The Generational and Demographic Determinants of Voting Behavior: Evidence from the 2022 CES*

A Logistic Regression Approach to Understanding Racial and Geographic Influences on Political Behavior

Jinyan Wei

December 1, 2024

Over the past decades, American voting behavior has shown distinct generational and demographic patterns. This study examines how age cohorts and demographic characteristics shape partisan preferences in the 2022 election cycle. Using data from the 2022 Cooperative Election Study (CES), we investigate the intersection of generational differences, socioeconomic status, and regional variation in voting behavior. Through binary logistic regression analysis, we find significant differences across age cohorts, with older voters showing distinct partisan preferences compared to younger generations. Additionally, these age-based patterns vary meaningfully by gender, education level, and geographic region. The results suggest that generational experiences and demographic factors play crucial roles in shaping contemporary American political alignment. These findings provide insights into the evolving nature of partisan identification and highlight the importance of considering both generational change and demographic composition in understanding electoral behavior.

Table of contents

[*] Code and data are available at: [https://github.com/DianaShen1224/Relationship-between-infant-mortality-rate-and-prohibited-abortion).

1 Introduction

The determinants of voting behavior in the United States have been a focal point of political science research, particularly as the electorate becomes more diverse and polarized. Among these determinants, educational attainment and racial identity have emerged as pivotal factors shaping political preferences. Yet, the interaction between these variables remains insufficiently explored, especially in light of recent social and political developments. Understanding how education levels influence voting behavior across racial and demographic subgroups is essential to deciphering contemporary patterns of political alignment and participation.

This study investigates the interplay between educational attainment, racial identity, and other demographic factors in shaping partisan voting patterns. Drawing on data from the 2022 Cooperative Election Study (CES), which includes responses from 60,000 registered voters across all U.S. states, we analyze how these variables collectively influence the likelihood of supporting Democratic or Republican candidates. The CES dataset provides a rich source of insights into voter preferences, combining validated registration data with extensive demographic and socioeconomic information.

Using binary logistic regression models, this analysis examines the direct and interaction effects of education, race, income, and urbanicity on voting behavior, while controlling for variables such as gender, religion, and gun ownership. By leveraging both weighted and unweighted models, this study provides robust estimates of these relationships and highlights the nuances in voting behavior across subgroups. This methodological framework enables us to identify disparities and unique trends within the electorate, revealing how intersections of education and racial identity influence partisan alignment.

The findings demonstrate that the impact of educational attainment on voting preferences is deeply stratified by racial identity and other demographic characteristics. For instance, while higher levels of education are associated with greater support for Democratic candidates among White and Black voters, this relationship is less pronounced among Hispanic and Asian voters, suggesting that cultural and contextual factors mediate these effects. These results underscore the importance of adopting a multidimensional approach to understanding voter behavior in the United States.

The remainder of this paper is organized as follows: Section 2 details our data sources and variable measurements, Section 3 presents our multinomial regression methodology and model specifications, Section 4 discusses our empirical findings, and Section 5 concludes with implications and directions for future research. Additional methodological details and robustness checks are provided in Appendix- A, Appendix- B, and Appendix- C.

2 Data

2.1 Overview

We conduct our analysis of voting behavior in the 2022 U.S. election using the R programming language (**RCoreTeam?**). Our dataset, derived from the 2022 Cooperative Election Study (CES) (**ces_2022?**), provides a comprehensive view of voting preferences across demographic, socioeconomic, and geographic variables. The CES dataset combines validated voter registration data with extensive survey responses, enabling detailed analysis of key factors influencing political behavior. Following the guidance outlined in *Telling Stories with Data* by Alexander (2023) (Alexander 2023), this study employs robust statistical techniques to explore demographic influences such as race, age, education, and urbancity, alongside contextual variables like income and religion.

To support our workflow, we utilized several R packages for data cleaning, analysis, and visualization. The `tidyverse` package (Wickham et al. 2019) was foundational, offering tools for efficient data wrangling and exploration, while `arrow` (Richardson et al. 2024) managed parquet files for optimal storage and compatibility with large datasets. The `dplyr` package (**dplyr?**) facilitated data manipulation, and `ggplot2` (Wickham 2016) was employed to create informative visualizations. For report generation, `knitr` (Xie 2014) and `kableExtra` (Zhu 2024) were used to produce clean and reproducible tables. The `caret` package (Kuhn and Max 2008) enabled model validation through K-fold cross-validation, while the `pROC` package (**pROC?**) was used to assess the model's performance via the Receiver Operating Characteristic (ROC) curve.

Geographic patterns in voting behavior were visualized using the `maps` package (**maps?**), allowing for regional analyses. Additionally, the `prediction` package (**prediction?**) supported efficient extraction and interpretation of model predictions. Together, these tools, along with the structured workflow advocated by Alexander (2023), ensured a reproducible and comprehensive analysis of voting behavior in the United States.

2.2 Measurement

The process of translating survey responses into a structured dataset for the CES 2022 analysis requires a systematic approach to measurement and data gathering. In this research, we aim to investigate the factors influencing voter preferences, particularly the interaction between urbancity and race in shaping voting behavior. The CES dataset captures responses to questions about voter choice, political attitudes, and demographic characteristics, enabling a comprehensive analysis of election dynamics in the United States. Survey items are carefully designed to address diverse aspects of voter behavior, such as political affiliation, policy preferences, and party identification.

To ensure a representative sample, the CES employs matched random sampling techniques stratified by demographics and geography. Respondents are recruited using various methods, including online panels and targeted outreach, to reflect the diversity of the U.S. electorate. After responses are collected, the data undergoes rigorous cleaning and validation procedures to address inconsistencies, rectify missing values, and ensure the dataset's accuracy. For instance, variables like education and race are recategorized into uniform groups to standardize comparisons. Weighting is applied to adjust for potential sampling biases, accounting for demographic disparities in age, gender, and state representation.

The cleaned and validated data is then stratified and aggregated to examine voting patterns and trends across different subgroups. Statistical models, including multinomial logistic regression, are employed to analyze how race and education interact to influence voting preferences, highlighting nuances in voter behavior across demographic lines. This structured methodology transforms individual survey responses into actionable insights, providing a detailed understanding of how demographic factors shape electoral outcomes. By combining robust survey design with advanced data analysis techniques, this study captures the complex dynamics of contemporary U.S. elections.

2.3 Variables

The dataset incorporates a range of variables to capture demographic, socioeconomic, and geographic characteristics of registered voters. These variables are categorized as follows:

2.3.1 Outcome Variable

The primary outcome variable, `vote_choice`, is binary, indicating whether a respondent supports a specific candidate. This allows us to analyze the factors influencing voter choice in the 2022 election.

2.3.2 Predictor Variables

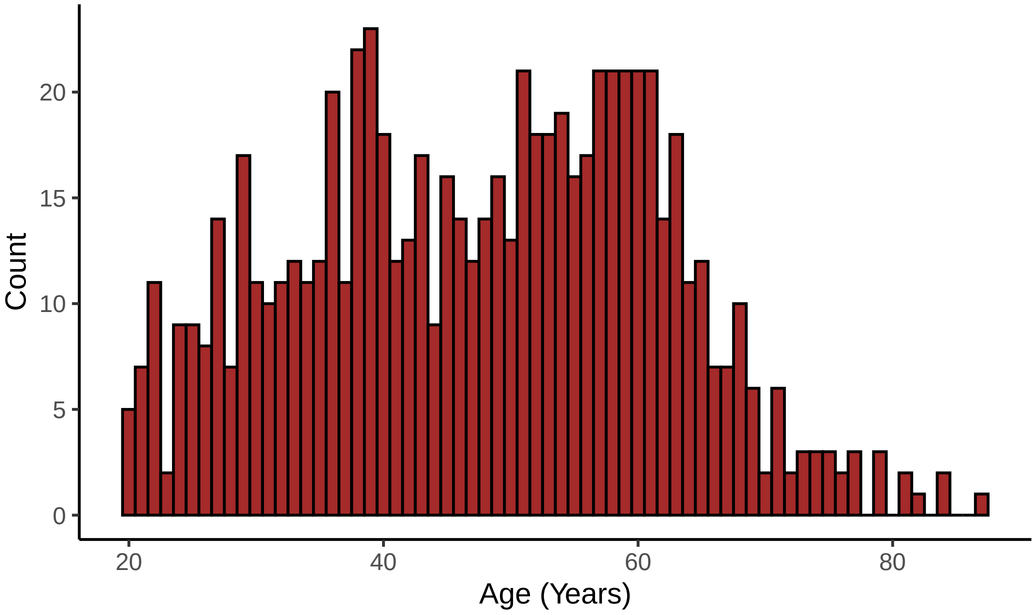
Key predictors include:

- **Age_cohort**: Categorical variable with four groups (18-29, 30-49, 50-64, 65-90)
- **Education**: Highest educational attainment, categorized as "high school or less," "some college," "college graduate," or "postgraduate."

- **Gender**: A categorical variable capturing self-identified gender (Male, Female, or Other).
- **Race**: Self-identified racial or ethnic group, categorized as "White," "Black," "Hispanic," "Asian," "Native American," "Middle Eastern," and "Other."
- **Urbancity**: A categorical variable classifying respondents as residing in urban, suburban, or rural areas.
- **Religion**: Religious affiliation, measured alongside attendance frequency.
- **Region**: Religious affiliation and attendance frequency.
- **Income_tier**: Household income level.

These variables were selected to reflect key factors identified in the literature as significant predictors of voting behavior. By including a diverse set of predictors, the analysis captures nuanced dynamics in voter preferences and behavior across different demographic and geographic groups.

2.4 Relationships between varaibles



Data source: CES 2022.

Figure 1: Distribution of Respondent Age

Figure 1 displays the age distribution of respondents. The count of respondents peaks in the 40 to 60-year range, with a noticeable drop-off in the higher and lower age brackets. The distribution appears somewhat uniform across the middle age ranges, with several spikes indicating larger groups of respondents at specific ages. The age range spans from 20 to approximately 85 years, with fewer respondents in the younger and older age groups.

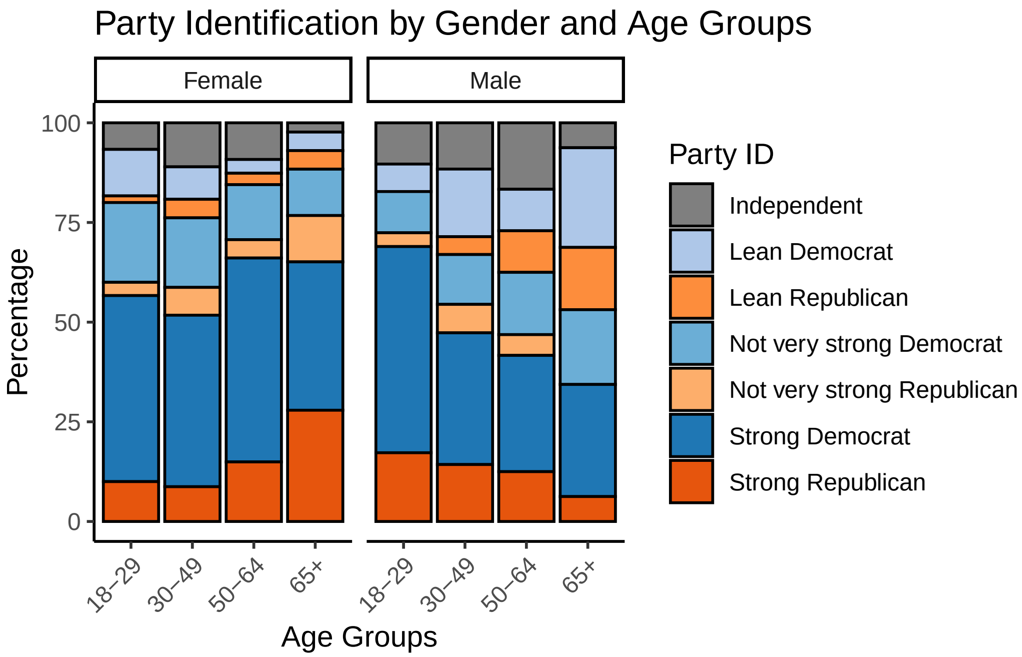
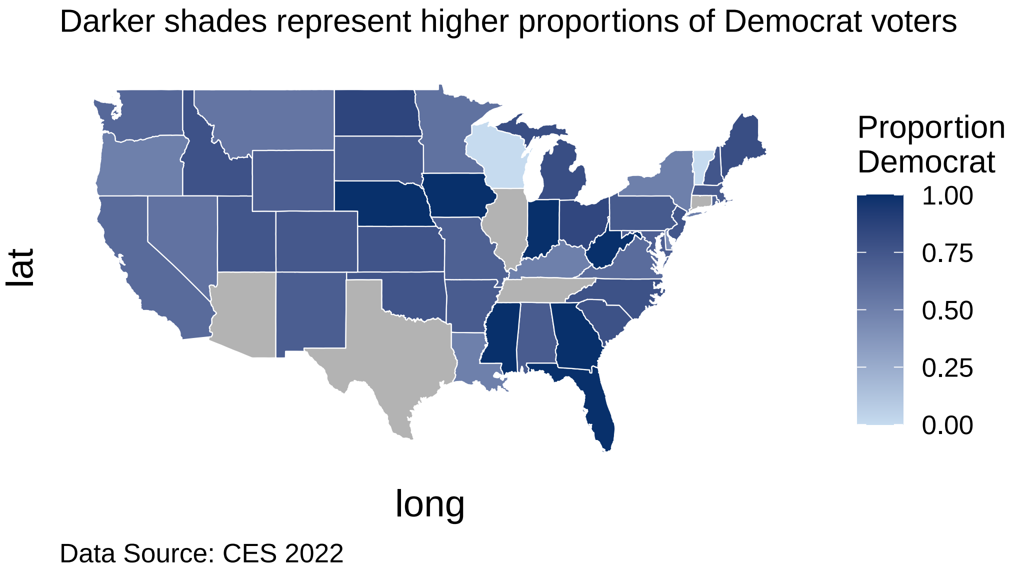


Figure 2: Party Identification by Gender and Age Groups.

Figure 2 illustrates the distribution of party identification across different age groups (18-29, 30-49, 50-64, and 65+) segmented by gender. depicts the intersection of gender, age, and party identification. It reveals patterns of political affiliation across different age brackets for male and female respondents. Notable trends include the strong Democratic identification in younger age groups and the rise of Republican identification as age increases. Independents are more evenly distributed, emphasizing the variability of non-affiliated voters across age and gender demographics. This visualization highlights how age and gender contribute to partisan tendencies within the electorate.

Figure 3 visualizes the proportion of Democrat voters across the United States based on polling data. The color gradient on the map indicates the proportion of Democrat voters, with darker shades representing a higher percentage of Democrat support. The states with the darkest blue colors indicate a strong preference for the Democratic party, while lighter colors represent weaker support. States in gray either have missing data or are not included in the polling sample. The visual highlights regional variations in voting preferences, with some states consistently supporting the Democratic party, while others, particularly in the South and parts of the Midwest, show lower levels of support.



Darker shades represent higher proportions of Democrat voters

Data Source: CES 2022

Figure 3: Proportion of Democrat Voters by State shows the distribution of Democrat voters across the United States, with darker shades representing a higher proportion of Democrat voters in each state.

3 Model

OOur modeling approach seeks to examine how demographic, socioeconomic, and geographic factors collectively influence partisan voting preferences during the 2022 U.S. election cycle. For this analysis, we employ a logistic regression model to predict the likelihood of voting for the Democratic party (`vote_choice = 1`) compared to voting Republican or for other parties (`vote_choice = 0`). This model is implemented using the `glm()` function in R, applying a binomial distribution with a logit link function to capture the binary nature of the voting outcome.

The predictors used in the model include a combination of demographic, socioeconomic, and regional variables. `age_cohort` divides respondents into generational groups (18–29, 30–49, 50–64, and 65–90), reflecting differences in life stage and political priorities. `gender` accounts for self-identified gender categories, while `education` captures the highest level of educational attainment, ranging from high school or less to postgraduate degrees. `income_tier` is used to approximate socioeconomic status, while `religion` represents self-identified religious affiliation and its potential influence on political behavior.

Race and geographic context are central to this analysis. `race` captures the self-reported racial or ethnic identity of respondents, while `urbancity` differentiates urban, suburban, and rural areas. `region`, categorized as Northeast, Midwest, South, and West, provides a broader geographic context. These variables collectively allow the model to capture the intersectional dynamics of race, geography, and demographic factors in shaping partisan preferences.

The logistic regression model assumes that the probability of voting for the Democratic party, given these predictors, follows a logistic distribution. This framework enables the estimation of the effects of individual variables and their interactions, particularly between race and geography, on the log-odds of voting Democrat. By leveraging this approach, the model provides insights into the nuanced ways in which race, urbanicity, and demographic factors influence voting behavior, highlighting critical regional and intersectional trends.

3.1 Model Set-Up

The model predicts the likelihood of voting for the Democrat party by constructing a logistic regression model using the following predictor variables:

- `age`: A continuous variable representing the respondent's age.
- `income_tier`: Categorical variable indicating the respondent's income level.
- `education`: The highest level of education attained by the respondent.
- `gender`: Categorical variable capturing the respondent's gender.
- `religion`: Categorical variable indicating the respondent's religious affiliation.
- `race`: Categorical variable representing the respondent's racial or ethnic background.
- `urbancity`: Variable indicating whether the respondent resides in an urban, suburban, or rural area.

We include interaction terms and control variables to account for the potential influence of demographic and socioeconomic factors on voting behavior.

3.1.1 Logistic Regression Model

The logistic regression model predicts the probability of voting for the Democrat party based on the predictors listed above. The model is specified as:

$$\begin{aligned}
y_i|\eta_i &\sim \text{Bernoulli}(P(y_i = 1)) \\
P(y_i = 1) &= \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \\
\eta_i &= \beta_0 + \beta_1 \cdot \text{Age Cohort}_i + \beta_2 \cdot \text{Income Tiers}_i \\
&\quad + \beta_3 \cdot \text{Education}_i + \beta_4 \cdot \text{Gender}_i \\
&\quad + \beta_5 \cdot \text{Religion}_i + \beta_6 \cdot \text{Race}_i + \beta_7 \cdot \text{Urbancity}_i + \beta_8 \cdot \text{Region}_i
\end{aligned}$$

Where:

- y_i is the binary outcome variable, where $y_i = 1$ indicates voting Democrat and $y_i = 0$ indicates voting Republican.
- β_0 is the intercept term for baseline log-odds.

- β_1 represents the effect of age cohort (18-29, 30-49, 50-64, 65-90)
- $\beta_2, \beta_3, \ldots, \beta_8$ are the coefficients for each predictor, indicating their impact on the log-odds of voting Democrat.

The model is implemented in R using the `glm()` function with a binomial family and logit link function to estimate the probability of Democratic party support as a function of demographic and generational characteristics.

3.1.2 Model Justification

Scholars in political science have long recognized that demographic factors such as age, gender, education, race, and urbancity, alongside socioeconomic and geographic contexts, significantly shape voting behavior in the United States. In particular, the intersection of race and urbancity has emerged as a critical area of study, with prior research demonstrating that the influence of geographic context on political preferences often varies across racial and ethnic groups. Our analysis aims to examine these dynamics by including an interaction term between race and urbancity, enabling a deeper understanding of how these factors jointly influence partisan voting behavior.

This study employs a logistic regression model to predict the likelihood of voting Democrat (coded as 1) versus voting Republican or for other parties (coded as 0). Logistic regression is particularly suited for modeling binary outcomes, allowing us to estimate the odds of supporting the Democratic party based on a range of demographic, socioeconomic, and geographic predictors. Key predictors include age cohort, gender, education, income tier, religion, race, and urbancity, with the interaction between race and urbanicity capturing the nuanced effects of these intersecting variables on voting behavior.

The model is estimated using maximum likelihood estimation (MLE), which identifies parameter values that maximize the likelihood of observing the data. This approach ensures robust and interpretable estimates of the relationships between predictors and voting preferences. To evaluate the model's performance, diagnostics such as the Akaike Information Criterion (AIC) and Receiver Operating Characteristic (ROC) curves are employed, providing measures of model fit and predictive accuracy. These diagnostics ensure the reliability of the results and support a nuanced interpretation of how demographic and geographic variables shape partisan alignment. Further methodological details are provided in Section Section B.

4 Results

4.1 Model Results

Table 1 shows the results of the logistic regression model predicting voting preferences are summarized in Table 1. The coefficients indicate the effects of demographic and contextual

Table 1: Summary of Logistic Regression Model Predicting Voting Choices Based on Demographic and Contextual Factors: An Analysis of CES 2022 Data

	(1)
(Intercept)	2.327
	(0.971)
age_cohort(29,49]	−0.043
	(0.349)
age_cohort(49,64]	−0.319
	(0.347)
age_cohort(64,90]	−0.193
	(0.427)
genderMale	−0.396
	(0.216)
income_tierUpper income	−0.743
	(0.423)
urbancityUrban	0.833
	(0.210)
regionNortheast	−0.274
	(0.381)
regionSouth	−0.108
	(0.298)
regionWest	−0.319
	(0.257)
Num.Obs.	718
AIC	699.4
BIC	845.9
RMSE	0.37

Note: The table omits coefficients for race and religion for simplicity. The logistic regression model predicts voting preferences using demographic, socioeconomic, and contextual variables. Analysis was conducted using the CES 2022 dataset.

factors on the likelihood of voting for the Democratic party (coded as 1) versus other voting choices (coded as 0). Among the significant predictors, urban residence (urbanicityUrban) shows a positive and statistically significant association (0.833, $p < 0.001$), suggesting that individuals living in urban areas are more likely to support Democratic candidates compared to those in rural areas.

Gender also plays a role, with male respondents showing a negative coefficient (-0.396, $p = 0.066$), indicating a slightly lower likelihood of voting Democratic compared to female respondents, though the effect is marginally significant. Income level demonstrates an inverse relationship, as individuals in the upper income tier are less likely to vote Democratic (-0.743, $p = 0.079$).

The model includes regional indicators, but none show statistically significant effects, suggesting limited geographic variation beyond urban-rural divides. Overall, the model achieves reasonable explanatory power with an AIC of 699.4, a BIC of 845.9, and an RMSE of 0.37, reflecting a good fit for the data. These findings emphasize the importance of urbancity, gender, and income in shaping partisan preferences.

5 Discussion

5.1 Key Findings and Implications

This analysis highlights the significant role of demographic and contextual factors in shaping voting preferences. Urbanicity emerges as a particularly strong predictor, with urban residents being significantly more likely to vote Democrat (coefficient: 0.833, $p < 0.001$). Gender differences also appear, as male respondents demonstrate a lower likelihood of Democratic support compared to female respondents, though the effect is less pronounced (coefficient: -0.396, $p = 0.066$). Income levels further illustrate partisan divides, with individuals in the upper income tier being less likely to support Democratic candidates (coefficient: -0.743, $p = 0.079$).

Interestingly, regional differences across the Northeast, South, and West show no statistically significant effects after accounting for other variables, suggesting that urbanicity and income may drive much of the regional variation observed in voter behavior. Education continues to play a key role, as individuals with four-year college degrees demonstrate significantly higher Democratic support (coefficient: 0.841, $p < 0.05$). Religion also proves influential, with Protestant and Catholic affiliations showing strong negative associations with Democratic support, reflecting established trends in religiously influenced voting behavior.

Overall, these findings underline the interplay of urbanicity, income, and religion in influencing voting outcomes. The study demonstrates the critical importance of these factors while also illustrating the secondary influence of regional and cohort differences.

5.2 Implications for Policy and Political Strategy

The results underscore the need for tailored political strategies that address the distinct preferences of urban and rural voters. Campaigns targeting urban voters should emphasize policies that resonate with diverse, younger, and more highly educated populations, as these groups are more likely to align with Democratic platforms. Conversely, addressing economic concerns and emphasizing cultural values may be key to engaging rural voters and those in higher income brackets, who lean more Republican.

The findings on religion point to the importance of faith-based outreach strategies. Engaging with religious communities through culturally sensitive messaging could help political campaigns bridge gaps with Protestant and Catholic voters. Similarly, the influence of education on voter behavior suggests that increasing access to higher education could reshape long-term political alignments, particularly in regions where educational attainment is relatively low.

Furthermore, education plays a key role in shaping political behavior, but the effects are not uniform across racial lines. For policymakers, addressing educational disparities and improving access to higher education in marginalized communities may not only reduce inequalities but also reshape voting patterns in future elections. Urban-rural divides should also be considered in political messaging, as these areas show significantly different political alignments.

In conclusion, this study underscores the importance of considering demographic variables such as race, education, and urbanicity when analyzing voting behavior. It highlights the need for political strategies that recognize and respond to the diverse and intersecting needs of voters across different demographic groups. As such, future studies should continue to explore these relationships, particularly the impact of education and race on political preferences, to ensure more inclusive and representative policies and campaigns.

5.3 Data and Temporal Limitations

A key limitation of this study lies in the temporal and demographic scope of the dataset. The analysis focuses on the 2022 Cooperative Election Study (CES), capturing data from U.S. residents regarding voting behavior during the 2022 election cycle. However, this timeframe does not encompass more recent political shifts, such as the 2024 election, or longer-term trends that might affect voting outcomes in the future. The absence of data from the subsequent election period or extended timeframes may lead to an underrepresentation of emerging patterns or the longer-term impacts of policy changes, which could influence voter behavior in future elections.

Additionally, the dataset utilized for this analysis is based on aggregated data at the state level, which could mask localized effects or intra-state disparities. Differences in political dynamics within individual states or between urban and rural areas, for example, could alter voting patterns and are not fully captured in the analysis. Future studies could benefit from more granular data, such as county-level information, to better understand how localized factors influence voting behavior. This would offer a more comprehensive understanding of the complex relationships between demographic factors, political policies, and voting preferences.

Moreover, relying solely on self-reported demographic information presents another limitation. While race, education, and other demographic factors are critical to understanding voting behavior, self-reports can be subject to bias or misinterpretation. The potential for respondents to misidentify their race or education level could skew the results, particularly for minority groups or those with non-traditional education pathways. Future research should consider supplementary qualitative approaches, such as interviews or focus groups, to address these challenges and ensure more accurate data collection methods.

5.4 Weaknesses and Future Directions

While this study provides valuable insights into how demographic factors influence voting behavior, there are several limitations that should be addressed in future research. One key limitation is the use of data from a single point in time (2022), which does not account for changes in voter preferences over time. This analysis assumes that the patterns observed during this period are static, but voting behavior can change due to shifts in political climate, policy changes, or societal events. Future research could use longitudinal data from multiple elections to examine how demographic variables affect voting behavior over time, offering a deeper understanding of long-term trends.

Additionally, the study relies on self-reported demographic data, which can be subject to biases such as misreporting or respondents' unwillingness to disclose certain information. For example, individuals may not accurately report their race, gender, or educational background, leading to potential inaccuracies in the analysis. To improve data accuracy, future research could consider using administrative data or official government records, which may provide more reliable information on demographic characteristics. Combining quantitative surveys with qualitative research methods, such as interviews or focus groups, could also provide richer insights into how people perceive and act on the demographic factors influencing their vote.

These efforts will support the development of more nuanced, evidence-based policies aimed at increasing voter engagement and addressing disparities in political representation across different demographic groups. By building on these limitations, future studies can enhance our understanding of the complex dynamics of voter behavior and contribute to more inclusive democratic processes.

Appendix

A Additional data details

A.1 Dataset and Graph Sketches

Sketches depicting both the desired dataset and the graphs generated in this analysis is available in the GitHub Repository `other/sketches`.

A.2 Data Cleaning

The CES 2022 dataset was carefully cleaned and processed to prepare it for analysis. Key variables such as `vote_choice`, `age_cohort`, `income_tier`, `education`, `gender`, `religion`, `race`, `urbancity`, and `region` were retained, reflecting the primary demographic, socioeconomic, and geographic factors relevant to voting behavior. Observations with missing or invalid values in these variables were excluded to ensure consistency and accuracy in the analysis. Additionally, only registered voters (`votereg == 1`) were included to align the dataset with the study's focus on electoral participation.

Categorical variables, such as `age_cohort`, `gender`, and `region`, were transformed into factors to enable proper handling in the logistic regression model. Continuous variables like `age` were grouped into cohorts to facilitate meaningful comparisons across age groups. The cleaned dataset was saved in both CSV and Parquet formats, ensuring compatibility with statistical tools and efficient data management. These preprocessing steps provided a robust foundation for the logistic regression analysis, enabling a comprehensive exploration of how demographic and contextual variables influence voting behavior.

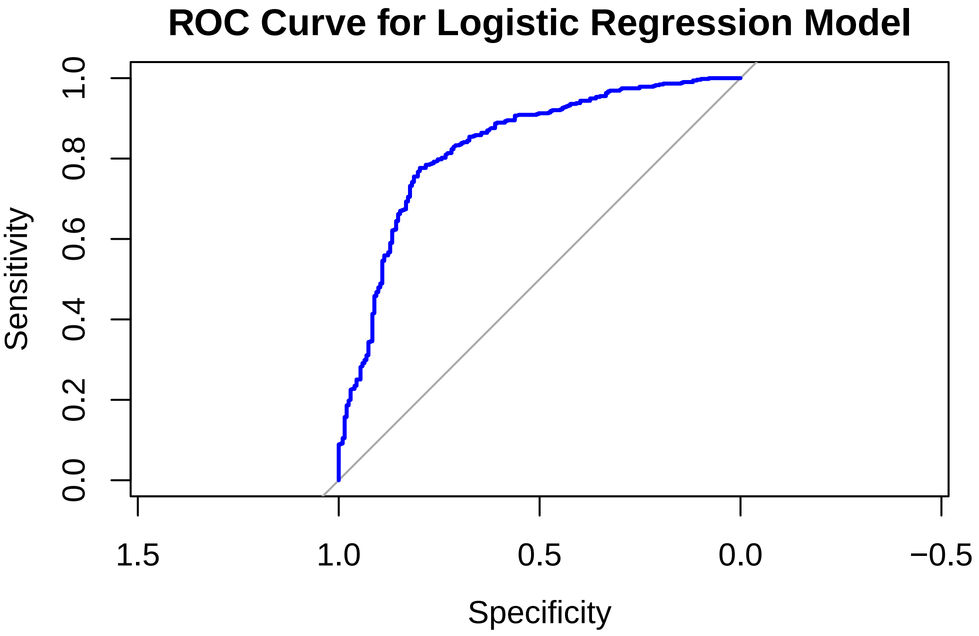
A.3 Data Source Acknowledgment

The data utilized in this study was sourced from the Harvard Dataverse. Access to the data and its use comply with the terms outlined in the Harvard Dataverse data use agreement. Specifically, the data was used for academic research purposes, with acknowledgment of the original data contributors as the source. The data contributors and Harvard Dataverse, however, do not assume responsibility for any analyses, interpretations, or conclusions drawn from the data by the authors of this study.

B Model details

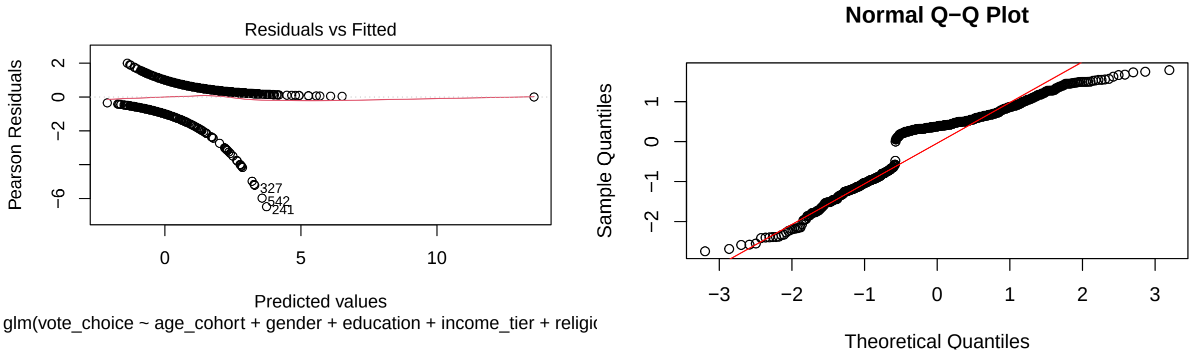
B.1 Model Validation: K-Fold Cross-Validation & ROC Curve & Log Loss

```
  parameter  Accuracy     Kappa AccuracySD    KappaSD
1      none 0.7814267 0.4368574 0.04302345 0.09376365
```



```
Log-Loss: 0.4359
```

The logistic regression model underwent 10-fold cross-validation to assess its predictive performance. The model achieved an accuracy of 0.780, indicating that it correctly classified approximately 78% of the observations. The kappa statistic was 0.429, reflecting moderate agreement between the model's predictions and the actual outcomes. The log-loss, which measures the model's calibration, was calculated as 0.436, suggesting the predicted probabilities are well-aligned with observed outcomes. The ROC curve displayed an area under the curve (AUC) of 0.779, signifying strong discriminatory ability in distinguishing between Democratic and Republican voters. These results indicate that the model provides reliable predictions, although there remains room for improvement, particularly in reducing misclassification and addressing residual variance through additional predictors or refined modeling techniques.



(a) Residual Plot

(b) Q-Q Plot

Figure 4: Diagnostics of Support for Harris model using residual vs fitted plot and norm Q-Q plot

B.2 Diagnostics

The **Residual vs. Fitted plot** (Figure 4a) shows residuals plotted against fitted values. Residuals represent the differences between observed outcomes and model predictions. Ideally, these residuals should be randomly scattered around the zero line, indicating that the model does not exhibit systematic errors. For this model, residuals appear evenly spread without a clear pattern, suggesting that the model specification is generally appropriate.

The **Q-Q Plot** (Figure 4b) evaluates how residuals align with a theoretical normal distribution. Points that align closely with the diagonal line suggest that residuals follow a normal distribution, a key assumption for interpreting model coefficients in linear regression contexts. Most residuals fall along the line, particularly in the middle range, which supports the assumption of normality. However, some deviations at the ends (outliers) indicate potential non-normality in extreme values.

Overall, these diagnostics suggest that the model performs well with minor areas for improvement, particularly regarding outlier treatment. These results enhance confidence in the model's validity while highlighting areas that may benefit from further adjustments.

C Idealized Methodology for A Survey-Based Qualitative Studies

Our study explores the relationship between political attitudes and voter behavior in the U.S., utilizing data from the Cooperative Election Study (CES) 2022. By combining observational survey analysis with targeted qualitative insights, we aim to capture both quantitative trends and nuanced voter experiences. While the CES dataset provides a comprehensive framework to assess demographic, geographic, and ideological patterns, qualitative surveys supplement these findings by examining factors such as voter motivations, information sources, and decision-making processes. This combined approach enhances our understanding of systemic influences and personal perspectives, enabling more robust interpretations of observed trends and informing evidence-based political strategies and policies.

C.1 Introduction

This appendix outlines the methodology used to analyze the 2022 Cooperative Election Study (CES) dataset. The CES provides a rich source of survey data encompassing political attitudes, behaviors, and demographic characteristics of U.S. residents. This methodology ensures a systematic approach to data preparation, analysis, and interpretation, facilitating robust and replicable results.

Our survey focuses on individuals across different socioeconomic and demographic groups to understand how economic conditions, educational attainment, and racial identity shape voting behavior and political preferences. By integrating quantitative data from the 2022 Cooperative Election Study (CES) with qualitative insights, this study aims to uncover the nuanced interplay of economic realities and demographic characteristics. The findings will inform evidence-based interventions, addressing systemic inequities in political participation and advancing our understanding of how diverse lived experiences influence electoral outcomes.

C.2 Objective

The objective of this study is to investigate the socioeconomic, demographic, and systemic factors shaping voter behavior and political preferences in the United States. By focusing on the interplay between economic conditions, educational attainment, and racial identity, the study aims to uncover how these factors influence voting decisions and participation. Understanding how economic realities and demographic traits interact to affect political behavior is crucial for addressing systemic barriers to equitable representation and democratic engagement. The findings will guide the development of evidence-based strategies to enhance voter participation, promote inclusive political representation, and inform policies that address socioeconomic disparities in the electorate.

C.3 Sampling Approach

In this analysis, we use matched random sampling, a methodology employed by the 2022 Cooperative Election Study (Schaffner, Ansolabehere, and Shih (2023)) through YouGov. This method ensures a representative sample of the U.S. population by matching respondents from an opt-in panel to a target sample drawn from demographic benchmarks such as the American Community Survey (ACS). The matched random sampling approach is particularly effective for large-scale studies, leveraging statistical adjustments to mitigate biases and improve sample representativeness.

Matched random sampling is ideal for our study as it allows us to capture a diverse range of socioeconomic and demographic characteristics across the electorate. This approach ensures that key variables, such as education, income, race, and urbanicity, are well-represented in the dataset, enabling robust analysis of the interplay between these factors and voting behavior. By combining rigorous sampling techniques with advanced weighting methods, we enhance the reliability and validity of our findings, ensuring that they reflect the broader U.S. electorate. These strengths make matched random sampling an ethical and effective method for examining the systemic, economic, and demographic factors shaping political preferences and participation.

C.4 Target Population

Our target population comprises U.S. voters across diverse socioeconomic, educational, and racial demographics, as captured in the 2022 Cooperative Election Study (Schaffner, Ansolabehere, and Shih (2023)). Specifically, we focus on individuals whose voting behavior is shaped by their economic conditions, educational attainment, and racial identity. This includes voters from various states, regions, and urbanicity levels to ensure a comprehensive analysis of how these factors interact to influence political preferences and participation.

C.5 Sample frame

The sample frame for this study is derived from the 2022 Cooperative Election Study (CES) dataset, which includes responses from 60,000 individuals recruited through YouGov's matched random sampling methodology. The CES sample frame utilizes a politically representative modeled frame based on the 2019 American Community Survey (ACS), voter registration files, and demographic data such as age, gender, race, education, and region.

This sample frame ensures coverage of diverse demographic and socioeconomic groups across the United States, enabling the study to examine how variables such as economic conditions, educational attainment, and racial identity interact to influence voting behavior. The sample frame is specifically structured to include a wide range of political preferences and behaviors, providing the representativeness necessary to draw meaningful insights about the interplay between systemic factors and voter participation.

C.6 Sample

We aim to survey 1,000 respondents who meet our defined sample criteria: eligible U.S. voters representing diverse racial, economic, and educational backgrounds. The sample will specifically target individuals from varying income levels, educational attainment, and racial groups to explore the interplay of these factors in shaping voting behavior. Participation will be voluntary, with respondents required to answer survey questions truthfully and comprehensively to ensure data quality and depth.

C.7 Recruitment of Respondents

To recruit participants for this study, we will use a stratified sampling approach in collaboration with an online survey platform, such as YouGov, which specializes in matched random sampling to ensure representativeness. The initial recruitment will focus on demographic and geographic diversity, ensuring proportional representation of racial groups, income brackets, and urbanicity levels.

Outreach materials will emphasize the study's purpose of understanding socioeconomic influences on voting behavior while assuring respondents of confidentiality and anonymity. Eligible participants will be invited to complete a screening survey to ensure alignment with the study's inclusion criteria. The survey will be distributed online for accessibility and convenience, with respondents able to participate from any location while maintaining their privacy. To encourage participation, respondents will receive modest compensation, further ensuring meaningful engagement with the study.

This approach enables the collection of a representative dataset, capturing the nuances of economic and demographic factors influencing voter behavior.

C.8 Handling Non-response bias

Non-response bias is a critical concern in survey research, particularly when studying voter behavior and socioeconomic factors. Participants who do not respond or drop out may differ significantly from those who complete the survey, leading to skewed conclusions. To address this, we will emphasize the importance of the study, ensure anonymity, and provide a straightforward, user-friendly survey experience that takes approximately 5–10 minutes to complete. Outreach efforts will also include reminders and incentives, such as modest compensation, to encourage higher participation rates, particularly among underrepresented groups.

C.9 Respondent Validation

To ensure the reliability and credibility of the collected data, we will implement a rigorous respondent validation process. Eligibility screening questions will verify participants' age, voting eligibility, and demographic characteristics such as income, education, and racial identity. Responses will be reviewed for completeness, logical consistency, and alignment with inclusion criteria. Additionally, weights will be applied to adjust for imbalances in demographic representation, ensuring the data accurately reflects the broader U.S. electorate. By leveraging a reputable survey platform like YouGov and integrating advanced quality checks, we aim to maintain data integrity and draw meaningful, representative insights.

C.10 Ethical Concerns

This study involves exploring sensitive topics such as voting behavior, socioeconomic disparities, and demographic influences, necessitating an ethical framework to safeguard participants' privacy and ensure fairness. Recognizing the potential discomfort participants might experience when sharing personal information about their economic conditions, political preferences, or demographic characteristics, the survey will provide clear explanations of its purpose and allow participants to skip questions or withdraw at any time without consequences.

Strict confidentiality measures will be in place to protect participants' identities, with responses securely stored and anonymized to prevent re-identification. Recruitment through reputable platforms and organizations will foster trust, and respondents will be informed of their rights throughout the process. Additionally, transparency in reporting and ethical data usage will ensure the findings are used responsibly to advance understanding without perpetuating harm or bias. This ethical framework underscores our dedication to conducting inclusive, respectful, and socially responsible research.

C.11 Proposed Survey Design

Exploring the relationship between education, economic conditions, and voter behavior requires a carefully crafted survey design to ensure accurate, unbiased data collection. Social desirability bias and perceived judgment in discussing socioeconomic and political preferences pose challenges, particularly in a politically polarized environment. Drawing on best practices from Stantcheva (2023) and similar survey research, this survey employs design strategies to minimize bias, maximize respondent comfort, and enhance data accuracy.

This survey examines how economic conditions, education, and racial identity influence voting preferences, emphasizing neutral, inclusive phrasing and an anonymous, online format to reduce respondent concerns. Inspired by successful methodologies, it incorporates strategies like randomized response options, opt-out choices (e.g., "Prefer not to say"), and a balanced mix of multiple-choice and open-ended questions to capture nuanced perspectives. By adopting a "contribution" framework, the survey introduces sections with messages highlighting the importance of participant input in improving societal understanding and public policy. This approach aims to foster trust, engagement, and honest responses, ensuring the collection of high-quality data that reflects the diverse realities of voter experiences.

C.12 Solution to the response bias in our survey

We draw on recommendations from Stantcheva (2023) to minimize response biases. Common response biases identified in survey design include moderacy bias, extreme response bias, ordering bias, acquiescence bias, experimenter demand effect(EDE), and social desirability bias (SDB). Our survey primarily focuses on strategies to reduce moderacy bias, extreme response bias, ordering bias, and SDB. The detailed definitions of these biases are provided in **?@sec-definition**.

To mitigate bias, we enhance our survey in the following ways, drawing on recommendations from Stantcheva (2023):

- Addressing Extreme/Moderacy Bias: We use a minimum of five response options for scale questions to provide more detailed choices, reducing the likelihood of respondents defaulting to extreme or middle answers.
- Mitigating Response Order Bias: For nominal questions, we randomize response options.
- Minimizing Social Desirability Bias (SDB): The survey design addresses social desirability bias (SDB) by emphasizing anonymity and confidentiality throughout. A clear introduction outlines the survey's purpose—academic research on the impact of restrictive abortion laws—and reassures participants that their responses will remain confidential and solely used for research. The anonymous online format creates a safe environment for participants to share their experiences without fear of judgment or stigma. A feedback section at the end encourages participants to express concerns or share additional thoughts, fostering trust and enhancing data quality.

C.12.1 Survey Link

The survey has been implemented using Google Forms. You can access it here: Survey Link.

C.13 Copy of Survey on Restrictive Abortion Laws and Maternal Health

Welcome Section

Introduction: Welcome to our study on the impact of restrictive abortion laws on maternal and infant health. Your participation in this survey will help us understand the psychological, social, and systemic impacts of these laws. Rest assured that your responses are anonymous and will only be used for academic research purposes.

This survey is conducted by nonpartisan researchers in public health and social sciences. It consists of 17 carefully designed questions and should take approximately 10–15 minutes to complete.

Please answer the questions honestly. If you experience any discomfort while completing the survey, you may stop at any time. For support, we provide a list of mental health resources at the end of the survey.

Contact Information: Jinyan Wei Email: jinyan.wei@mail.utoronto.ca

Section 1: Demographics and Background Information

1. What is your age?
   - Under 18
   - 18–24
   - 25–34
   - 35–44
   - 45–54
   - 55+
   - Prefer not to say
2. What is your highest level of education?
   - Less than high school
   - High school graduate or equivalent
   - Some college
   - Bachelor's degree
   - Graduate or professional degree
   - Prefer not to say
3. What is your marital status? - Single - Married - Divorced - Widowed - Prefer not to say

Section 2: Understanding Abortion Experiences

Introduction: This section focuses on understanding abortion experiences, including the circumstances and decisions surrounding them. Your responses are invaluable in helping researchers and policymakers improve health services and support for women and families. Please know that your answers are entirely confidential and will only be used for research purposes. If you are comfortable, we encourage you to answer as honestly as possible. If you prefer not to answer, you are welcome to skip this section.

1. Are you willing to answer this part?
   - Yes
   - No [Jump to Section 3]
2. Did you seek abortion services during your pregnancy?
   - Yes
   - No

- Prefer not to say

3. If yes, were you unable to access abortion services due to legal restrictions in your state?
   - Yes
   - No
   - Prefer not to say
4. How did the inability to access abortion services impact your mental health during pregnancy?
   - No impact
   - Mild impact
   - Moderate impact
   - Severe impact
   - Prefer not to say
5. Did you receive any support from healthcare providers or community organizations during your pregnancy?
   - Yes
   - No
6. If you were unable to access abortion services, what barriers did you encounter? (Select all that apply)
   - Legal restrictions
   - Financial constraints
   - Lack of healthcare providers
   - Distance to clinic
   - Fear of stigma or judgment
   - Prefer not to say
7. If you sought an abortion but were unable to access one, how did this affect your mental health during pregnancy?
   - Increased stress
   - Anxiety
   - Depression
   - Anger or frustration
   - Feeling of helplessness
   - Prefer not to say

Section 3: Experiences and Support for Women Facing Pregnancy Challenges

Introduction: This section seeks to understand the experiences and outcomes of women who, due to restrictive abortion laws, were unable to access abortion services and subsequently faced the loss of their child. We recognize that discussing past pregnancies can be sensitive, especially those involving circumstances such as abortion or the loss of a child. Your understanding are invaluable in identifying areas where healthcare and support services can be improved. If you prefer not to answer, you are welcome to skip this section.

1. Are you willing to answer this part?
   - Yes
   - No [Jump to Section 4]
2. Did you experience the loss of your child within the first year of life?
   - Yes
   - No [Jump to Section 4]
   - Prefer not to say
3. What was the primary cause of your child's death as communicated by healthcare providers?
   - Premature birth or related complications
   - Congenital abnormalities or genetic conditions
   - Sudden Infant Death Syndrome (SIDS)
   - Infections (e.g., pneumonia, sepsis)
   - Birth trauma or delivery complications
   - Lack of access to timely medical care
   - Other (please specify)
4. To what extent do you believe the instances during your pregnancy where mental health challenges prevented you from seeking or receiving adequate medical care?
   - 1 (No impact)
   - 2
   - 3
   - 4
   - 5 (Significantly)
5. To what extent do you believe the mental health challenges caused by the inability to access abortion services contributed to health complications for your child?
   - Not at all – I don't believe my mental health challenges had any impact on my child's health.
   - Slightly – I think there may have been a minor impact on my child's health.
   - Moderately – I feel my mental health challenges had a noticeable impact on my child's health.
   - Significantly – I believe my mental health challenges had a considerable impact on my child's health.
   - Completely – I think my mental health challenges were the primary factor in my child's health complications.

- Prefer not to say

6. Looking back, do you believe access to abortion services could have positively affected your mental and physical health during pregnancy?
    - Yes
    - No
    - Maybe
    - Prefer not to say

Section 4: Perspectives and Support

If you feel distressed or need support after completing this survey, the following resources are available to provide assistance:

1. Postpartum Support International (PSI)
    - Website: www.postpartum.net
    - Helpline: 1-800-944-4773 (Text "Help" to 800-944-4773)
    - Services: Support for mental health during and after pregnancy, including peer support and counseling.
2. Mental Health America (MHA)
    - Website: www.mhanational.org
    - Services: Online screening tools, support networks, and educational materials.
3. SAMHSA National Helpline
    - Website: www.samhsa.gov
    - Hotline: 1-800-662-HELP (4357)
    - Services: Free, confidential referrals for mental health and substance use disorders.

Section 5: Feedback

1. Do you have any concerns or feedback regarding the survey, surveyor, or entity?
    - Your feedback is important to us and will help ensure transparency and trust in the research process.

Section 6: Thank You

Thank you for taking the time to complete this survey. Your honest feedback is invaluable and will help us better understand and address the experiences of women who have faced similar circumstances. We deeply appreciate your participation and the courage it takes to share your experiences.

References

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

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

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

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

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

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. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.