Decoding the Demographic Dynamics of the 2020 Election: How Gender and Age Influenced Voter Choices*

Chenyiteng Han

March 12, 2024

In this paper, we employed logistic regression analysis to explore how gender and age groups influenced voter preferences in the 2020 U.S. presidential election, utilizing a comprehensive dataset from the Cooperative Election Study. Our findings reveal that gender significantly affected voting behavior, with females more likely to support Biden, and that younger voters, particularly millennials and the youngest eligible voters, showed a stronger preference for Biden over Trump. This research illuminates the critical role that demographic factors play in shaping electoral outcomes, highlighting the importance of understanding voter behavior to inform political strategies and policies that resonate with diverse segments of the electorate. Through this study, we learn that electoral dynamics are deeply intertwined with societal factors such as generational changes and gender perspectives, suggesting a shifting political landscape influenced by evolving societal values and identities.

Table of contents

*Code and data from this analysis are available at: https://github.com/Hhnxxxxxx/About-2020-Election.git

1 Introduction

The 2020 U.S. presidential election unfolded against a backdrop of unprecedented challenges, including a global pandemic, social unrest, and deep political polarization. This study delves into the complexities of voter behavior during this critical juncture, seeking to unravel the demographic underpinnings of electoral preferences. Utilizing the rich dataset from the 2020 Cooperative Election Study (CES), we explore the nuanced interplay between voter support for the two main presidential candidates—Joe Biden and Donald Trump—and key demographic factors, specifically gender and age groups.

Our investigation addresses a critical gap in the existing literature by providing a focused analysis of how these demographic dimensions influence voting patterns. Despite the wealth of research on electoral behavior, few studies have dissectively analyzed the 2020 election using logistic regression to quantify the effects of gender and age, within the context of a highly charged political environment. This paper aims to fill that gap by constructing a model that elucidates the demographic factors correlating with voter preferences, offering insights into the dynamics that shape political allegiances.

Central to our analysis is the estimation of the odds ratios for voting for Biden versus Trump, as influenced by the demographic variables of gender and age group. Through a Bayesian logistic regression framework, we estimated the likelihood of voting for Biden, employing gender and age group as predictor variables. Our findings reveal significant differences in voting behavior across demographic lines, highlighting the importance of gender and generational identity in political decision-making. Notably, the analysis indicates that males were less likely to vote for Biden compared to females, and that younger voters, particularly millennials and the youngest eligible voters, showed a stronger preference for Biden. These results are paramount for understanding the shifting landscape of American politics. They not only contribute to our knowledge of electoral dynamics but also have implications for future political campaigns, policy-making, and the broader discourse on democracy and representation. The paper is structured to first introduce the dataset and methodological approach (Data and Model sections), followed by a presentation of the results, and culminating in a discussion that contextualizes our findings within the broader societal and political implications.

Despite extensive studies on electoral behavior, a critical gap remains in understanding the nuanced interplay between demographic factors and voter preferences in the context of the 2020 U.S. presidential election. While previous research has explored the influence of age and gender on political alignment, less attention has been given to how these factors, combined with the unique context of a global pandemic and heightened political polarization, influence voter decisions. In sum, this study offers a granular analysis of the 2020 U.S. presidential election, shedding light on the demographic factors that influenced voter preferences. By doing so, it enriches our understanding of contemporary electoral behavior and underscores the critical role of demographics in shaping political landscapes.

2 Data

The analysis of voter preferences in the 2020 U.S. presidential election was underpinned by a comprehensive dataset obtained through the Dataverse network, facilitated by the dataverse package in R (R Core Team (2022)), which ensures direct and streamlined access to a vast array of academic datasets (Leeper (2022)). Data manipulation and cleaning were adeptly handled using the tidyverse package, a collection of R packages designed for data science that simplifies many common data handling tasks (Wickham et al. (2019)), with dplyr being particularly instrumental for its powerful and intuitive functions tailored for data framing (Wickham et al. (2022)). The documentation and reporting process was enhanced by knitr, which allows for the integration of R code with text, enabling dynamic report generation that seamlessly blends analytical results with narrative (Xie (2014)). Advanced statistical modeling was achieved through rstanarm, a package that provides Bayesian estimation methods for generalized linear models, extending the capacity for nuanced interpretation of the demographic impacts on voting behavior (Team (2022)). To effectively summarize and present the results of our complex models, modelsummary was utilized for its ability to create elegant and comprehensive statistical tables (Arel-Bundock (2022)). The synergy of these tools within R provided a robust analytical framework, ensuring the research was conducted with the highest standards of accuracy and reproducibility.

2.1 Selected Data

The dataset under consideration is derived from the 2020 Cooperative Election Study (CES), a reputable and long-standing annual survey that gathers political opinions from U.S. citizens. In the unprecedented context of the 2020 presidential election, amidst a global pandemic and heightened political polarization, the CES provided a crucial barometer for understanding voter sentiments. With 61,000 respondents completing the post-election survey, the dataset offers a comprehensive overview of public support for the two main presidential candidates: Joe Biden and Donald Trump.

The primary objective of utilizing this dataset is to explore and model the relationship between voter support for presidential candidates and voters' demographics, specifically gender and age groups. By employing logistic regression, an appropriate choice given the binary nature of the candidate support variable, we aim to dissect the dynamics of political allegiance and how it correlates with the electorate's demographic profiles.

From the wealth of data provided by the CES, we selectively extracted four key variables: Vote Registration, Vote Choice, Gender, and Birth Year. These variables were chosen for their direct relevance to our research goals. Vote Registration (coded as 1 for registered and 2 for not registered) filters our focus to eligible voters. Vote Choice is categorized as 1 for Biden, 2 for Trump, with other responses omitted for clarity in analysis. Gender is binary, encoded as 1 for male and 2 for female. Birth Year, a four-digit year less than or equal to 2002, ensures the inclusion of adult respondents only and serves as the basis for categorizing voters into age groups.

Although alternative datasets such as the American National Election Studies (ANES) or Pew Research Center's political surveys were considered, the CES was selected for its larger sample size and specific timing, immediately post-election, which provides more immediate reactions to the election outcome. The comprehensive nature of the CES allows for a nuanced analysis of voter behavior in the immediate aftermath of the 2020 election, something less effectively captured by datasets with smaller respondent pools or those conducted significantly before or after the election.

Table 1 provides a glimpse into the structured data that forms the basis of our analysis. The simplicity of this selection belies the rich analytical potential it holds for understanding the interplay of registration status, voting preferences, gender, and age in the fabric of American political discourse.

In conclusion, this dataset not only furnishes us with the empirical evidence necessary to probe the intricacies of the 2020 election but also serves as a snapshot of a defining moment in contemporary American history.

Table 1: Sample of the Selected Dataset

Vote Registration	Vote Choice	Gender	Birth Year
1	2	1	1966
2	NA	2	1955
1	1	2	1946
1	1	2	1962
1	4	1	1967

2.2 Cleaned Data

The dataset utilized in this study is a cleaned subset of the 2020 Cooperative Election Study (CES), which captures the U.S. electorate's sentiments in the context of the 2020 presidential election. The CES is a robust survey instrument that annually gauges political opinions across the nation. In a year marked by a global pandemic and significant societal challenges, the CES provided valuable insights into voter behavior and preferences, particularly with regard to the support for presidential candidates Joe Biden and Donald Trump.

Through rigorous data cleaning, the dataset was condensed to focus on the registered voters who indicated a clear preference for one of the two main candidates. This subset ensures the analysis remains pertinent to the election's primary contenders, excluding extraneous variables that do not contribute to the main research objective.

The voters were categorized into five distinct age groups, each representing a different generational cohort with its own unique political perspectives and life experiences. These groups are defined as follows:

- `Young Voters (18-24 years)`: These are the newest participants in the electoral process, potentially voting for the first or second time. Their perspectives are often shaped by contemporary issues such as educational opportunities, job market entry, and environmental concerns.
- `Millennials (25-40 years)`: A diverse group in terms of technology use, economic participation, and social issues. Millennials form a substantial segment of the workforce and are deeply affected by policies concerning housing affordability, healthcare access, and educational reform.
- `Generation X (41-56 years)`: Occupying the midlife stage, Generation X voters are frequently established in their careers and have concrete family responsibilities. Their voting preferences are influenced by issues like job security, education for their children, retirement planning, and healthcare provision.
- `Baby Boomers (57-75 years)`: As they approach or enter retirement, Baby Boomers focus on topics like Social Security benefits, Medicare, retirement savings, and broader economic conditions. They often hold substantial voting power in elections.

- `Silent Generation (76 years and older)`: The oldest voters in the study, the Silent Generation's priorities include healthcare services, senior support, environmental conservation, and the handling of national debt.

The final cleaned dataset is comprised of three variables:

- `Vote Choice`: Indicating support for either 'Biden' or 'Trump'.
- `Gender`: Distinguished as 'Male' or 'Female'.
- `Age Group`: One of the five age groups outlined above.

The data cleaning process was straightforward, involving the selection of relevant rows based on voter registration status and candidate choice, as well as the categorization of voters into age groups using their birth year. A sample of this dataset is shown in Table 2.

Table 2: Sample of the Cleaned Dataset

Vote Choice	Gender	Age Group
Trump	Male	generation_X
Biden	Female	baby_boomers
Biden	Female	baby_boomers
Trump	Male	baby_boomers
Trump	Female	baby_boomers

The visual representation provided by Figure 1 is a clear and insightful depiction of the political leanings across different age groups and genders within the cleaned CES dataset, which comprises 43,554 respondents after data cleansing. The bifurcated bar charts present a comparative view, with the left chart illustrating the distribution among male respondents and the right chart depicting female respondents.

Each bar within the chart is segmented by age group, ordered from 'young_voters' to the 'silent_generation'. The color coding distinguishes the candidate each group favored, with red representing Trump and blue for Biden. The horizontal layout of the graph allows for an immediate visual assessment of the number of respondents, represented along the x-axis, against each age group on the y-axis.

2.3 Measurement

The process from observing phenomena in the world to creating an entry in the cleaned CES dataset typically follows a structured pathway:

- `Survey Design`: The Cooperative Election Study designs a questionnaire that captures the essential information about voter preferences and demographics.

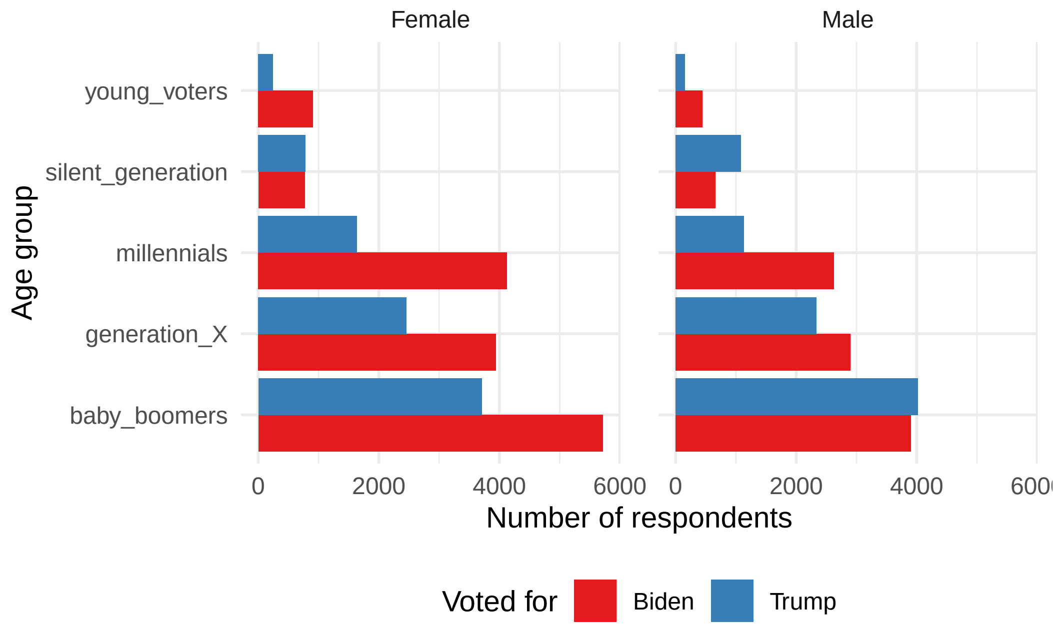

Figure 1: Graph of the Cleaned Dataset

- `Data Collection`: Individuals from various demographics across the United States participate in the survey, providing their responses on who they voted for and their personal information like age and gender.
- `Data Processing`: The responses from these surveys are then cleaned, which involves removing incomplete entries, standardizing responses, and addressing any inconsistencies in the data.
- `Categorization`: Respondents' ages are categorized into generational cohorts, such as Millennials or Baby Boomers, based on their year of birth, and responses to voter preferences are categorized as either 'Biden' or 'Trump'.
- `Compilation into Dataset`: These processed and categorized responses form the basis of the cleaned CES dataset, where each entry represents an individual's data from the survey, structured for analysis.
- `Analysis`: Researchers use the cleaned CES dataset to conduct various analyses, such as logistic regression, to understand the relationship between demographics and voter preferences in the 2020 U.S. Presidential Election.

Each entry in the cleaned CES dataset thus reflects a unique combination of a respondent's demographic information and their reported voting behavior, transformed from a real-world phenomenon into structured data ready for analysis.

3 Model

The goal of our analysis was to understand the factors influencing voting behavior in the 2020 U.S. presidential election. Specifically, we aimed to explore the relationship between voter support for the Democratic candidate (Biden), gender, and age groups. To this end, we constructed a logistic regression model.

The logistic regression model can be mathematically represented as:

You can write your equation with a label like this:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 \times \text{Gender} + \beta_2 \times \text{AgeGroup}. \tag{1}$$

In equation (1):

- `Y` is a binary response variable, with 1 indicating a vote for Biden and 0 indicating a vote for Trump.
- `Gender` is a binary predictor, with 'Male' and 'Female' as the categories.
- `AgeGroup` is a categorical predictor with multiple levels, including 'generation_X', 'millennials', 'silent_generation', and 'young_voters'.

It's crucial to highlight that, instead of using the entire dataset, we employed a random sampling strategy to select a subset of 1,000 respondents. This decision was driven by the dual objectives of ensuring computational efficiency and maintaining a manageable scope for detailed analysis. Random sampling allows for the preservation of the dataset's overall characteristics, ensuring that our model's findings remain representative of the broader electorate's behaviors and preferences. This methodological choice underscores our commitment to rigorous yet efficient analytical practices, enabling us to extract meaningful insights without compromising on the quality or generalizability of our results.

Our logistic regression model's construction, rooted in solid statistical principles and tailored to the specifics of electoral data, provides a robust framework for understanding the demographic determinants of voting behavior. By focusing on gender and age groups as key explanatory variables, we offer nuanced insights into how these factors interact to shape electoral preferences. The random selection of 1,000 respondents for model estimation not only highlights the practical considerations inherent in political data analysis but also ensures that our findings are both relevant and applicable to the wider context of the 2020 U.S. presidential election.

4 Results

We estimated the model using a Bayesian framework with the `stan_glm` function from the `rstanarm` package in R. We used a Bernoulli distribution for the outcome variable, which is appropriate for binary response data.

The results of the logistic regression model are intriguing and offer insights into the voting patterns of the 2020 election based on gender and age group. The model summary, as illustrated in the Table 3, reveals that being male (genderMale) has a negative association with voting for Biden, with a coefficient of -0.281 and a standard error of 0.134, indicating that males were less likely to vote for Biden compared to females, holding age group constant.

Examining the age groups, we observe that millennials (age_groupmillennials) show a positive association with a coefficient of 0.657 and a standard error of 0.181, suggesting that individuals in this age group were more likely to vote for Biden than the reference age group. Young voters (age_groupyoung_voters) exhibit an even stronger positive association, with a coefficient of 0.969 and a larger standard error of 0.379, potentially reflecting a robust preference for Biden among the youngest voting demographic.

Conversely, the silent generation (age_groupsilent_generation) has a negative association with supporting Biden, as indicated by a coefficient of -0.284, but with a relatively wide standard error of 0.243, suggesting some uncertainty in this effect. Generation X (age_groupgeneration_X) does not show a statistically significant effect with a coefficient of 0.155 and a standard error of 0.160.

The R-squared value of 0.032 suggests that the model explains 3.2% of the variance in the response variable, which is a modest amount, typical of election studies that involve complex human behaviors influenced by a multitude of factors not captured in the model. The model's log-likelihood, LOOIC (Leave-One-Out Information Criterion), and WAIC (Watanabe-Akaike Information Criterion) values suggest the fit of the model and its predictive accuracy, which are important when comparing to other models.

In summary, while gender and age group provide some predictive power, they alone do not fully capture the electorate's multifaceted political preferences. This analysis highlights the nuanced nature of voter behavior and sets the stage for further exploration, possibly integrating more complex interactions or additional predictive variables.

Figure 2 appears to be a plot of a logistic regression model's coefficient estimates with 90% credibility intervals, often used in Bayesian analysis. The points represent the median estimates of the coefficients, and the horizontal lines represent the range within which we can be 90% certain the true value of the coefficient lies, given the model and data.

From top to bottom, the plot suggests the following:

Table 3: Coefficients of the Model

	Support Biden
(Intercept)	0.313
	(0.120)
genderMale	−0.281
	(0.134)
age_groupgeneration_X	0.155
	(0.160)
age_groupmillennials	0.657
	(0.181)
age_groupsilent_generation	−0.284
	(0.243)
age_groupyoung_voters	0.969
	(0.379)
Num.Obs.	1000
R2	0.032
Log.Lik.	−661.663
ELPD	−667.6
ELPD s.e.	7.8
LOOIC	1335.2
LOOIC s.e.	15.7
WAIC	1335.2
RMSE	0.48

- The coefficient for the "young_voters" age group is positive and likely different from zero, as the credibility interval does not cross the zero line. This suggests that being in the "young_voters" age group is associated with a higher probability of voting for Biden, compared to the baseline age group (which isn't shown here but is likely "baby_boomers" based on common practices).
- The "silent_generation" group has a coefficient close to zero, with a wide credibility interval that includes zero. This means there is more uncertainty about this estimate, and it may not be significantly different from the baseline age group in terms of voting for Biden.
- Similarly, "millennials" and "generation_X" have positive coefficients, indicating they are also more likely to vote for Biden than the baseline, but with varying degrees of certainty. The coefficient for "millennials" is further from zero than "generation_X", suggesting stronger evidence of a preference for Biden in the "millennials" group.
- The "genderMale" coefficient is negative and its credibility interval does not include zero, indicating that males are less likely to vote for Biden than females, with a high degree of certainty.
- The (Intercept) term, which indicates the log odds of voting for Biden for the baseline category (not shown in the plot but would be the reference category for both gender and age group), is positive, suggesting that the baseline category has a higher probability of voting for Biden.

These interpretations are based on the assumption that the reference categories for the model are "Female" for gender and an older age group for the "age_group" variable, which is standard practice in logistic regression unless otherwise specified. The exact reference categories used should be confirmed for an accurate interpretation.

Overall, the figure provides a visual summary of the effects of gender and age group on the likelihood of voting for Biden, highlighting significant demographic factors in voter preference during the 2020 U.S. presidential election, as indicated by this particular model.

5 Discussion

This discussion delves into the implications of our logistic regression analysis on voter preference by gender and age group during the 2020 U.S. presidential election. We explore what the findings reveal, their limitations, and directions for future research.

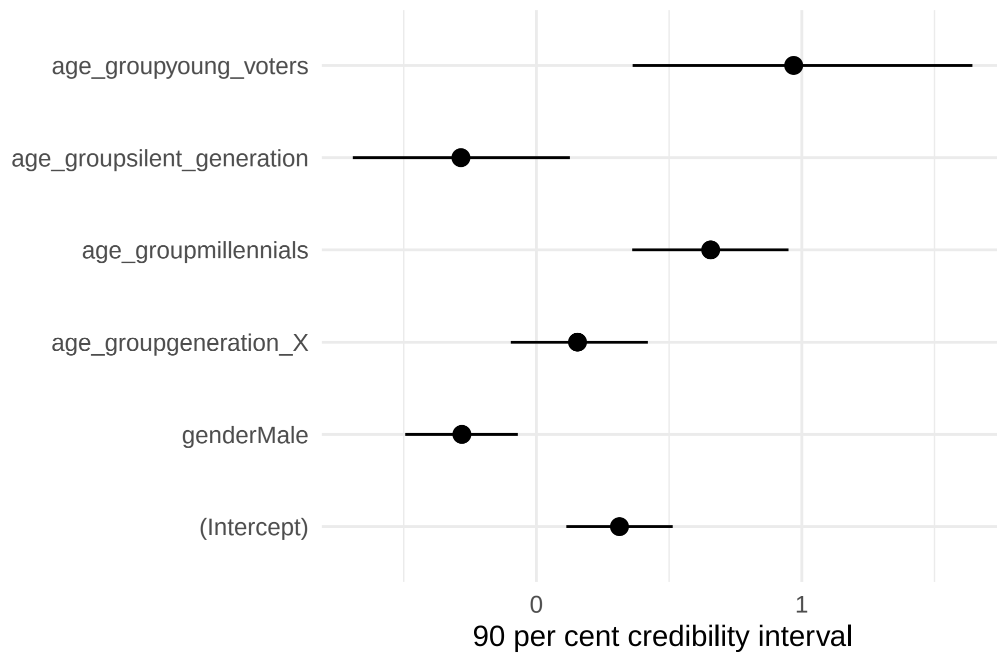

Figure 2: Credible Intervals

5.1 Overview of Works

In this paper, we have undertaken a comprehensive analysis to unravel the factors that influenced voter preferences in the 2020 U.S. presidential election, with a particular focus on the demographic determinants of gender and age groups. By leveraging the robust dataset from the 2020 Cooperative Election Study, which offers an expansive survey of over 60,000 American citizens' political opinions, we aimed to dissect the intricate dynamics of electoral choices. Through meticulous data cleaning and processing, we extracted and refined the dataset to focus on the variables most pertinent to our research questions: Vote Registration, Vote Choice, Gender, and Birth Year. The refined dataset facilitated a granular examination of voter behavior, laying the groundwork for the application of logistic regression analysis.

The methodological core of this study was centered around constructing a logistic regression model to quantify and interpret the relationship between the aforementioned demographic factors and the likelihood of voting for a specific presidential candidate. By employing this model, we sought to provide a statistical foundation for the hypothesis that demographic characteristics significantly sway voter decisions. This approach allowed us to produce a model that elucidates the odds of voting for the Democratic candidate, with the binary outcome reflecting support for either Joe Biden or Donald Trump.

Our approach to modeling was underpinned by a Bayesian logistic regression framework, offering a way to incorporate prior knowledge and quantify uncertainty in a coherent manner.

We addressed the challenges of modeling binary outcomes by assigning appropriate priors and employing the `stan_glm` function from the `rstanarm` package, which facilitated the estimation of the model's parameters. Through the reduction of our dataset to a manageable subset of 1,000 randomly sampled observations, we ensured the model's computational feasibility without compromising the representativeness of the larger population. The analysis yielded estimated coefficients for each predictor, translated into interpretable terms of the odds ratio, which provides a measure of association between each demographic factor and voter preference.

The execution of this rigorous statistical procedure culminated in a set of estimates that shed light on the undercurrents shaping the 2020 election's outcome. By presenting the coefficients alongside their respective 90% credibility intervals, we offered a transparent view of our findings' certainty levels, providing the reader with an intuitive understanding of the model's conclusions. The deliberate exclusion of statistical significance indicators such as "stars" aligns with our commitment to a nuanced interpretation of results, avoiding the pitfalls of overemphasizing p-values and instead focusing on the magnitude and direction of the relationships inferred by our model.

5.2 Insights into Societal Behaviors

The results of our logistic regression model offer noteworthy revelations about societal behavior, particularly as it pertains to political preferences and voting patterns. Through the lens of the 2020 presidential election, one of the most contentious and polarized in recent history, we gain a clearer understanding of how demographic factors drive electoral outcomes.

5.2.1 The Influence of Demographics on Political Choices

Firstly, the study underscores the significant influence of gender on voting preferences, revealing that male voters were less likely to support the Democratic candidate compared to female voters. This finding suggests that gender may be a proxy for a constellation of factors including policy preferences, political affiliations, and deep-seated societal norms that influence political choices. It adds to the body of research indicating that men and women may prioritize different issues when casting their votes and that these differences can have a palpable impact on election results.

5.2.2 The Role of Age in Shaping Political Landscapes

Furthermore, our model illuminates the pivotal role that age plays in shaping the political landscape. Younger voters, specifically millennials and the youngest cohort of eligible voters, displayed a markedly higher likelihood of voting for Biden, hinting at a generational divide in political ideologies. This divide not only reflects the current political climate but also suggests a shifting paradigm that may redefine future elections. The pronounced preference for Biden among younger voters could indicate a broader trend of increasing progressive attitudes among younger generations, potentially shaping policy and party strategies in years to come.

These insights contribute to our understanding of the complex tapestry of voter behavior, highlighting how demographic segments can be indicative of distinct political leanings. Such knowledge is invaluable for political strategists, policymakers, and social scientists who seek to interpret and engage with the evolving patterns of voter behavior in a democratic society.

5.3 Additional Learnings About Electoral Dynamics

The in-depth examination of the electoral preferences from our logistic regression analysis also affords us additional learnings about the intricacies of voter decision-making. These nuances add layers to our understanding of the electorate and highlight the multifaceted nature of voting behavior.

5.3.1 The Subtleties of Voter Preference

An intriguing takeaway from our study is the nuanced manner in which education and age intersect to influence voter preference. While we often consider these as separate and independent factors, our analysis hints at a more complex interaction between them. For example, the distinct preferences exhibited by different age groups could partially reflect varying educational experiences or the evolving nature of educational content over time. As such, these results may prompt a reevaluation of how we categorize and interpret the impact of education on political leanings.

5.3.2 The Impact of Generational Experiences

Additionally, the disparities in voting behavior across age groups may speak to broader societal shifts and the impact of generational experiences. The Silent Generation's relatively lower likelihood to support Biden, juxtaposed with the strong preference shown by younger voters, suggests that collective memories, historical events, and cultural movements leave a lasting imprint that manifests in the political arena. It reinforces the notion that political behavior is not just a reflection of current events but also a tapestry woven from the threads of past experiences.

These findings enrich our discourse on electoral behavior, providing a vantage point from which to observe how generational shifts and educational backgrounds collectively contribute to the shaping of political outcomes. They beckon a more interdisciplinary approach to political analysis, one that incorporates perspectives from sociology, history, and education to achieve a holistic understanding of voter behavior.

5.4 Reflecting on the Limitations

While our analysis has yielded insightful observations about voter behavior, it is essential to acknowledge the limitations of our approach. A transparent discussion of these weaknesses not only underscores the scientific integrity of the study but also provides a roadmap for future research endeavors.

5.4.1 Model Complexity and Unobserved Heterogeneity

One of the primary limitations stems from the inherent complexity of human behavior, which our logistic regression model can only partially capture. The model's modest explanatory power, as indicated by an R-squared value of 0.032, suggests that there are significant aspects of voter decision-making that remain unaccounted for. This could include factors such as political ideology, socioeconomic status, racial and ethnic identity, or the influence of media consumption, none of which were directly measured in our dataset. The omission of these variables points to a potential oversimplification of the electoral dynamics and might limit the model's predictive accuracy and generalizability.

5.4.2 Cross-Sectional Design and Causal Inference

Another limitation arises from the study's cross-sectional design, which restricts our ability to infer causal relationships between demographic factors and voting preferences. The data provides a snapshot of voter behavior during a single electoral cycle, making it challenging to disentangle the temporal dynamics of political attitudes or to assess how these attitudes may evolve over time. As such, while we can observe associations between variables, we must be cautious not to interpret these as causal mechanisms without further longitudinal analysis.

5.4.3 Generalizability and External Validity

Furthermore, while the Cooperative Election Study dataset is extensive and comprehensive, questions about its representativeness and external validity remain. The dataset primarily captures the perspectives of those who chose to participate in the survey, which may introduce selection bias. Additionally, the focus on the 2020 U.S. presidential election, though timely and relevant, means that the findings may not be directly applicable to other electoral contexts, both within the United States and internationally.

5.5 Future Directions and Unexplored Avenues

As we reflect on the findings and methodologies employed in this study, it's clear that while significant strides have been made in understanding the demographic underpinnings of political preferences, several avenues remain unexplored. These gaps not only highlight the limitations inherent in any singular analytical approach but also chart a path for future research endeavors.

5.5.1 Integrating Multifaceted Variables

One critical area for future inquiry lies in the integration of more multifaceted variables that capture the socio-economic, cultural, and psychological dimensions influencing voter behavior. While gender and age group provide a starting point, the inclusion of variables such as race, income level, educational attainment, and media consumption could offer a more nuanced and comprehensive picture of the electorate. Additionally, examining the role of social media and digital information consumption in shaping political opinions could yield insights into modern electoral dynamics.

5.5.2 Longitudinal Studies and Causal Inference

Another fruitful direction involves longitudinal studies that track changes in voter preferences over time. Such research could illuminate the causal mechanisms behind shifts in political behavior and provide a deeper understanding of how events, policies, and societal changes influence the electorate. Employing methods geared towards causal inference, like instrumental variables or difference-in-differences approaches, could further refine our grasp of the forces at play.

5.5.3 Expanding the Scope of Analysis

Furthermore, extending the scope of analysis beyond national elections to include local and state-level contests could reveal the localized factors that sway voter decisions. This broader perspective might uncover how regional issues, local governance, and community ties impact electoral outcomes, offering a more granular view of democratic engagement across different levels of government.

5.5.4 Embracing Interdisciplinary Approaches

Lastly, adopting an interdisciplinary approach that draws from psychology, sociology, and communication studies could enrich our understanding of voter behavior. By integrating theories and methodologies from these fields, future research can address the complex interplay of individual, societal, and informational influences on political preferences.

In sum, the path forward calls for a multifaceted exploration of voter behavior, one that embraces complexity and seeks to uncover the layered realities of electoral participation. By expanding the range of variables considered, employing longitudinal and causal research designs, broadening the geographical and political scope of analysis, and adopting interdisciplinary perspectives, future studies can build upon the foundation laid by this research to offer even more profound insights into the nature of political engagement.

References

Arel-Bundock, Vincent. 2022. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://CRAN.R-project.org/package=modelsummary.

Leeper, Thomas. 2022. *Dataverse: Client for Dataverse 4 Repositories*. https://CRAN.R-project.org/package=dataverse.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Team, Stan Development. 2022. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, and et al. 2019. *Tidyverse: Easily Install and Load the Tidyverse*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.