A Multifaceted Examination of the 2020 United States Presidential Election*

Understanding and Comparing how Identity-Based, Socio-Economic, and Regional Variables Shaped the Outcome

Sima Shmuylovich

March 16, 2024

This study examines the factors that shaped support for the United States of America's Presidential Candidates, Joseph Biden and Donald J. Trump, during the 2020 federal election. Analyzing identity-based, socio-economic, and regional variables, we discovered that factors such as gender, age, race, education level, employment status, income, region, urban status, and state significantly influenced voter preferences. Our findings reveal that higher education levels and unemployment were positively associated with support for Biden, while older age groups, males, and certain racial demographics showed varied levels of support between the candidates. Regionally, urban areas and specific states like the District of Columbia showed stronger preferences for Biden, highlighting the geographical and socio-economic divides within the electorate. These insights offer a comprehensive understanding of the 2020 election dynamics, enabling more informed predictions for the upcoming 2024 election (for which many analysts predict the same candidates). This research contributes to a better grasp of American political affiliations, assisting citizens and policymakers in adapting to evolving electoral trends.

Table of contents

*Code and data are available at: https://github.com/sshmuylovich/us-election-2020.git

1 Introduction

The 2020 federal elections saw Joseph Biden winning 306 of 538 electoral votes, defeating Donald J. Trump to become the 46th President of the United States (Politics 2020). With the 47th federal election quickly approaching in 2024 and both Biden and Trump being the presumptive nominees for their respective political parties, this paper conducts an analysis on support for these two politicians, specifically examining identity-based, socio-economic, and regional variables. The estimand of interest in this research is the average causal effect of these variables on the likelihood of individuals expressing support for either Democratic Candidate Biden or Republican Candidate Trump.

By delving into identity-based (gender, age, and race), socio-economic (education level, employment status, and income), and regional factors (region, urban status, and state), this paper seeks to illuminate the multifaceted influences that shape political support via a case study in the 2020 US Election. The objective is to unravel the complex dynamics that dictate electoral outcomes, thereby providing a better understanding of the American political landscape and a better ability to forecast potential shifts in voter alignment.

The subsequent sections follow a structured format. Section 2 outlines the source and variables central to our analysis. Section 3 details the construction and methodology of the statistical models used. Section 4 presents the key findings of our analysis, while Section 5 critically reviews the content, addresses the implications of the results, acknowledges model limitations, and suggests potential research directions.

2 Data

The data used in this paper was gathered from the 2020 Cooperative Election Study (CES) hosted on the Harvard Dataverse (Stephen Ansolabehere and Luks 2021) and analyzed using R (R Core Team 2023) with help from `tidyverse` (Wickham et al. 2019), `rstanarm` (Goodrich et al. 2022), `modelsummary` (Arel-Bundock 2022), `testthat` (Wickham 2011), `here` (Müller 2020), `knitr` (Xie 2023), and `kableExtra` (Zhu 2021).

2.1 Cooperative Election Study 2020

The Cooperative Election Study (CES), previously known as the Cooperative Congressional Election Study, is an extensive research project aimed at understanding the intricacies of American electoral behavior (Stephen Ansolabehere and Luks 2021). The 2020 CES survey was conducted over the Internet by YouGov, a British international Internet-based market research and data analytics firm (Meltwater 2023). Participants were interviewed from September 29 to November 2, 2020 (for pre-election data), and from November 8 to December 14, 2020 (for post-election data). A total of 61000 participants were interviewed for pre-election and 51551 returned for post-election interviews. The dataset includes 717 variables, many of which could have been included in the analysis but was narrowed down to 9 variables that could correlate with candidate support. These variables are gathered from the CPS portion of the survey with no open-ended answers and assigned numerical values with labels.The data was released on March 26, 2021 (Stephen Ansolabehere and Luks 2021).

2.2 Identity-Based Variables

Gender was picked as one of the identity-based variables because gender can influence an individual's policy preferences. For instance, women might prioritize issues such as healthcare, reproductive rights, and gender equality more than men. Understanding gender differences in voting behavior can help in analyzing how these issues impact elections. A visualization of the proportion of votes by gender in the 2020 election can be found here, Figure 6.

Age was picked as one of the identity-based variables because different age groups often hold distinct values and perspectives shaped by their generational experiences. For instance, younger voters may prioritize climate change, education, and jobs, while older voters may focus on healthcare, social security, and national security. A visualization of the proportion of votes by age in the 2020 election can be found here, Figure 1.

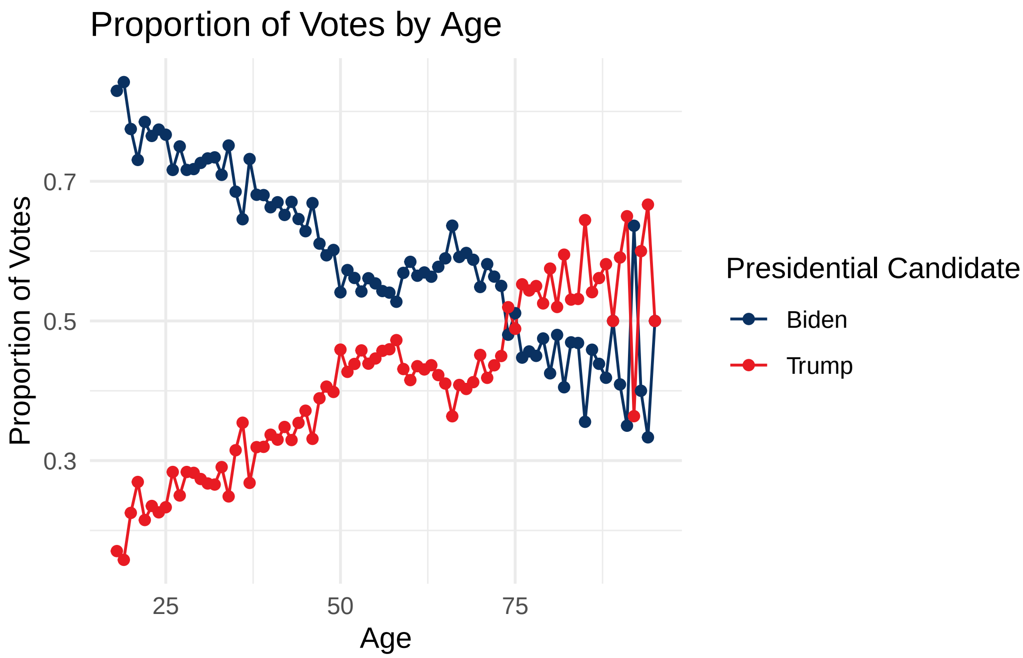

Figure 1: Proportion of Votes by Age in the 2020 Election

Race was picked as one of the identity-based variables because different racial and ethnic groups have unique historical and social experiences that influence their political views and behaviors. For example, policies on immigration, law enforcement, and affirmative action may be viewed differently by voters of different racial backgrounds. The Pew Research Center found that in the 2022 midterms, Black voters supported Democrats by overwhelming margins: 93% voted for Democrats while only 5% supported Republicans, similar to levels of support in 2020, 2018 and 2016 (Center 2023). A visualization of the proportion of votes by race in the 2020 election can be found here, Figure 7.

2.3 Socio-Economic Variables

Education was picked as one of the socio-economic variables because education level is closely linked to an individual's policy preferences and political awareness. Higher education levels often correlate with more liberal attitudes on social issues and a greater engagement in political processes. A visualization of the proportion of votes by education in the 2020 election can be found here, Figure 8.

Employment status was picked as one of the socio-economic variables because employment status directly affects an individual's economic security and outlook, influencing their priorities at the polls. For example, unemployed or underemployed voters might prioritize job creation, economic recovery, and social safety nets more highly than those securely employed. A visualization of the proportion of votes by employment status in the 2020 election can be found here, Figure 9.

Income was picked as one of the socio-economic variables because income levels influence voters' economic interests, with higher-income individuals potentially prioritizing tax policies and economic strategies that favor wealth preservation, while lower-income voters might focus on income redistribution, minimum wage increases, and access to affordable healthcare. A visualization of the proportion of votes by income in the 2020 election can be found here, Figure 10.

2.4 Regional Variables

Region was picked as one of the regional variables because different regions in a country often have unique cultural and historical contexts that shape residents' values, attitudes, and political leanings. For example, historical voting patterns, regional industries, and local issues can significantly influence regional voting behaviors. Furthermore, specific issues may be more pressing in certain regions than others, such as environmental concerns in areas prone to climate-related disasters or economic policies in regions dominated by particular industries. A visualization of the proportion of votes by region in the 2020 election can be found here, Figure 11.

Urban status was picked as one of the regional variables because urban, suburban, and rural areas differ markedly in their socio-economic and demographic compositions. These differences can lead to distinct political priorities and voting behaviors, with urban areas often leaning more towards progressive policies and rural areas favoring conservative stances, influenced by factors like population density, diversity, and economic opportunities. The concentration of services and infrastructure in urban areas, as opposed to their scarcity in rural regions, can influence voter concerns and priorities, such as public transportation, education, and healthcare services. A visualization of the proportion of votes by urban status in the 2020 election can be found here, Figure 12.

State was picked as one of the regional variables because state-level policies and governance can significantly affect residents' lives, influencing their political preferences. Issues such as education, healthcare, taxation, and environmental regulation can vary widely by state, affecting voting behavior. Some states also have a higher electoral significance due to their size, demographic composition, or status as swing states. Understanding the political dynamics at the state level is crucial for predicting and analyzing election outcomes, especially in systems like the United States' Electoral College. A visualization of the proportion of votes by state in the 2020 election can be found here, Figure 2.

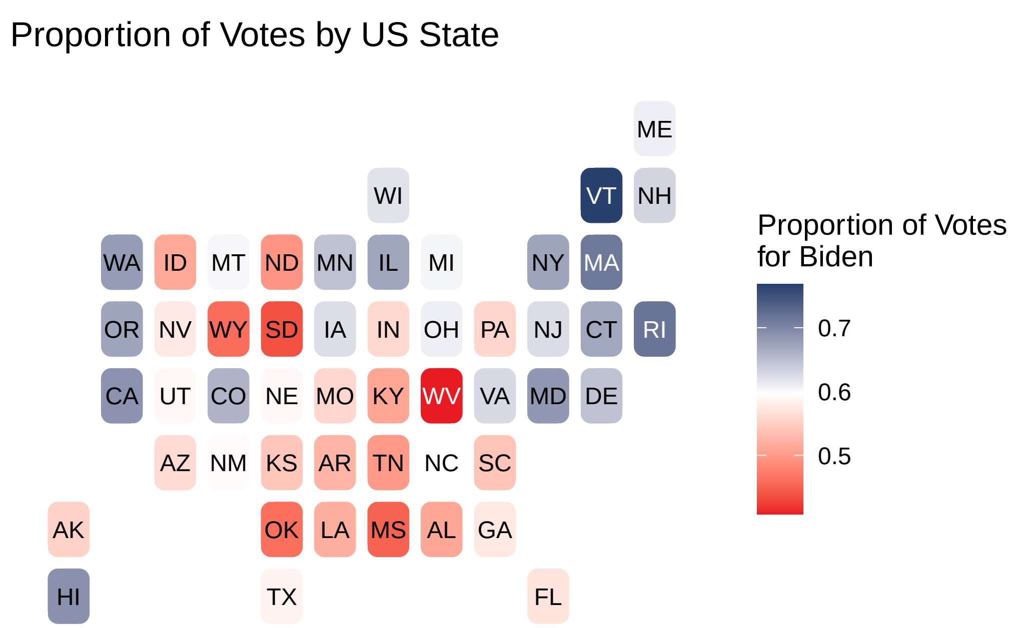

Figure 2: Proportion of Votes by State in the 2020 Election

2.5 Measurement

Age was calculated in years based on the year of birth the respondent inputted. Gender and Race were given a number with a corresponding label based on which radio button the respondent selected. When cleaning the data the following changes were made to the raw data: age was categorized as seen in Table 7, gender was limited to Male and Female, and race was mapped from a number to the categories seen in Table 8.

A preview of the identity-based variables used in this paper can be seen in Table 1.

Table 1: Sample of Identity-Based Data

gender	age_bucket	race
Male	45-64	White
Female	45-64	White
Male	45-64	White
Female	65+	White
Female	45-64	White

Education, employment status, and income were all given a number with a corresponding label based on which radio button the respondent selected. When cleaning the data the following changes were made to the raw data: education was mapped from a number to the categories seen in Table 9, employment status was mapped from a number to the categories seen in Table 10, and income was mapped from a number to the categories seen in Table 11.

A preview of the socio-economic variables used in this paper can be seen in Table 2.

Table 2: Sample of Socio-Economic Data

education	employment_status	income
Some college or assoc. degree	Not in the Workforce	Less than 30,000
College graduate	Not in the Workforce	100,000 - 199,999
Some college or assoc. degree	Employed	30,000 - 49,999
Some college or assoc. degree	Not in the Workforce	Less than 30,000
High school or less	Employed	50,000 - 99,999

Region, urban status, and state were all given a number with a corresponding label based on which radio button the respondent selected. When cleaning the data the following changes were made to the raw data: region was mapped to the categories seen in the CES survey (Northeast, Midwest, South, West), urban status was mapped from a number to the categories seen in the CES survey (City, Suburb, Town, Rural Area), and state was mapped from a number to the categories seen in the CES survey (all 50 states and the District of Columbia).

A preview of the regional variables used in this paper can be seen in Table 3.

Table 3: Sample of Regional Data

region	urban_status	state
Northeast	Suburb	Connecticut
Northeast	Rural Area	Massachusetts
Midwest	Suburb	Ohio
Midwest	Rural Area	South Dakota
Midwest	Rural Area	Ohio

It is important to note that this study is only interested in participants who were registered to vote in the 2020 election and who voted for either Biden or Trump, this was reflected in the data cleaning process.

3 Model

For a comprehensive analysis of how identity-based, socio-economic, and regional factors influenced voting behavior in the 2020 US Election, a generalized linear model (GLM) is an effective statistical approach. Given the nature of the dependent variable which is binary (voted for Joe Biden vs. Donald Trump), logistic regression is a suitable model within the GLM framework. This model predicts the log-odds of the outcome as a linear combination of the independent variables.

3.1 Identity-Based Model Setup

Define y_i as who the respondent voted for and equal to 1 if Joe Biden and 0 if Donald Trump. Then gender, age, and race are the respective answers of the respondent.

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{age}_i + \beta_3 \times \text{race}_i && (2)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (6)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`. The use of Normal(0, 2.5) priors is a conservative choice that imposes minimal prior beliefs on the magnitude of the coefficients. It is used in

Bayesian logistic regression to reflect a lack of strong prior knowledge while still allowing the data to inform the final posterior distributions.

3.2 Socio-Economic Model Setup

Define y_j as who the respondent voted for and equal to 1 if Joe Biden and 0 if Donald Trump. Then education, employment status, and income are the respective answers of the respondent.

$$
\begin{aligned}
y_j|\pi_j &\sim \text{Bern}(\pi_j) && (7)\\
\text{logit}(\pi_j) &= \beta_0 + \beta_1 \times \text{education}_j + \beta_2 \times \text{employment status}_j + \beta_3 \times \text{income}_j && (8)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (9)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (10)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (11)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (12)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`. The use of Normal(0, 2.5) priors is a conservative choice that imposes minimal prior beliefs on the magnitude of the coefficients. It is used in Bayesian logistic regression to reflect a lack of strong prior knowledge while still allowing the data to inform the final posterior distributions.

3.3 Regional Model Setup

Define y_k as who the respondent voted for and equal to 1 if Joe Biden and 0 if Donald Trump. Then region, urban status, and state are the respective answers of the respondent.

$$
\begin{aligned}
y_k|\pi_k &\sim \text{Bern}(\pi_k) && (13)\\
\text{logit}(\pi_k) &= \beta_0 + \beta_1 \times \text{region}_k + \beta_2 \times \text{urban status}_k + \beta_3 \times \text{state}_k && (14)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (15)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (16)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (17)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (18)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`. The use of Normal(0, 2.5) priors is a conservative choice that imposes minimal prior beliefs on the magnitude of the coefficients. It is used in Bayesian logistic regression to reflect a lack of strong prior knowledge while still allowing the data to inform the final posterior distributions.

4 Results

4.1 Identity-Based Results

Table 4: Model of support for the 2020 Presidential Candidates based on gender, age, and race

term	estimate	std.error	conf.low	conf.high
(Intercept)	1.89	0.15	1.63	2.15
genderMale	-0.27	0.04	-0.34	-0.20
age_bucket30-44	-0.37	0.09	-0.52	-0.22
age_bucket45-64	-0.94	0.09	-1.08	-0.80
age_bucket65+	-1.03	0.09	-1.18	-0.88
raceBlack	1.70	0.19	1.39	2.02
raceHispanic	-0.24	0.16	-0.52	0.02
raceOther	-1.09	0.26	-1.53	-0.66
raceWhite	-0.77	0.14	-1.01	-0.54

This table presents the estimates for how gender, age, and race affected voter preferences. Key observations include: being male (genderMale) negatively impacted the likelihood of voting for Biden by -0.27, suggesting a gender divide; age groups 30-44, 45-64, and 65+ all had negative impacts on Biden support, with coefficients of -0.37, -0.94, and -1.03, respectively, indicating a trend where older voters were more likely to support Trump; Black voters were significantly more likely to support Biden (+1.70), while Hispanic, Other, and White races showed less likelihood (with coefficients of -0.24, -1.09, and -0.77, respectively) to vote for Biden over Trump.

This figure visualizes the coefficient estimates from Table 4, showing how each identity-based factor (gender, age, and race) influences the odds of voting for Biden versus Trump. The negative coefficients for genders, age groups, and certain races would appear below the zero line, indicating a decrease in the likelihood of voting for Biden. The positive coefficient for raceBlack and the negative coefficients for other races indicate a significant racial divide. Gender also played a crucial role, with males being less likely to vote for Biden. Age differences further accentuated these divides, with older voters being more inclined towards Trump.

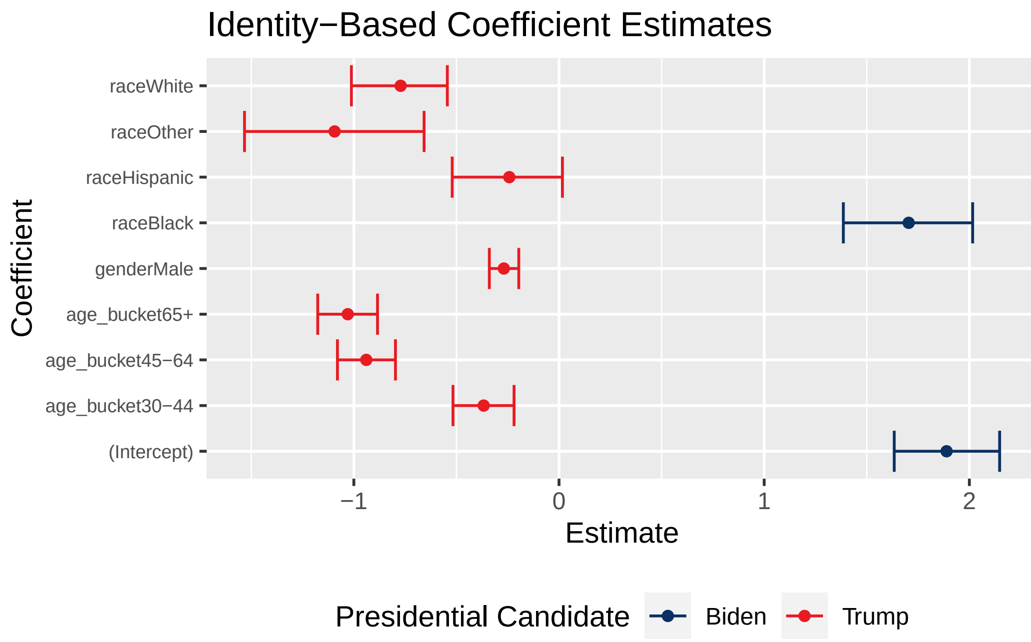

Figure 3: Graph of support for the 2020 Presidential Candidates based on gender, age, and race

4.2 Socio-Economic Results

Table 5: Model of support for the 2020 Presidential Candidates based on education, employment status, and income

term	estimate	std.error	conf.low	conf.high
(Intercept)	0.66	0.06	0.57	0.76
educationHigh school or less	-1.09	0.06	-1.20	-0.99
educationPostgraduate study	0.25	0.07	0.14	0.37
educationSome college or assoc. degree	-0.49	0.06	-0.58	-0.40
employment_statusNot in the Workforce	-0.09	0.05	-0.16	-0.01
employment_statusUnemployed	0.41	0.08	0.27	0.55
income200,000 or more	0.04	0.11	-0.13	0.23
income30,000 - 49,999	0.28	0.07	0.17	0.40
income50,000 - 99,999	0.11	0.06	0.01	0.21
incomeLess than 30,000	0.56	0.07	0.45	0.68

This table examines the effects of education, employment status, and income on voting preferences. Key observations include: higher education levels (e.g., postgraduate study +0.25) positively influenced Biden support, while lower education levels (high school or less -1.09) were detrimental; unemployment was associated with a higher likelihood of voting for Biden (+0.41), reflecting economic grievances; lower income brackets (<$30,000 +0.56) showed stronger support for Biden, suggesting economic factors played a significant role in voting behavior.

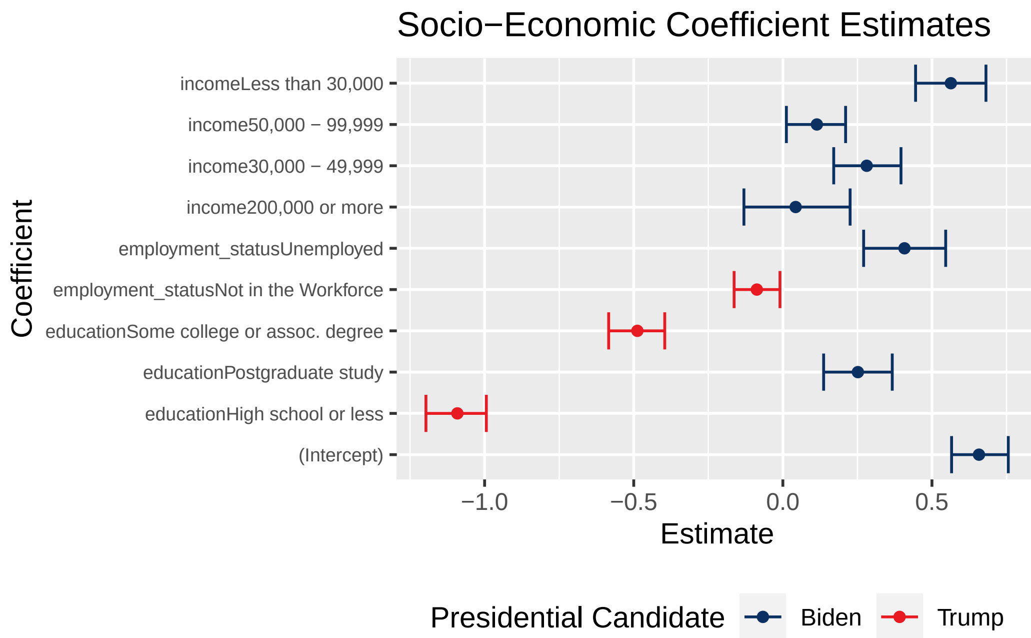

Figure 4: Graph of support for the 2020 Presidential Candidates based on education, employment status, and income

This figure graphically represents the findings from Table 5, illustrating the varying impacts of socio-economic factors on voter preferences. The positive and negative coefficients indicate increased or decreased likelihoods of supporting Biden, visualizing the socio-economic divides within the electorate. Education emerged as a strong predictor, with higher education levels correlating with Biden support. Employment status and income levels further highlighted economic concerns among voters, with lower income brackets and unemployed individuals more likely to support Biden.

4.3 Regional Results

Table 6: Model of support for the 2020 Presidential Candidates based on region, urban status, and state

term	estimate	std.error	conf.low	conf.high
(Intercept)	1.11	2.96	-4.06	5.85
regionNortheast	0.04	4.46	-7.13	7.70
regionSouth	-0.41	2.97	-5.09	4.75
regionWest	0.05	4.50	-7.09	7.60
urban_statusRural Area	-1.36	0.07	-1.48	-1.25
urban_statusSuburb	-0.60	0.06	-0.70	-0.51
urban_statusTown	-0.85	0.07	-0.97	-0.73
stateAlaska	-0.65	4.40	-7.61	6.59
stateArizona	-0.21	4.31	-7.05	6.96
stateArkansas	0.12	0.29	-0.37	0.60
stateCalifornia	0.02	4.34	-6.84	7.18
stateColorado	-0.03	4.30	-6.86	7.15
stateConnecticut	0.28	4.47	-7.08	7.60
stateDelaware	0.44	0.34	-0.12	1.00
stateDistrict of Columbia	2.10	0.80	0.95	3.75
stateFlorida	0.25	0.19	-0.06	0.55
stateGeorgia	0.23	0.21	-0.11	0.57
stateHawaii	0.40	4.32	-6.65	7.65
stateIdaho	-0.39	4.31	-7.23	6.85
stateIllinois	0.14	2.98	-4.61	5.34
stateIndiana	-0.12	2.96	-4.86	5.03
stateIowa	0.18	3.03	-4.56	5.39
stateKansas	-0.80	2.99	-5.58	4.35
stateKentucky	-0.01	0.23	-0.39	0.36
stateLouisiana	0.09	0.26	-0.34	0.54
stateMaine	0.29	4.53	-6.95	7.60
stateMaryland	0.61	0.24	0.22	0.99
stateMassachusetts	0.42	4.51	-6.92	7.73
stateMichigan	0.17	2.96	-4.56	5.38
stateMinnesota	0.36	3.00	-4.37	5.56
stateMississippi	-0.01	0.31	-0.53	0.49
stateMissouri	-0.13	2.99	-4.84	4.98
stateMontana	0.31	4.32	-6.73	7.44
stateNebraska	0.14	3.03	-4.62	5.37
stateNevada	-0.37	4.31	-7.33	6.74
stateNew Hampshire	-0.11	4.50	-7.32	7.18

term	estimate	std.error	conf.low	conf.high
stateNew Jersey	0.05	4.50	-7.24	7.33
stateNew Mexico	-0.15	4.30	-7.07	7.05
stateNew York	0.21	4.52	-7.12	7.52
stateNorth Carolina	0.55	0.20	0.21	0.89
stateNorth Dakota	-0.23	3.09	-5.08	4.96
stateOhio	-0.03	3.03	-4.78	5.12
stateOklahoma	-0.28	0.30	-0.76	0.21
stateOregon	0.62	4.35	-6.17	7.81
statePennsylvania	-0.22	4.50	-7.51	7.06
stateRhode Island	0.10	4.54	-7.10	7.48
stateSouth Carolina	0.26	0.23	-0.12	0.62
stateSouth Dakota	0.03	3.02	-4.81	5.27
stateTennessee	-0.01	0.21	-0.37	0.33
stateTexas	0.28	0.19	-0.03	0.59
stateUtah	0.10	4.30	-6.86	7.33
stateVermont	1.32	4.60	-6.07	8.46
stateVirginia	0.37	0.21	0.03	0.71
stateWashington	0.27	4.33	-6.58	7.40
stateWest Virginia	0.01	0.30	-0.49	0.51
stateWisconsin	0.04	3.00	-4.68	5.18
stateWyoming	-0.45	4.34	-7.34	6.75

This table provides a detailed analysis of how region, urban status, and specific states influenced voting behavior. Key observations include: certain regions (e.g., the Northeast +0.04 and West +0.05) and urban statuses (Suburb -0.60) had distinct effects on voting preferences, with rural areas significantly less likely to support Biden (-1.36); the District of Columbia (+2.10) showed strong support for Biden, while rural states like West Virginia had negative coefficients, indicating a preference for Trump.

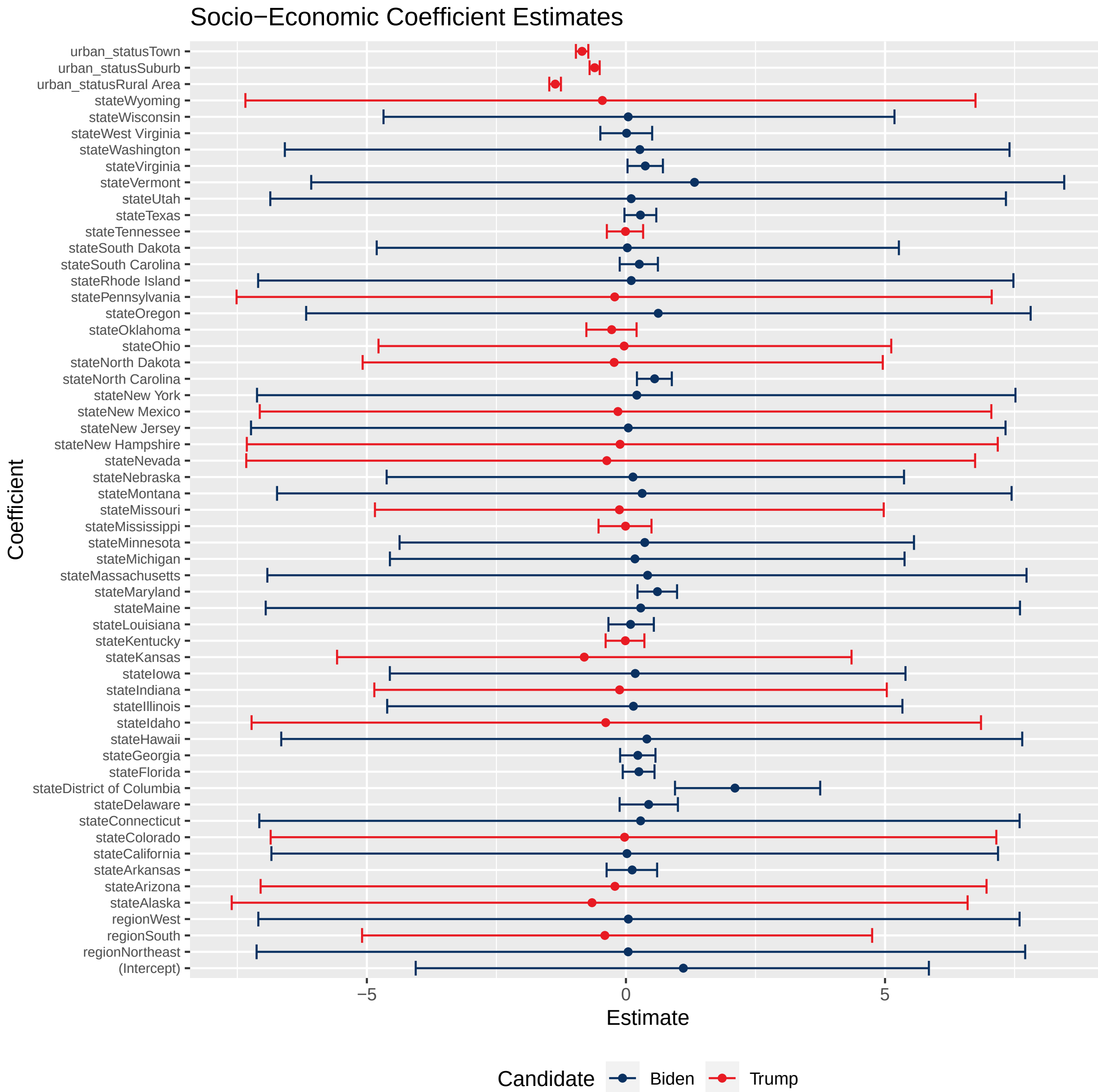

Figure 5: Graph of support for the 2020 Presidential Candidates based on region, urban status, and state

This figure graphically represents the findings from Table 6, showcasing the geographical patterns in voting behavior. Regional variables demonstrated the geographical polarization of voter preferences, with stark differences between urban and rural areas, as well as among different states and regions. The strong negative coefficient for urban_statusRural Area and significant state-specific variations (e.g., District of Columbia vs. West Virginia) illustrate the geographical and cultural divides.

5 Discussion

This study presents a comprehensive examination of the factors influencing voter preferences for Joseph Biden and Donald J. Trump in the 2020 US Presidential Election. By analyzing identity-based, socio-economic, and regional variables through a series of logistic regression models, the paper uncovers insights into the electorate's behavior. Using data from the Cooperative Election Study 2020, we employed a series of logistic regression models to examine the influence of gender, age, race, education level, employment status, income, geographical region, urbanization level, and state-specific factors on electoral outcomes. This approach allows for a detailed understanding of how various demographic, economic, and geographical factors contributed to the election's outcome.

5.1 Identity-Based Findings

One key finding is the significant role of identity-based variables, such as race and gender, in shaping voter preferences. The analysis demonstrates that Black voters were markedly more likely to support Biden, while white and male voters showed a higher propensity to support Trump. This highlights the enduring impact of racial divisions within the American political landscape, underscoring the need for policies that address these deep-seated disparities. Furthermore, the gender gap observed, with males showing a higher likelihood of supporting Trump, points to differing political priorities and perceptions between men and women. These insights into the electoral impact of race and gender highlight the ongoing challenges of addressing inequality and fostering unity within the political landscape.

Age emerged as another pivotal factor, revealing generational divides in political alignment. Younger voters' inclination towards Biden could reflect concerns about issues like climate change, social justice, and economic opportunity, areas where they may feel the Democratic platform offers more robust solutions. Conversely, older voters' support for Trump might be influenced by different priorities, such as economic policies, national security, and conservative social values. This generational split underscores the need for political strategies that bridge age-related divides, offering policies that cater to the diverse concerns and aspirations across age groups.

These identity-based insights into the 2020 election reveal the complex interplay between social identities and political preferences, highlighting the need for inclusive, responsive policies that address the concerns of a diverse electorate. Understanding the nuanced influences of race, gender, and age on voting behavior is crucial for developing political strategies and policies that foster unity, address systemic inequalities, and meet the varied needs of the American people.

5.2 Socio-Economic Findings

Another key finding was the socio-economic dynamics on electoral outcomes, underscoring the significance of socio-economic classes in shaping political preferences. Our analysis demonstrates clear distinctions in voting behavior across different education levels, eployment statuses, and incomes. For instance, employed versus unemployed divides emerge as a critical factor, with stronger support for Biden among unemployed individuals. Thus highlights the critical concern for economic recovery strategies and social safety nets, reflecting anxieties about job security and the desire for policies that promise economic stability and support.

Income levels provided another layer of insight, revealing a divide based on economic interests and perceived benefits from policy proposals. Lower income voters' inclination towards Biden may reflect a belief in the Democratic Party's commitment to economic policies aimed at wealth redistribution, affordable healthcare, and social welfare programs. In contrast, higher-income individuals, who might prioritize tax policies and economic strategies favorable to wealth preservation, exhibited a more varied political alignment, suggesting a nuanced calculation of economic interests and policy preferences.

These socio-economic insights indicate a clear demand for political leadership that addresses the complex economic realities facing different segments of the American populace. They highlight the need for policies that bridge the economic divide, offering solutions that resonate with both lower and higher-income Americans while addressing the systemic issues of unemployment and economic insecurity.

5.3 Regional Findings

A third key finding was the regional dynamics on electoral outcomes, underscoring the significance of geographic location in shaping political preferences. Our analysis demonstrates clear distinctions in voting behavior across different regions, urbanization levels, and states. For instance, urban versus rural divides emerge as a critical factor, with voters in urban areas showing a stronger inclination towards Biden, whereas those in rural areas were more likely to support Trump. This urban-rural polarization reflects broader socio-economic and cultural differences, suggesting that issues pertinent to urban centers resonate differently with those in rural communities.

Furthermore, the state-specific analysis provides insights into the localized nature of political support, illustrating how state-level policies, historical voting patterns, and local issues can influence voter preferences. States with diverse populations and a mix of urban and rural areas exhibited varied voting behaviors, highlighting the intricate interplay between state identity and political affiliation. For example, the strong support for Biden in the District of Columbia contrasted sharply with the preferences observed in more rural states, pointing to the significant influence of local contexts on electoral decisions.

These regional insights are crucial for understanding the geographic segmentation of the American electorate and the challenges of crafting political messages that resonate across such a diverse landscape. They emphasize the need for policies and political strategies that address the specific concerns and priorities of different regions, ensuring that the voices of both urban and rural communities are heard and valued in the political discourse.

5.4 Study Weaknesses

Despite its comprehensive analysis, the study has limitations, notably in its reliance on self-reported data, which may introduce bias. Respondents may not always accurately recall their voting behavior or may portray it in a socially desirable manner, leading to discrepancies between reported and actual voting patterns. Furthermore, self-selection bias could arise if the individuals who choose to participate in the survey are not representative of the broader electorate, skewing the data towards particular demographic or political groups.

Additionally, the models might not fully capture the complexity of voter behavior, as they cannot account for all possible confounding factors, such as media influence or personal political ideologies. While logistic regression models offer valuable insights into the factors influencing voter behavior, they have inherent limitations in capturing the full spectrum of human decision-making. Voter behavior is influenced by a complex interplay of factors, including emotional responses, social influences, and dynamic political landscapes, which may not be fully accounted for by the models used. The assumption of linear relationships between variables and voting behavior overlooks the possibility of non-linear dynamics and interactions between factors that could significantly influence electoral outcomes. Additionally, the models' inability to incorporate all potential confounding factors, such as the effect of media campaigns, political events leading up to the election, and individual political ideologies, may result in an incomplete understanding of voter behavior.

Furthermore, the decision to condense the sample size to 10,000 random participants, while necessary to manage time and space limitations, introduces another layer of complexity to the study's limitations and warrants a deeper examination of its implications. Reducing the sample size to a subset of the original dataset may impact the representativeness of the study. Although random sampling methods are designed to create a sample that reflects the larger population, decreasing the sample size increases the margin of error and the potential for sampling bias. This means that the smaller sample may not adequately capture the diversity and variability of the broader electorate, particularly the nuances of smaller demographic groups or regions with less representation.

5.5 Next Steps

Future research should aim to incorporate more dynamic models that can better account for the multifaceted nature of political behavior and the influence of evolving societal trends. The advent of digital platforms has transformed how information is disseminated and consumed, significantly impacting political discourse and voter behavior. A next step would be to delve into the nuances of this digital revolution, examining the role of social media, online news outlets, and digital campaigning in shaping political opinions and electoral outcomes. This includes studying the effects of misinformation, echo chambers, and the role of algorithms in curating political content, as well as investigating strategies for fostering digital literacy and critical engagement with online content among the electorate.

Studies could also expand the scope of research to include cross-cultural comparisons can enrich our understanding of electoral behavior in a global context. Examining how different political systems, cultural norms, and historical contexts influence voter behavior can uncover universal patterns and unique variations. This approach can facilitate a deeper understanding of democracy, governance, and political engagement across diverse societies, contributing to a more comprehensive global political science discourse.

6 Appendix

Table 7: Age bucket variable values

age_bucket
45-64
65+
30-44
18-29
NA

Table 8: Race variable values

race
White
Black
Hispanic
Asian
Other

Table 9: Education variable values

education
Some college or assoc. degree
College graduate
High school or less
Postgraduate study

Table 10: Eployment status variable values

employment_status
Not in the Workforce
Employed
Unemployed

Table 11: Income variable values

income
Less than 30,000
100,000 - 199,999
30,000 - 49,999
50,000 - 99,999
200,000 or more

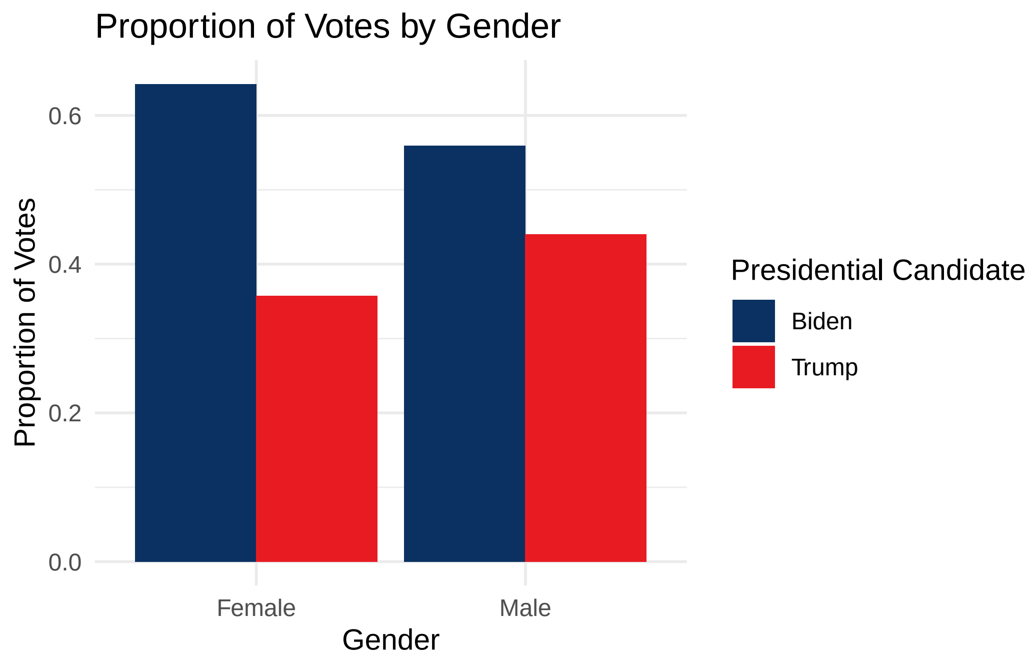

Figure 6: Proportion of Votes by Gender in the 2020 Election

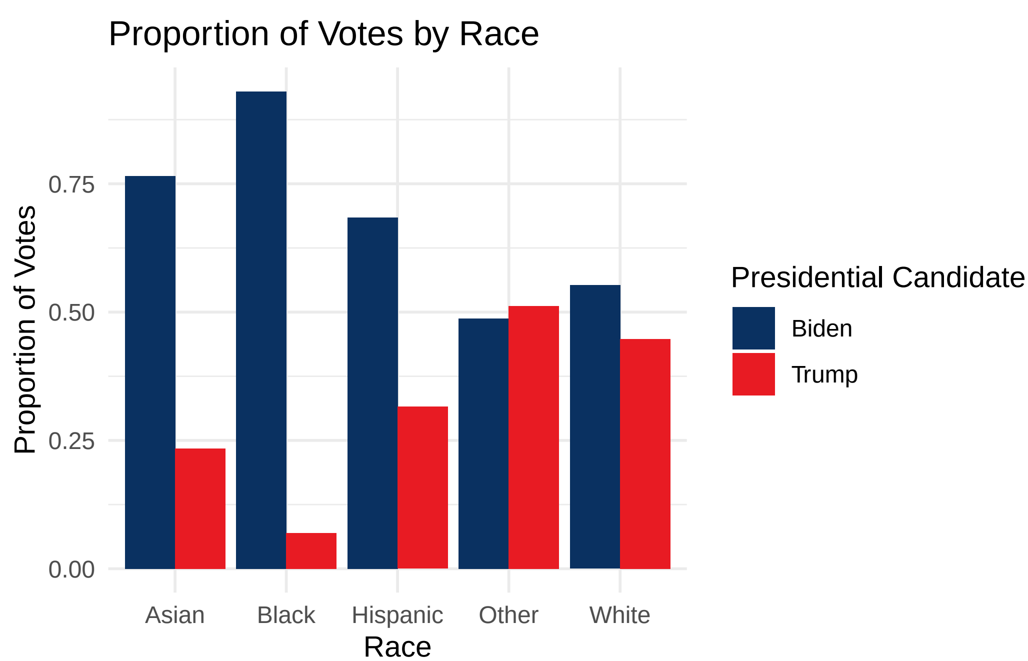

Figure 7: Proportion of Votes by Race in the 2020 Election

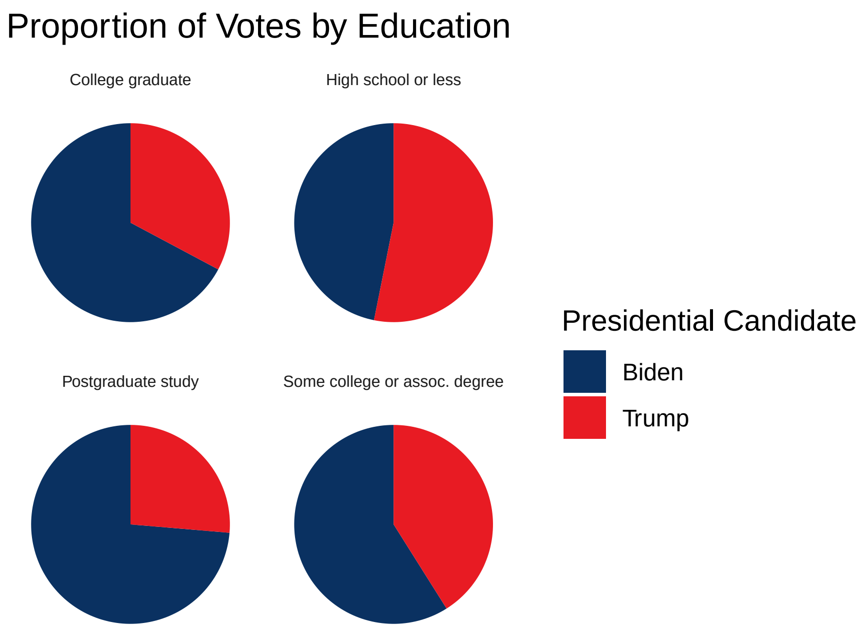

Figure 8: Proportion of Votes by Education in the 2020 Election

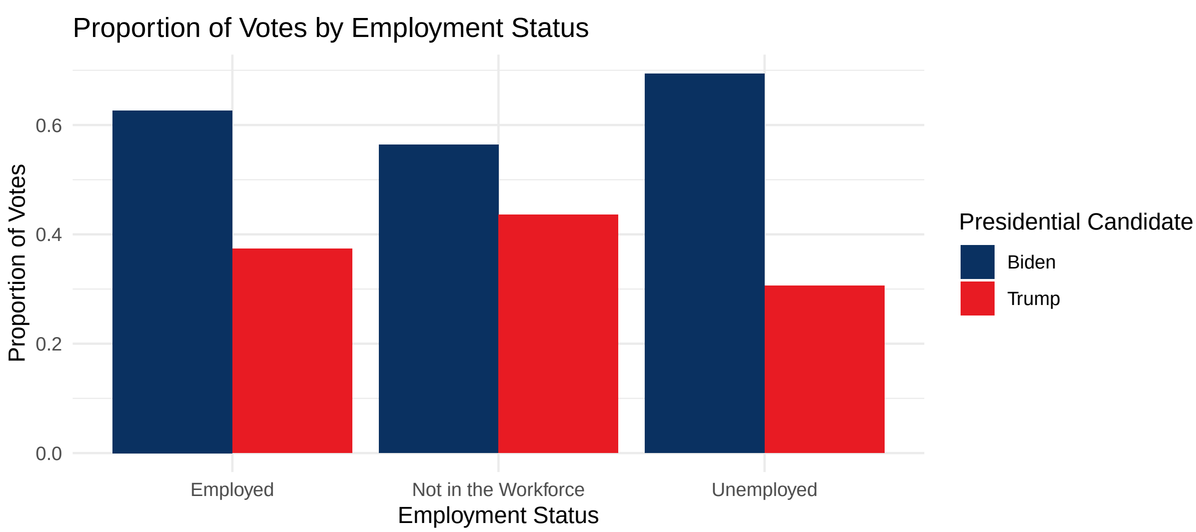

Figure 9: Proportion of Votes by Employment Status in the 2020 Election

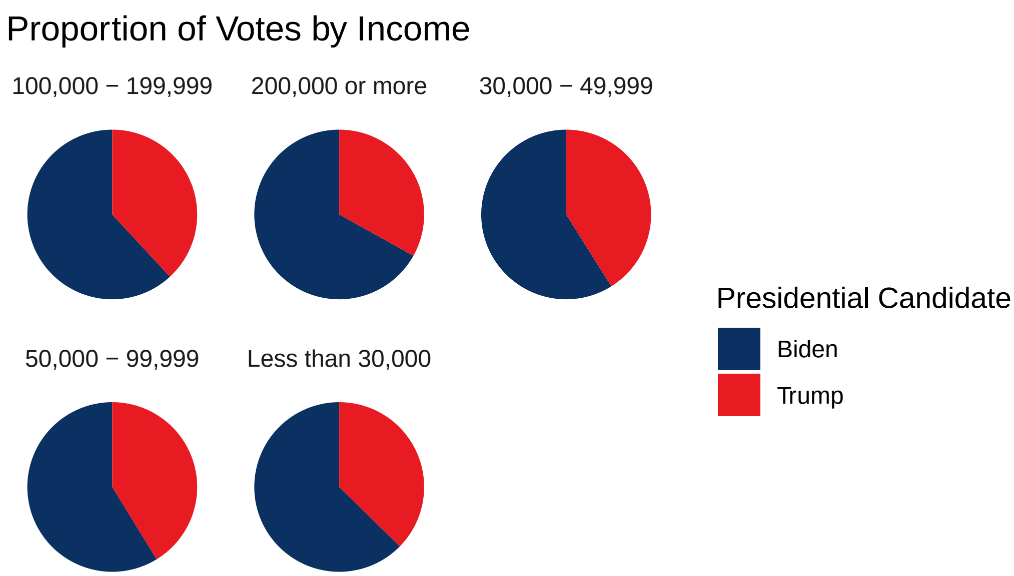

Figure 10: Proportion of Votes by Income in the 2020 Election

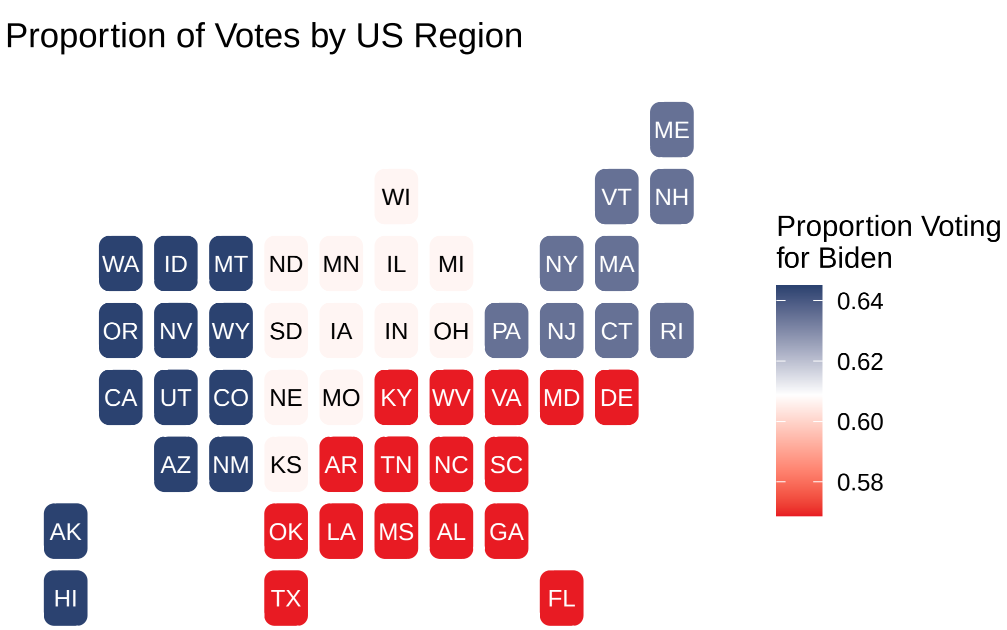

Figure 11: Proportion of Votes by Region in the 2020 Election

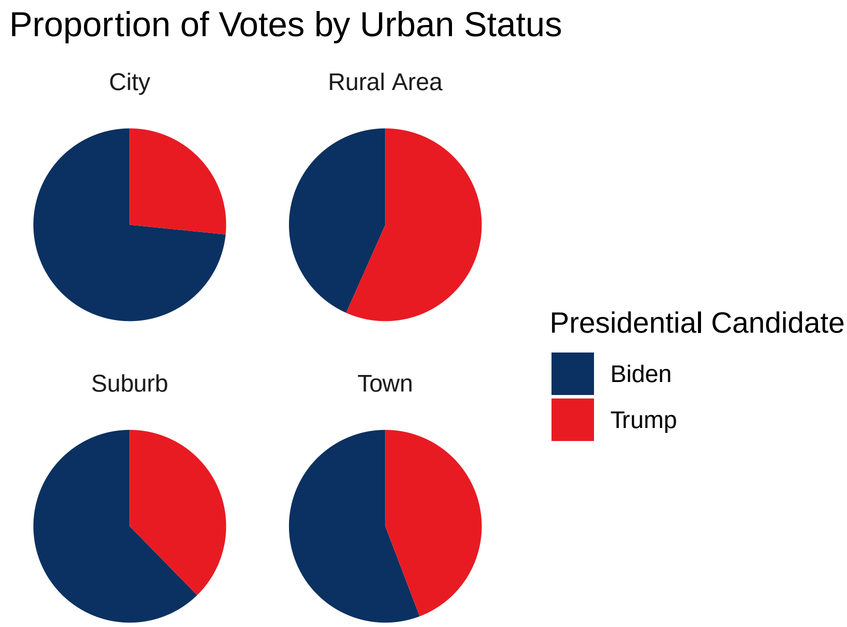

Figure 12: Proportion of Votes by Urban Status in the 2020 Election

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Center, Pew Research. 2023. *Voting Patterns in the 2022 Elections*. *Pew Research Center*. https://www.pewresearch.org/politics/2023/07/12/voting-patterns-in-the-2022-elections/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Meltwater. 2023. *About YouGov*. *YouGov*. https://today.yougov.com/about.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Politics, CNN. 2020. "Presidential Results." *CNN*. https://www.cnn.com/election/2020/results/president.

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

Stephen Ansolabehere, Brian Schaffner, and Sam Luks. 2021. "COOPERATIVE ELECTION STUDY, 2020: COMMON CONTENT." Harvard Dataverse. https://doi.org/10.7910/DVN/E9N6PH.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal* *of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.