Voting Patterns Unveiled*

The Impact of Demographics and Economic Outlook in the 2020 Election

Carl Fernandes Lexi Knight Raghav Bhatia

March 12, 2024

This paper investigates voter demographics and economic outlook on preferences for candidates in the 2020 US presidential election. It was found that race and education played pivotal roles in determining support for Biden over Trump. Moreover, economic factors influenced voter sentiment with perceptions of economic hardship impacting candidate favorability. Awareness of these complex dynamics is imperative for grasping the landscape of American politics especially with the upcoming US election this November.

Table of contents

*Code and data are available at: https://github.com/1raghav-bhatia/Voter-Outcomes.git

1 Introduction

As the curtains rose on the 2020 presidential election stage, the dynamic interplay between demographics, economic outlook and voter preferences took center stage. With divergent voices and values at play, the battle between Donald Trump and Joe Biden underscored the intricate tapestry of American society, where race, gender and the state of the economy painted a complex picture of electoral dynamics. On November 3rd, 2020, the United States presidential election took place where the top two candidates were Joe Biden of the democratic party and Donald Trump of the republican party. Biden won the majority with 306 electoral votes over Trump with 232. Our estimand is how certain voter characteristics and opinions affected voter preferences in the 2020 US presidential election. In this paper, we analyze how demographics including race, income and education affected which voters were more inclined to vote for Biden over Trump. Additionally, we investigate the effect that two economic variables, namely

economic outlook and income change, had on a respondents inclination to vote in favor of Biden.

Previous research uses the American Trends panel conducted shortly after the 2020 US general election as well as surveys conducted in 2018 on 10,000 panelists. This research studies race, illustrating Black, Asian and other races supported Biden with an overwhelming 92% of Blacks voting for Biden (Igielnik 2021). On the other hand White and Hispanic voters were found to support Trump. In terms of gender, men supported Biden and women supported Trump. When looking at a combination of race and gender, there is a similar trend, with white men supporting Biden and white women supporting Trump. Education's impact on voting preferences showcase that Biden supporters had at least a four-year college education or higher in comparison to Trump voters who had less than that. The amalgamation of race and education that voted for Trump were Hispanics and whites without a four-year educational accomplishment. Conversely, Biden supporters were strictly those that had a minimum four year college experience and race did not impact this relationship (Igielnik 2021).

Our data portrays that a significant share of Bidens voters had a very negative view on the economy. Additionally it was found that there was more female support toward Biden. Black, Hispanic, and Asian voters voted for Biden. Furthermore, individuals with a four-year college degree or higher were in support of Biden. Finally, in looking at education by race, both white and non-white college graduates as well as non-white voters without a college degree voted for Biden, leaving whites without a college degree to support Trump. Majority of voters who perceived that the economy had stayed stable or gotten worse overwhelmingly supported Biden. A comparison of our current economy and the 2020 economy offers an idea as to whether voters' outlook on the economy will play a role in the upcoming 2024 presidential election, as it did in 2020.

The paper starts with the data section to understand and visualize the demographic and economic outlook variables. Next, we introduce the model in order to observe the variable's relationship with voter preference. Moreover, we provide visual representations of the findings in the results section. Finally, in the discussion, we summarize the main takeaways, suggest areas of improvement as well as potential areas of future research.

2 Data

The analysis is based on the Cooperative Election Study (CES), from 2020 which contains a wealth of data on voter registration, preferences, demographics, and various opinions on national issues. The coding language and associated packages used are `R` (R Core Team 2023), the `tidyverse` (Wickham, Henry, and Müller 2023), `boot`(Canty and Ripley 2023), `broom.mixed` (Bolker and Robinson 2023), `collapse`(Lingl 2023) ,`dataverse` (Blume-Kohout et al. 2023), `arrow` (Neal et al. 2023), `rstanarm` (Goodrich et al. 2023), `knitr` (Xie 2023), `janitor` (Firke 2023), `marginaleffects` (Arel-Bundock 2023), `here` (Müller 2023), and `modelsummary` (Arel-Bundock and Pilon 2023), `readr` (Wickham and Hester 2023) and `Kable Extra` (Zhu

2023).The CES was conducted during a time of increased political activity and social discussions in the United States, providing a snapshot of the electorate during the 2020 Presidential election—a year marked by significant social and economic changes.

This dataset is notable for its extensive coverage across geographical regions and the depth of topics covered. With over 60,000 respondents it offers a broadened perspective on voters, allowing for detailed examination of voter behavior and preferences beyond just party affiliations. It delves into the details of voter demographics and their nuanced views on matters.

The main goal of utilizing this dataset is to understand how conditions around the 2020 election influenced voter preferences. The analysis focuses on determining whether individuals' views on economics and personal financial situations played a role in their decision between President Trump and former Vice President Biden. The emphasis on how economic factors influence voting patterns becomes crucial considering the economic conditions experienced in 2020 marked by significant changes triggered by the impact of the COVID 19 outbreak.

2.1 Description of Variables and Context

2.1.1 Voter Registration (votereg)

The variable `votereg` indicates if participants are registered to vote. Being registered to vote is an important indicator of political involvement and a necessary step for taking part in elections. This variable plays a role in filtering the data to focus on those who vote, ensuring that subsequent analyses accurately represent the behaviors and preferences of real participants in the voting process.

2.1.2 Preference for Presidential Candidates (voted_for)

The `voted_for` variable shows the choice made by respondents between President Donald Trump and former Vice President Joe Biden in the 2020 election. This distinction offers insights into the political landscape of the United States during that period and acts as a primary measure for studying how other factors impact voter choices.

2.1.3 Gender (gender)

The recording of `gender` aims to understand the composition of respondents and its potential impact on voting behavior. Gender dynamics often influence preferences and priorities, making this variable significant for our analysis. The equal distribution between female respondents, in this dataset mirrors the overall voting populace allowing for a balanced examination of gender related trends or differences.

2.1.4 Education Level (educ)

The `educ` category indicates the level of education attained by individuals ranging from "No High School Diploma" to "Postgraduate Degree." Education plays a role in shaping awareness, policy preferences and voting tendencies. Examining the backgrounds of voters can provide insights into the complexity of the electorate and potential relationships between achievements and support for candidates.

2.1.5 Ethnicity (race)

Ethnicity encompasses an array of identities, such as White, Black, Hispanic, Asian, Native American, Middle Eastern and others. As an aspect of group identity, ethnicity can influence political beliefs and alliances. Analyzing the `race` variable helps in understanding how racial demographics may have influenced leanings in elections given the increased attention to racial issues during the 2020 campaign season.

2.1.6 Household Income Changes (CC20_303)

The `CC20_303` delves into how individuals have experienced shifts by inquiring about their household's yearly income changes. It classifies these changes into five levels; "Increased significantly", "Increased", "Remained stable", "Decreased" and "Decreased significantly". Given the fluctuating economic climate impacted by the events of 2020, especially due to the COVID 19 pandemic, these responses hold significant relevance. They offer a foundation to gauge peoples outlook and may be linked to their voting preferences as financial well-being or challenges often play a role in shaping political inclinations.

2.1.7 National Economic Perception (CC20_302)

The `CC20_302` variable captures people's views on the economy's performance. Respondents were asked to assess whether within the year the country's economy has "Improved markedly" "Improved", "Stayed consistent", "Declined moderately", "Declined significantly" or if they were uncertain about it. This important economic indicator is vital as it could mirror the contentment or discontent of voters with how the current administration's managing economic matters and may play a significant role in their selection of a candidate.

Adding these factors with voter registration and demographic details offers a set of data allowing for a thorough examination of the 2020 U.S. Presidential Election, from various angles. The subsequent investigation will dive into the specifics of these factors, uncovering the intricacies of voters' views on the economy and how they might impact election results.

2.2 Alternative Datasets and Justification

Researchers studying elections have access to a variety of datasets, each with its own focus and level of detail. While the American National Election Studies (ANES) and the General Social Survey (GSS) are well regarded for their insights into voter behavior and attitudes, the Cooperative Election Study (CES) from 2020 was chosen for this analysis for several reasons.

The CES dataset is known for its breadth and timeliness, capturing a range of variables directly related to the 2020 election. Unlike ANES and GSS which provide perspectives over extended periods CES offers a view of the political landscape during a critical election year. Its large sample size allows for an examination of voter demographics and behaviors during a time when the U.S. was grappling with social and economic challenges.

On the other hand, while ANES offers longitudinal data it may not offer as much specificity on economic variables crucial to our hypothesis—that voters' economic perceptions and experiences leading up, to the 2020 election significantly influenced their voting choices. GSS, known for its range, in sciences may not capture the immediacy of economic sentiments as quickly as CES does.

Furthermore, CES's methodological strengths lie in its sample size that allows for breakdowns by demographics while still representing a broad cross section of the American public. It delves into nuanced factors such as changes in household income and the national economic outlook, which are crucial to the hypothesis being explored.

When it comes to the dataset construction, CES's approach to gauging economic perceptions is particularly relevant as it reflects both nationwide trends and individual experiences. It goes beyond asking respondents about their opinions on the economy to inquire about shifts in their own household income creating a direct link between macroeconomic patterns and personal financial situations.

Ultimately selecting CES for this analysis was a decision based on its provision of relevant data that is up to date for investigating the research question at hand. The dataset's emphasis on the 2020 election and incorporation of factors corresponds with our goal to analyze the economic factors that might have impacted voting patterns.

2.3 Data Cleaning and Preparation

Ensuring the accuracy of the analysis involves cleaning and preparing the data. In the case of the 2020 Cooperative Election Study (CES) dataset various steps were taken to improve the quality and dependability of the findings.

2.3.1 Initial Data Review

The first step was to examine the dataset for any discrepancies, missing data or unusual values that could impact the analysis. This included checking how responses were distributed across variables to confirm they fell within expected ranges and followed the guidelines outlined in the CES codebook.

2.3.2 Data Cleaning and Modification

Transforming data played a role in preparing it for analysis. Numeric variables were converted into formats for interpretation. For instance:

- The `gender` variable was changed from codes (1, for male, 2, for female) to categories "Male" and "Female."
- The variable 'education' was changed from numbers to categories, with labels like "No High School", "High School Graduate", "Some College", "2 Year College", "4 Year College" and "Post Graduate."
- The factors `CC20_303` and `CC20_302` which represent shifts in household income and views on the economy were converted from values to descriptive categories ensuring each option was labeled meaningfully based on the survey choices.

2.3.3 Subsetting

The focus of the analysis was on registered voters who expressed a preference for one of the candidates. Therefore the dataset was filtered to include individuals with a `votereg` value of 1 (registered voters) and those whose `CC20_410` values indicated support for either Biden or Trump.

2.3.4 Handling Missing Data

Entries containing incomplete information were managed appropriately—either through imputation or exclusion depending on the severity and type of missing data. For example if there were a few cases with missing data, for the race variable those instances were excluded to ensure the analysis accuracy.

2.3.5 Validation and Testing

Throughout the data cleaning process we conducted tests to confirm the changes and ensure that the data followed formats. These tests checked for factor levels, no missing values after imputation and proper application of recording rules.

We meticulously documented the data cleaning and preparation procedures to maintain transparency and reproducibility of our findings. In summary, by cleaning and preparing the CES dataset we have established a foundation for subsequent analysis. The next section will explore exploratory data analysis where visualizations of cleaned data will be used to uncover patterns and insights to behaviors and economic sentiments of the 2020 American electorate.

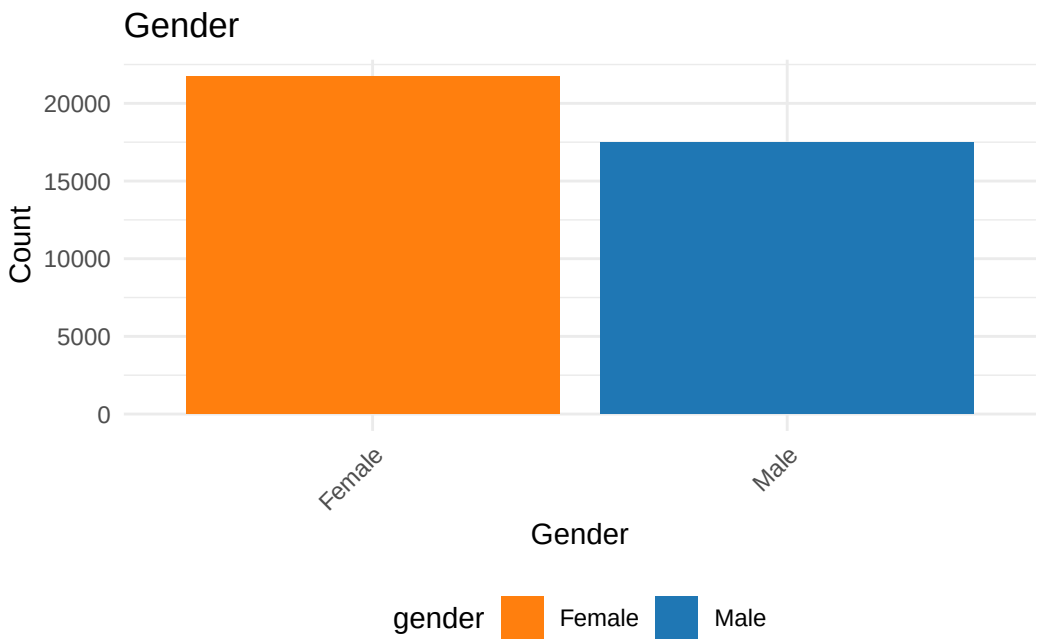
2.4 Summary Statistics and Graphical Representation

The summary data from the Cooperative Election Study (CES) 2020 dataset provides an overview of the voter demographics and their economic perspectives.

2.4.1 Gender Representation

In the dataset there are 24,303 female participants making up around 55.8% of the sample and 19,251 male participants accounting for approximately 44.2%. These numbers reflect trends showing that women tend to be politically engaged and active in voting during recent elections.

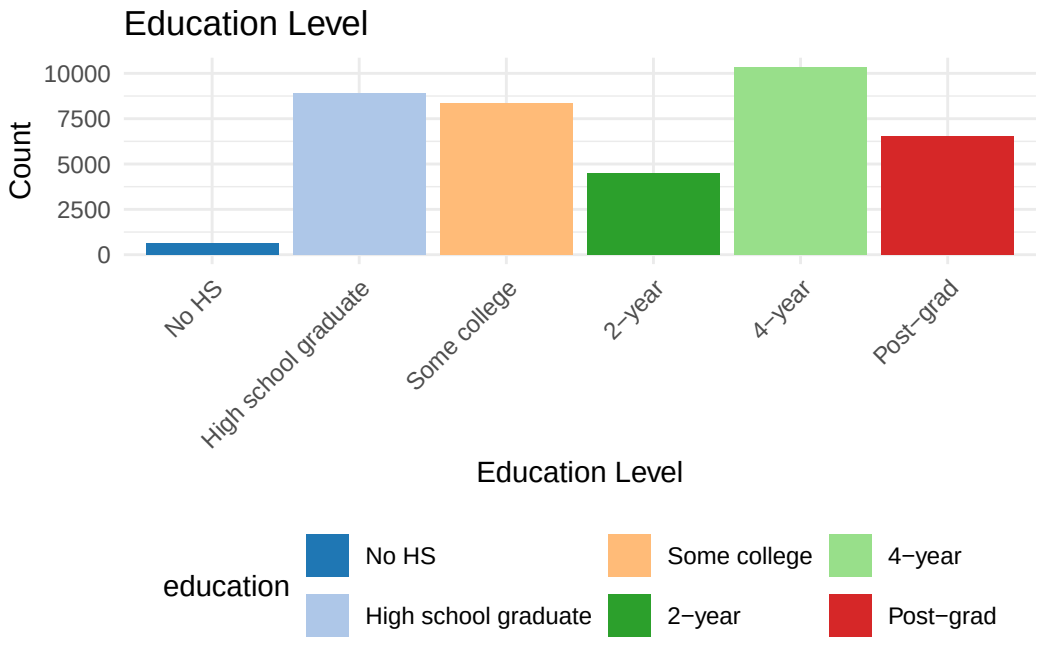
Category	n	Percent
Female	21722	55.40902
Male	17481	44.59098
Total	39203	100.00000



2.4.2 Educational Background

The educational profile of respondents is varied with the majority holding a '4 year' college degree (26.4%) followed by high school graduates (22.5%) and individuals with some college education (21.3%). Those with graduate degrees make up 16.7% of the sample while '2 year' college graduates represent 11.4%. Participants without a high school diploma constitute a proportion at 1.6%. This distribution highlights the significance of considering education levels in analysis as they can influence awareness, participation, in civic activities and policy preferences.

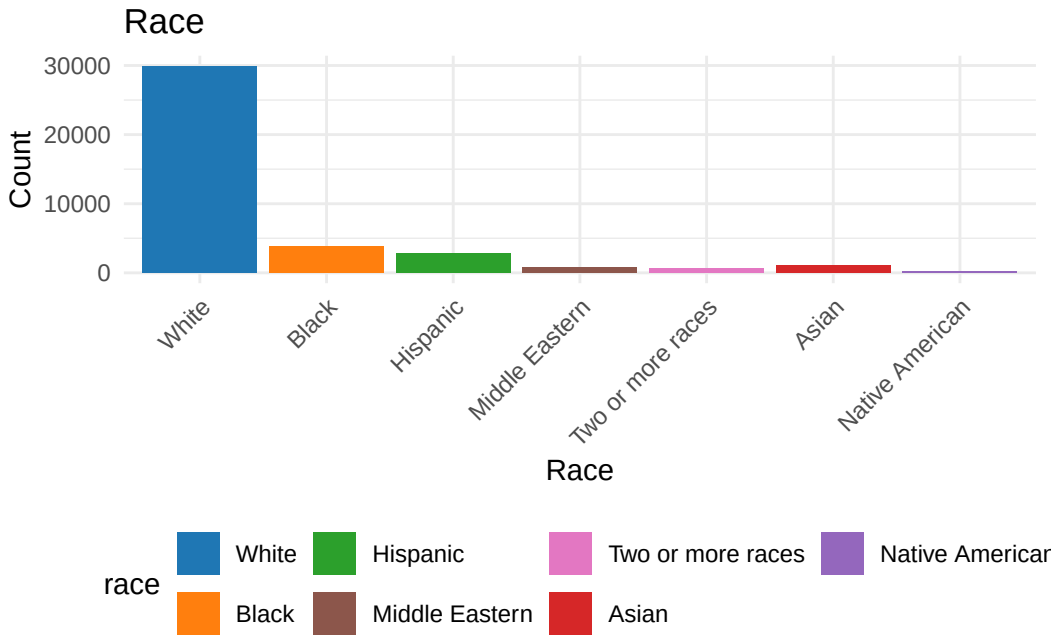
Category	n	Percent
No HS	621	1.584062
High school graduate	8894	22.687039
Some college	8326	21.238170
2-year	4500	11.478713
4-year	10343	26.383185
Post-grad	6519	16.628829
Total	39203	100.000000



2.4.3 Race

The majority of survey participants are White (76.3%) followed by African individuals at 9.5% and Hispanic or Latino respondents at 7%. A smaller portion of the group includes individuals (2.6%) those of 'Two or more races' (1.7%) Eastern individuals (1.7%) and Native Americans (0.7%). The presence of various ethnic backgrounds among voters highlights the importance of considering how race, policy issues and voting patterns intersect.

Category	n	Percent
White	29916	76.310486
Black	3797	9.685483
Hispanic	2840	7.244344
Middle Eastern	752	1.918221
Two or more races	599	1.527944
Asian	1013	2.583986
Native American	286	0.729536
Total	39203	100.000000



2.4.4 National Economic Sentiment

When it comes to perceptions of the economy a large percentage of respondents feel that the economy has either "Gotten worse" (41.4%) or "Gotten somewhat worse" (27.3%). Those who believe it has "Stayed about the same" make up 8.9% of participants while a smaller percentage think it has "Gotten somewhat better" (10.2%) or Gotten much better" (8.2%). A small fraction is uncertain about the outlook with 3.9% stating they are "Not sure". These findings suggest a view towards national economic conditions, during the election year which could impact voter decisions.

Category	n	Percent
Gotten much better	3410	8.698314
Gotten somewhat better	3925	10.011989
Stayed about the same	3166	8.075913
Gotten somewhat worse	10492	26.763258
Gotten much worse	17413	44.417519
Not sure	797	2.033008
Total	39203	100.000000

2.4.5 Change in Household Income

People's experiences with changes in household income over the years vary widely. 31.2% mentioned that their income remained relatively stable while 18.3% saw a decrease and 10% noticed a slight increase. A smaller percentage, 10% faced a decrease while only 3.7% reported a substantial increase. These individual financial shifts could have effects on voting decisions in an election marked by economic uncertainty. These summarized figures provide a glimpse into the economic landscape of the dataset offering context, for understanding the outcomes of the 2020 election. The forthcoming analysis will explore how these elements might have influenced voting behavior in detail.

Category	n	Percent
Increased a lot	1341	3.420657
Increased somewhat	6464	16.488534
Stayed about the same	21137	53.916792
Decreased somewhat	6829	17.419585
Decreased a lot	3432	8.754432
Total	39203	100.000000

2.5 Measurement

Measuring variables in a dataset is crucial for ensuring the reliability and accuracy of research results. In the Cooperative Election Study (CES) 2020 important concepts like outlook, personal income changes, gender, education level and race were defined through crafted survey questions. These questions aimed to capture the characteristics and perspectives of the electorate.

2.5.1 Defining Key Terms

Economic Outlook (National Economics. CC20_302): Survey participants were asked to evaluate whether they believed the national economy had improved or worsened in the year. Their responses were rated on a six point scale from "Gotten much better" to "Gotten much worse" with an option for "Not sure." This measure reflects perceptions of conditions rather than objective economic data.

Household Income Changes (CC20_303): This factor assesses how respondents view changes in their household incomes over the year. Using a five point scale from "Increased a lot" to "Decreased a lot" it captures both negative shifts in situations. This measure offers insights into voters' direct experiences.

Gender: Participants identified themselves as either “Male” or “Female” in a manner. Understanding gender in this way is crucial for delving into how gender dynamics impact preferences and voting behavior.

Education Level: The educational background of respondents was divided into six categories ranging from “No HS” to “Post grad” enabling an analysis of the relationship between education and various aspects such as beliefs, economic viewpoints and candidate preferences.

Race and Ethnicity: Respondents self-reported their race and ethnicity which were then categorized into groups like “White” “Black,” “Hispanic,” “ ” “Native American” “Middle Eastern” and those identifying with multiple races. This categorization is essential for exploring the influence of ethnic diversity on voting trends.

2.5.2 Considerations for Measurement

The subjective nature of variables can lead to response biases since individual perceptions are shaped by experiences, media exposure and political affiliations. However, relying on self reported data offers insights into the electorate’s mindset that objective data alone may not fully capture. The way education and race were measured in the study followed standards to ensure consistency with other research and data sets. However it’s important to recognize that there can be experiences, within these broad categories, which calls for a thoughtful analysis of the findings.

In general, the methods used to measure variables in the CES 2020 dataset allow for an exploration of voter demographics. While understanding the limitations of self reported data the datasets structure supports an examination of how factors economic views and voting patterns intersect. The thorough measurement approach sets a foundation for analyses offering a dependable framework for uncovering meaningful insights, into the dynamics of the 2020 Presidential election.

3 Model

3.1 Model set-up

In this section, we frame the estimand in terms of a logistic regression model and subsequently use the model to find relationships between the response variable and input variables. We first specify the model and then justify its appropriateness to conduct the analysis.

3.1.1 Model Specifications

The model used in this analysis is a multiple logistic regression. The variables along with their distributions are given as below:

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{education}_i + \beta_3 \times \text{race}_i && (2)\\
&\quad + \beta_4 \times \text{economic outlook}_i + \beta_5 \times \text{income change}_i && (3)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (5)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (6)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (7)\\
\beta_4 &\sim \text{Normal}(0, 2.5) && (8)\\
\beta_5 &\sim \text{Normal}(0, 2.5) && (9)
\end{aligned}
$$

where:

1. y_i takes the value 1 if the respondent voted Biden and 0 if they voted Trump
2. π_i is the conditional probability of voting for Biden given respondent i
3. $\text{Bern}(\pi_i)$ is the Bernoulli Distribution with parameter π_i
4. $logit(x)$ is the logit function that maps $(0, 1)$ to $(-\infty, \infty)$
5. gender_i is the i^{th} respondents gender which takes the value 1 for male and 0 for female
6. education_i is the i^{th} respondents education level which takes values from 1 to 6 depending on the respondents education level
7. race_i is the i^{th} respondents race which takes values from 1 to 7 depending on the respondents race
8. $\text{economic outlook}_i$ is the i^{th} respondents view on the economy which takes values on a scale from 1 to 6 depending on the respondents view of the economy
9. income change_i is the i^{th} respondents change in income which takes values on a scale from 1 to 5 depending on if the respondents income changed for the worse or better
10. β_0 is the intercept of the logistic regression equation
11. β_1 is the coefficient of the gender variable
12. β_2 is the coefficient of the education variable
13. β_3 is the coefficient of the race variable
14. β_4 is the coefficient of the economic outlook variable
15. β_5 is the coefficient of the income change variable

The above model will be used later in the results section to find the impact of the input variables on voter preferences. For now, we justify certain characteristics of the model that capture certain real life aspects of the analysis.

3.1.2 Model Justification

Our goal is to find the effect that certain voter specific characteristics and non-partisan views have on a voter voting for Joe Biden. For that, we consider the framework of the model by considering the response variable, the input variables, and how they come together within the model structure.

3.1.2.1 Response Variable

Our variable of interest is voter preferences for Joe Biden. We assume that voting preference for Biden is a random variable and has a Bern(π_i) distribution with π_i being the 'likelihood of voting for Biden'. We choose specifically the random variable characterization because voting outcomes aren't deterministic but rather phenomenon subject to chance. This also aligns with the voting patterns we observe in the real world.

We estimate the 'likelihood of voting for Biden' by the using the conditional probability of voting for Biden given a certain respondent.

3.1.2.2 Input Variables

The voter specific characteristics that we consider are:

1. Gender
2. Race
3. Education
4. Change in Income

The non-partisan view of the respondents that we consider is:

1. Economic Outlook

We assume for this analysis that all 5 variables have Normal prior distributions as based of our prior information, we can't conclude any specific form for any of the variables' distributions. All we can conclude is that the extreme values taken by each variable are less likely than their intermediate values and hence a suitable distribution to describe flat tails with a peak in the center is the normal distribution.

3.1.2.3 Model Structure

We construct a model that gives the conditional probability of voting for Biden from only the above 5 input variables. Probability is a measure taking values from $[0, 1]$ which is why we use the logit function to transform the conditional probability to the unbounded real line. The logit function is appropriate in this context because it preserves the ordering of probabilities while expanding them to the entire real line.

Once we have the $logit(\pi_i)$, we define a linear relationship between this quantity and the 5 categorical variables. Taking the inverse of the logit function gives probability as a function of the 5 categorical variables. We then use this model to study the impact of the 5 categorical variables on voting for Biden.

$$\begin{aligned} \text{logit}(\pi_i) = \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{education}_i + \beta_3 \times \text{race}_i \quad (10) \\ + \beta_4 \times \text{economic outlook}_i + \beta_5 \times \text{income change}_i \quad (11) \\ \pi_i = \text{logit}^{-1}(\beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{education}_i + \beta_3 \times \text{race}_i \quad (12) \\ + \beta_4 \times \text{economic outlook}_i + \beta_5 \times \text{income change}_i) \quad (13) \end{aligned}$$

3.1.2.4 Parameter Estimation

The model has 6 parameters β_0, β_1, β_2, β_3, β_4, β_5 which need to be estimated from the cleaned_data dataset. These parameters are estimated by defining the model within the stan_glm() function from the (Goodrich et al. 2023) package. The function uses Markov Chain Monte Carlo (MCMC) algorithms to create a chain of samples from the posterior distribution of each coefficient. This not only gives us point estimates, but also the posterior distribution of values for each coefficient.

The posterior distribution contains information about all of the coefficients moments, giving us estimates for not only the central values, but also the uncertainty associated with each estimate. Using the central moments as well as the 95% credibility intervals for the coefficients, helps us determine whether the input variables have an effect on voting preferences.

4 Results

This section looks at the results of the regression performed using the Model from the Model section and the cleaned data set. We first look at the results of our regression analysis, which includes, the parameter estimates and their credibility intervals. Subsequently, we explore select variables of interest and their effects on the probability of voting for Biden. Finally, we assess model fit using the R^2 metric and compare our final model's R^2 to the same model with certain terms dropped in order to assess the explanatory power of our model.

4.1 Regression Results

Table 6 contains the parameter estimates for the regression analysis. The regression analysis was done on a random sample of 10,000 observations out of 39,203 in order to speed up computational time while preserving the properties of the dataset. The first column of the Table 6 contains the parameter name. Since all the input variables are categorical variables, the regression creates indicator terms for each value the variable takes and estimates a unique coefficient based of that.

The second column contains the actual parameter point estimates. These estimates are based of the stan_glm() function from the (Goodrich et al. 2023) package. The third column gives the standard error while the 4th and 5th columns give the lower and upper limit for the credibility intervals. The table was created from the model using the tidy() function from the (Bolker and Robinson 2023) package, while the 'kable()' and 'kable_styling()' functions were used from (Xie 2023) and (Zhu 2023).

Figure 1 creates a plot of Table 6 with the estimate and standard errors shown. The plot shows the credibility intervals for each parameter.

Table 6: Parameter Estimates of the Multiple Logistic Regression

Parameter	Estimate	Standard Error	Credibility Interval Lower Limit	Credibility Interval Upper Limit
(Intercept)	-3.99	0.31	-4.51	-3.49
genderMale	-0.35	0.06	-0.44	-0.25
educationHigh school graduate	0.22	0.23	-0.15	0.61
educationSome college	0.59	0.23	0.22	0.98
education2-year	0.53	0.24	0.15	0.92
education4-year	1.01	0.23	0.64	1.41
educationPost-grad	1.17	0.23	0.79	1.56
raceBlack	2.96	0.17	2.70	3.24
raceHispanic	0.76	0.11	0.57	0.94
raceMiddle Eastern	0.78	0.23	0.41	1.16
raceTwo or more races	-0.15	0.23	-0.53	0.25
raceAsian	0.95	0.18	0.65	1.26
raceNative American	-0.26	0.31	-0.75	0.23
economic_outlookGotten somewhat better	0.94	0.20	0.62	1.27
economic_outlookStayed about the same	2.30	0.19	2.00	2.61
economic_outlookGotten somewhat worse	3.20	0.18	2.91	3.50

Parameter	Estimate	Standard Error	Credibility Interval Lower Limit	Credibility Interval Upper Limit
economic_outlookGotten much worse	5.11	0.18	4.82	5.42
economic_outlookNot sure	2.64	0.23	2.27	3.02
income_changeIncreased somewhat	-0.01	0.19	-0.34	0.29
income_changeStayed about the same	0.30	0.18	0.01	0.60
income_changeDecreased somewhat	0.50	0.19	0.19	0.80
income_changeDecreased a lot	0.47	0.21	0.12	0.81

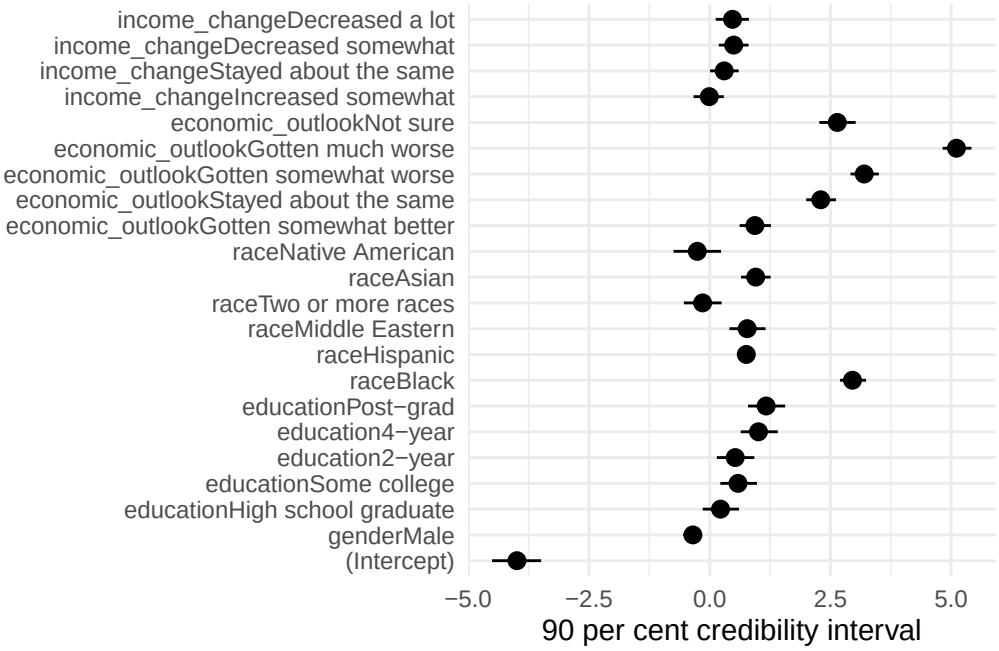


Figure 1: Parameter Estimates and Confidence Intervals

From the table and the plot, we observe that the gender variable has a negative coefficient estimate, indicating that given the respondent is male holding all else fixed, they have a lower probability of voting Biden compared to Trump. The coefficient estimates for education are all positive indicating that a respondent with a high school degree or higher has a higher probability of voting for Biden over Trump. The coefficient estimates for Race are positive for all

categories except 'two or more races' and 'native american' indicating that given a respondent is either 'Black', 'Asian', 'Middle Eastern' or 'Hispanic', they have a higher probability of favoring Biden.

The economic variables which are 'economic outlook' and 'income change' have distinct patterns compared to the demographic variables which are 'Gender', 'Education', and 'Race'. For one, the economic outlook coefficients, given in Table 6 are all positive but much larger in magnitude than any of the other variable coefficients. This indicates that given a respondent who's outlook on the economy is between neutral and poor, we find that the respondent has a very high probability of favoring Biden over Trump.

The same holds for change in income as most of the income coefficients are positive but negligible. This indicates that the respondents preferences based of income change are not very significant.

4.1.1 Model Fit

We now assess the fit of the model to the data. The statistic that validates model fit in this situation is R^2. We compare our models R^2 with the R^2 of the model:

1. Taking only demographic variables (Gender, Race, Education)
2. Taking only economic variables (Economic Outlook, Income Change)

These 2 different variants of the model are taken to show the improvement that our final model makes over both of them. We consider R^2 as a measure of model fit as it's a measure of the variation within the data explained by the model. The R^2 of all 3 models are given in Table 7. The first column lists the parameters along with different statistical measures, including R^2. The second column gives the gives the coefficient estimates and R^2 of the main model. The third column gives the coefficient estimates and R^2 of the model taking only demographic variables as the input variables. The third column gives the same but for economic variables.

Table 7 shows that the R^2 for the main model is 0.458 while the R^2 for the demographic only model and economic only model is 0.104 and 0.383 respectively. This shows that the demographic only model's fit is significantly less than our final model's, showing that considering economic variables in the analysis significantly boosted model fit.

When we look at the model with only economic variables, we find that a good amount of variation in model is explained by these 2 variables, evidenced by the comparatively high R^2 of 0.383. The model though lacks some interpretability and can be improved by considering the demographic variables on top. This leads us back to our main model, with the highest R^2 between the 3.

Table 7: Models Coefficients and R-Squared Estimates

	Main Model	Economic Only	Demographic Only
(Intercept)	−3.992	0.185	−2.837
genderMale	−0.348	−0.424	
educationHigh school graduate	0.224	−0.307	
educationSome college	0.586	−0.075	
education2-year	0.529	−0.066	
education4-year	1.010	0.507	
educationPost-grad	1.168	0.881	
raceBlack	2.957	2.413	
raceHispanic	0.758	0.597	
raceMiddle Eastern	0.776	0.765	
raceTwo or more races	−0.147	0.182	
raceAsian	0.953	0.707	
raceNative American	−0.257	−0.854	
economic_outlookGotten somewhat better	0.936		0.424
economic_outlookStayed about the same	2.300		2.120
economic_outlookGotten somewhat worse	3.200		2.945
economic_outlookGotten much worse	5.110		4.790
economic_outlookNot sure	2.643		2.884
income_changeIncreased somewhat	−0.009		−0.239
income_changeStayed about the same	0.300		0.032
income_changeDecreased somewhat	0.496		0.347
income_changeDecreased a lot	0.472		0.133
Num.Obs.	10 000	2000	2000
R2	0.458	0.104	0.383
Log.Lik.	−4068.493	−1237.760	−922.412
ELPD	−4090.8	−1251.0	−933.0
ELPD s.e.	57.2	16.0	25.3
LOOIC	8181.6	2502.0	1865.9
LOOIC s.e.	114.4	31.9	50.5
WAIC	8181.5	2501.9	1865.9
RMSE	0.36	0.46	0.39

5 Discussion

5.1 Demographic Characteristics Effect on Voter Preference

In this paper we examined how race, education, gender, economic outlook and change in income impacted voters' preferences for Biden over Trump in the 2020 US presidential election. Our analysis showed that Biden voters thought the nation's economy got much worse. Furthermore, voters that supported Biden were female, Black, Hispanic, and Asian. They possessed a minimum of a four-year college degree or higher. Finally, they were white and non-white college graduates as well as non-white without a four-year college degree. Overall, Biden voters were women, more racially diverse with greater education attainment.

Our data and other research we conducted on education as well as education by race was consistent. There were a few discrepancies in the data we analyzed and previous research with regards to gender and race.

First, voters' gender on voting preferences were contradictory. Our data demonstrated a greater percent of women voted for Biden whereas (Igielnik 2021) points out that Biden received more votes from men. A plausible explanation for this difference is a small difference in gender preferences. Further research should be done as well as observing previous election results to determine whether voters' gender really is a reliable predictor of voting preferences in US elections.

Secondly, our data on race was congruent with Whites, Blacks and Asians. Our data indicated Hispanics voted for Biden however other research suggests Hispanics were more inclined to vote for Trump. Another incongruence involves the 'other' race category in that our data signifies the other group voted Trump however research alludes this group as being Biden supporters. A possible reason for this inconsistency is that in our data, the other category consists of only three percent of electorates and Hispanics ten percent. Having such a small representation of these groups increases the likelihood of discrepancies in results.

5.2 Economic Characteristics Effect on Voter Preference

Perhaps the strongest effect on voting preference, was through the economic outlook variable. The coefficients were on orders of magnitude 2-3 times larger than the coefficients for the other variables. This indicates the prevalence of a poor economy sentiment among majority of the voter base. This also shows that the prevelance of a weak economy during the pandemic influenced the political preferences of voters and shifted a large amount of the populace in favor of Biden.

This, coupled with the fact that a majority of voters who saw their incomes reduce favored Biden, alludes to the possibility that a weak economy during election season prompted an anti-incumbency sentiment amongst many voters, influencing them to vote for Biden.

The effect of both these variables on voting preference for Biden is interesting because, for one, they had the strongest combined effect, and also are characteristics which aren't easily predictable based of a voters demographics. Also, when we ran our regression with only these 2 variables, they had better predictive power than the model with only the demographic variables involved. Therefore, economic characteristics may be the key to understanding voter preferences, and combined with demographic characteristics, tell a compelling story of voter outcomes.

5.3 Weaknesses

A major weakness of our analysis is assuming a linear model for voting preferences as a function of demographic and economic characteristics. Assuming linearity hids a lot about the mechanism through which these characteristics influence voting outcomes. Though model interpretability is increased, we lose any sound understanding of the mechanism underlying the phenomenon that we aim to model.

5.4 Next steps

Despite taking place during COVID-19, the 2020 election showed the highest turnout in the 21st century, where 66.8% of eligible voters voted. This is up from 60.1% in the previous election in 2016, most likely due to the addition of nontraditional methods of voting such as early voting and mailing ballots to ensure adequate voter turnout (Fabina 2022). A potential probe is looking at the demographics behind those that voted by mail, absentee, early or on the traditional day. For example, (Igielnik 2021) states Hispanic and whites voted by mail whereas blacks voted early in person.

A possible area of additional research could question the correlation between voters' race and the race of the candidate they vote for. For instance, investigating the different demographics that voted for Biden, a white male candidate compared to his running mate, Kamala Harris a black female candidate. Additionally, white respondents who showed high levels of radical resentment were more likely to vote for Joe Biden as opposed to Kamala Harris. Not only was Harris at a disadvantage because of her race but also because of her gender. (Nelson 2021) states that "hostile sexism" may play a role in Harris's success in the election (2021). The combination of these two variables could lead to interesting findings.

Additional research could look in the direction of age as a predictor in the US election. (Lees 2024), with the use of 1000 research participants through a multi-step design, finds voters aged 49 and younger were in support of Biden whereas the older population favored Trump, however, not by a significant amount. Another trend portrayed the insignificance of candidates age on voter preference. Young adults aged 30 and below indicated that voting for a candidate that is over the age of 70 did not change their voting decision. At the time of the election,

both Biden and trump were over the age of 70 (Lees 2024). Thus further research is necessary to determine the reliable relationship between age and voting preferences.

References

Arel-Bundock, Vincent. 2023. *Marginaleffects: Marginal Effects for Model Objects*. https://CRAN.R-project.org/package=marginaleffects.

Arel-Bundock, Vincent, and Antoine Pilon. 2023. *Modelsummary: Summary Tables and Plots for Statistical Models and Data*. https://CRAN.R-project.org/package=modelsummary.

Blume-Kohout, Margaret, Matthew Arch, Casey Stanton, Timothée Poisot, and Danny Lee. 2023. *Dataverse: Client for Dataverse Repositories*. https://CRAN.R-project.org/package=dataverse.

Bolker, Benjamin, and David Robinson. 2023. *Broom.mixed: Tidying Methods for Mixed Models*. https://CRAN.R-project.org/package=broom.mixed.

Canty, Angelo, and Brian Ripley. 2023. *Boot: Bootstrap Functions (r-Forge)*. https://CRAN.R-project.org/package=boot.

Fabina, Scherer, J. 2022. "U.s. Age and Voting Patterns in the 2020 Presidential Election." *Census Bureau. (Publication No. P20-585)*. https://www.census.gov/content/dam/Census/library/publications/2022/demo/p20-585.pdf.

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. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://CRAN.R-project.org/package=rstanarm.

Igielnik, Keeter, R. 2021. "Behind Biden's 2020 Victory." *Pew Research Center - U.S. Politics & Policy*. https://www.pewresearch.org/politics/2021/06/30/behind-bidens-2020-victory/.

Lees, Praino, C. 2024. "Young Voters, Older Candidates and Policy Preferences: Evidence from Two Experiments." *International Political Science Review*. https://doi.org/https://doi.org/10.1177/01925121221139544.

Lingl, Daniel. 2023. *Collapse: Advanced and Fast Data Transformation*. https://CRAN.R-project.org/package=collapse.

Müller, Kirill. 2023. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Neal, Romain et al. 2023. *Arrow: Integration to 'Apache Arrow'*. https://CRAN.R-project.org/package=arrow.

Nelson, K. 2021. "You Seem Like a Great Candidate, but…: Race and Gender Attitudes and the 2020 Democratic Primary." *The Journal of Race, Ethnicity, and Politics*. https://doi.org/https://doi.org/10.1017/rep.2020.53.

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

Wickham, Hadley, Lionel Henry, and Kirill Müller. 2023. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, and Jim Hester. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://CRAN.R-project.org/package=knitr.

Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.