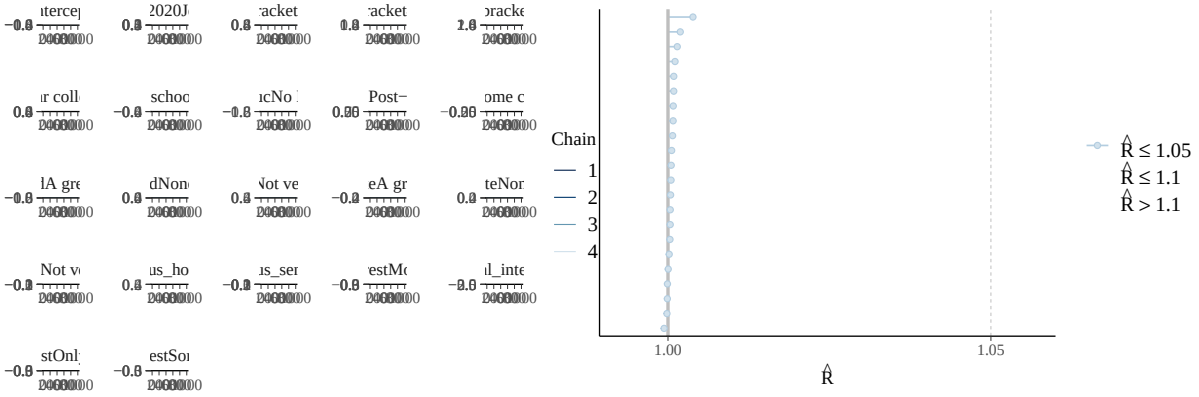
(a) Trace plot (b) Rhat plot

Figure 14: Checking the convergence of the MCMC algorithm

C.4.2 Turnout model

Figure 15a is a trace plot. It shows horizontal lines that bounce and overlap in between chains. This suggests that the turnout model does not have any major issues (Alexander 2023).

Figure 15b is a Rhat plot. It shows values very close to 1.0. This suggests that the predictors used in the turnout model do not have any major issues and the model does not need to be simplified (Alexander 2023).



(a) Trace plot (b) Rhat plot

Figure 15: Checking the convergence of the MCMC algorithm

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

CNN. 2024. *Election Day in America*. https://www.cnn.com/election/2024/results/president?election-data-id=2024-PG&election-painting-mode=projection-with-lead&filter-key-races=false&filter-flipped=false&filter-remaining=false.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

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/.

Picciotto, Rebecca. 2024. "Why Election Polls Were so Wrong in 2016 and 2020 — and What's Changing to Fix That." https://www.cnbc.com/2024/05/04/why-election-polls-were-wrong-in-2016-and-2020-and-whats-changing.html.

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.

Schaffner, Brian, and Shiro Kuriwaki. 2022. *Frequently Asked Questions*. Harvard University. https://cces.gov.harvard.edu/frequently-asked-questions.

Silver, Nate. 2024. "Nate Silver: Here's What My Gut Says about the Election, but Don't Trust Anyone's Gut, Even Mine." https://www.nytimes.com/2024/10/23/opinion/election-polls-results-trump-harris.html?unlocked_article_code=1.UU4.pFkQ.F2hD-woxmiEj&smid=url-share.

Stantcheva, Stefanie. 2023. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." *Annual Review of Economics* 15: 205–34. https://www.annualreviews.org/content/journals/10.1146/annurev-economics-091622-010157.

USAgov. 2024. *Congressional Elections and Midterm Elections*. United States Government. https://www.usa.gov/midterm-elections.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.