The Educational Divide: Understanding the Interaction Between Educational Attainment and Racial Identity in American Voting Patterns*

A Multinomial Analysis of the Cooperative Election Study Common Content Dataset

Jinyan Wei

November 29, 2024

This study examines the impact of the Dobbs decision and subsequent abortion bans on infant mortality rates in the United States, focusing on data from 2021 to 2022. Using a Difference-in-Differences (DID) approach, we analyze trends in states where abortion remained legal compared to those where it became illegal after June 2022. Our results indicate that states enforcing abortion bans experienced an increase of 0.285 infant deaths per 1,000 live births, with compounded effects observed during the post-injunction period. Maternal age and race also emerged as significant predictors, with Black mothers facing disproportionately higher infant mortality rates. These findings underscore the compounded public health implications of restrictive abortion policies and the urgent need for targeted interventions to address racial and demographic disparities. By utilizing recent, high-quality data and rigorous modeling, this study provides critical insights into the intersection of reproductive rights and infant health outcomes.

Table of contents

1 Introduction **2**

2 Data **3**

2.1 Overview 3

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

2.2 Measurement . . . 4
2.3 Variables . . . 5

3 Model **8**
3.1 Model Parameters . . . 8
3.2 Model Set-Up . . . 9
3.2.1 Model justification . . . 10

4 Results **11**
4.1 Model Results . . . 11

5 Discussion **12**
5.1 Key Findings and Implications . . . 12
5.2 Racial and Socioeconomic Disparities . . . 12
5.3 Addressing Public Health and Long-Term Implications of Restrictive Abortion Laws . . . 13
5.4 Data and Temporal Limitations . . . 13
5.5 Weaknesses and Future Directions . . . 13

Appendix **15**

A Additional data details **15**
A.1 Dataset and Graph Sketches . . . 15
A.2 Data Cleaning . . . 15

B Model details **15**
B.1 Diagnostics . . . 15

References **16**

1 Introduction

The relationship between educational attainment and voting behavior has long been a subject of scholarly interest, particularly as the American electorate becomes increasingly diverse and polarized. Understanding how education levels interact with racial identity to shape political preferences is crucial for comprehending contemporary voting patterns and their implications for democratic participation.

This study examines the complex interplay between educational attainment and racial demographics in determining voting choices, using comprehensive data from the 2022 Cooperative Election Study (CES). The CES dataset, comprising 60,000 respondents through matched random sampling by YouGov, provides detailed insights into voter preferences, demographic characteristics, and political behavior across all U.S. states. Our analysis addresses a critical gap in the literature by investigating how the impact of education on voting decisions varies across racial groups, accounting for socioeconomic differences and access to opportunities.

Using a multinomial logistic regression framework, we analyze how educational attainment interacts with racial identity to influence voting patterns, while controlling for key demographic and socioeconomic factors including gender, urbanicity, gun ownership, and education loan status. The model incorporates both direct effects and interaction terms to capture the nuanced relationships between these variables and voting preferences. This methodological approach allows us to estimate the differential impacts of education across racial groups while accounting for other important determinants of political behavior.

The findings indicate that the relationship between educational attainment and voting choice varies significantly across racial groups, suggesting that the political implications of education are not uniform across demographic categories. This research contributes to our understanding of contemporary American electoral behavior by providing current evidence on how educational experiences and racial identity jointly shape political preferences. These insights have important implications for understanding voter behavior, demographic voting patterns, and the broader dynamics of political participation in an increasingly diverse electorate.

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 **?@sec-survey**.

2 Data

2.1 Overview

We conduct our voting pattern analysis using the R programming language (R Core Team 2023), leveraging data from the 2022 Cooperative Election Study (CES) (Schaffner, Ansolabehere, and Shih 2023). The CES dataset, constructed through matched random sampling by YouGov, offers a comprehensive overview of voter preferences, demographic details, and political behavior in the United States. The survey, designed to ensure representation at both state and national levels, incorporates pre-election and post-election responses from 60,000 participants. This robust dataset enables the analysis of voting behavior, party identification, policy preferences, and their interactions with demographic variables.

In this study, we utilized several R packages to enhance data manipulation, modeling, and visualization capabilities. The `tidyverse` package (Wickham et al. 2019) provided a suite of tools for efficient data wrangling and exploration, while `arrow` (Richardson et al. 2024) facilitated the management of large parquet files, ensuring seamless data input and output. The `janitor` package (Firke 2023) supported data cleaning by offering tools to identify and rectify quality issues within the dataset, and `lubridate` (Grolemund and Wickham 2011) simplified the handling of date-time variables. To streamline file organization and improve reproducibility, we relied on the `here` package (Müller 2020).

For modeling voting patterns, the `nnet` package (**nnet?**) was used to fit multinomial logistic regression models, capturing the influence of demographic factors and their interactions on vote choice. Results were summarized and visualized using the `modelsummary` package (Arel-Bundock 2022) and `ggplot2` (Wickham 2016), respectively, ensuring clarity and interpretability. Additionally, we employed `knitr` (Xie 2014) and `kableExtra` (Zhu 2024) to generate clean, reproducible tables and reports. Together, these packages supported an efficient and transparent workflow for analyzing the CES dataset, enabling a deeper understanding of the complex relationship between race, education, and voting behavior.

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 education 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

Our analysis focuses on the following variables, with a specific emphasis on `vote_choice` as the dependent variable:

- `vote_choice`: The voting preference of individuals, serving as the dependent variable in our analysis. The categories include:
  - **Democrat**: Individuals who voted for the Democratic candidate.
  - **Republican**: Individuals who voted for the Republican candidate.
  - **Independent**: Individuals who voted for a third-party or independent candidate.
  - **Did not vote**: Individuals who did not participate in voting.
- `education`: The highest level of education completed by individuals. Categories include:
  - **High school or less**: Individuals with a high school diploma or lower education level.
  - **Some college**: Individuals with some college education but no degree.
  - **College graduate**: Individuals who have completed a bachelor's degree.
  - **Postgraduate**: Individuals with education beyond a bachelor's degree.
- `race_group`: The racial or ethnic group individuals identify with. Categories include:
  - **White**: Individuals identifying as White.
  - **Black**: Individuals identifying as Black or African American.
  - **Hispanic**: Individuals identifying as Hispanic or Latino.
  - **Asian**: Individuals identifying as Asian.
  - **Other**: Individuals identifying with any other racial or ethnic group.
- `gender`: The self-reported gender of individuals, categorized as:
  - **Male**
  - **Female**
- `urbancity`: The type of community individuals reside in, categorized as:
  - **Urban**: Dense metropolitan areas.
  - **Suburban**: Residential areas near urban centers.
  - **Rural**: Sparsely populated regions or countryside.
- `gunown`: Indicates whether an individual owns a gun, categorized as:
  - **Yes**
  - **No**
- `edloan`: Indicates whether an individual has education loans, categorized as:
  - **Yes**
  - **No**

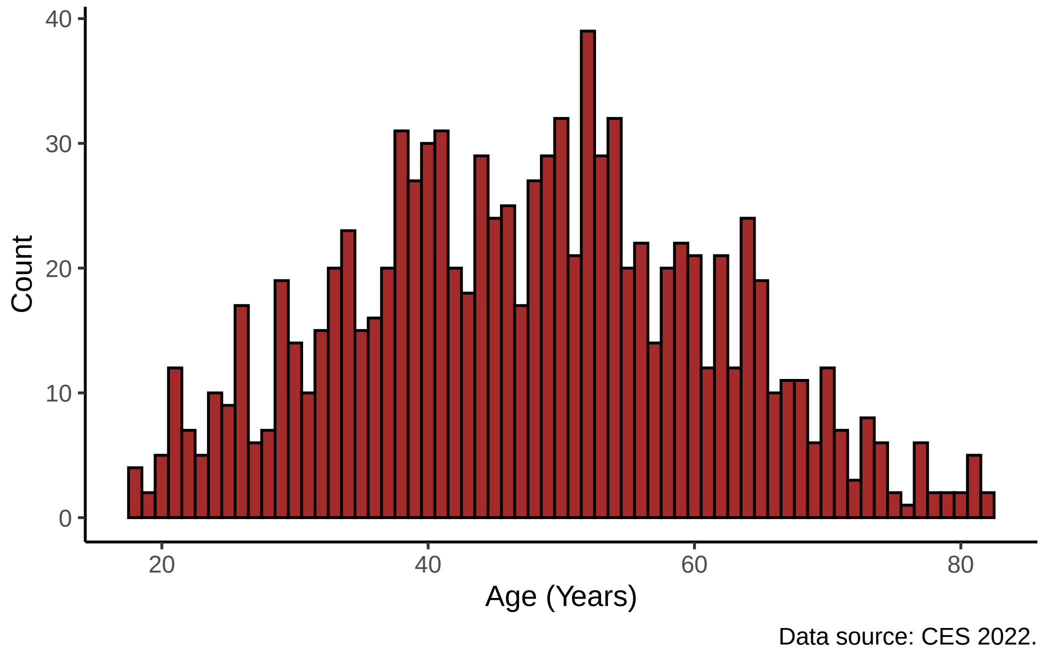

Figure 1: Distribution of Respondent Age

- `age`: The age of the individual, recorded as a continuous variable.

Figure 1 shows the distribution of respondent ages, with most individuals concentrated between the ages of 30 and 60. The distribution appears relatively symmetric, with no extreme outliers, suggesting that the sample covers a broad range of adult ages. This balance ensures that the analysis includes perspectives from various life stages.

Figure 2 shows the distribution of respondents' voting choices across different age groups, categorized by race. The visualization highlights three main categories of vote choice: Republican, Democrat, and Other. Each point represents an individual, with its position along the x-axis corresponding to their age and its color indicating their race group. This graph illustrates how voting preferences may vary across demographics, showing a significant spread for both Republican and Democrat voters across all ages, while the "Other" category remains sparsely populated. Observing the clustering patterns within race groups can provide insights into potential demographic trends influencing voting behavior.

Figure 3 visualizes the distribution of educational attainment across racial and ethnic groups, highlighting disparities that may influence political behavior. "College graduate" is the largest educational category across most groups, with a particularly high proportion among Middle Eastern (64%) and Other (63%) populations. Meanwhile, "High school or less" is notably prevalent among Native American (33%) and Hispanic (20%) respondents, suggesting barriers to higher education for these groups. The graph underscores how the interaction between race and education shapes opportunities and outcomes, making it a crucial factor for analyzing

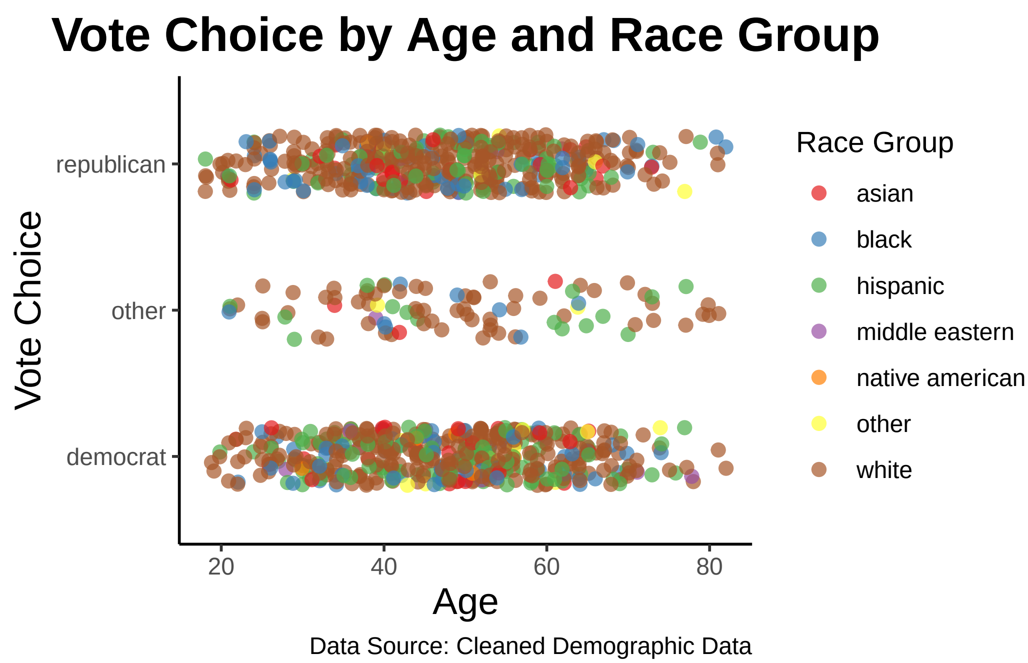

Figure 2: Distribution of Vote Choice: The distribution of respondents' voting choices in the dataset highlights that the majority voted for either the Republican or Democrat candidates, with a smaller proportion selecting other candidates.

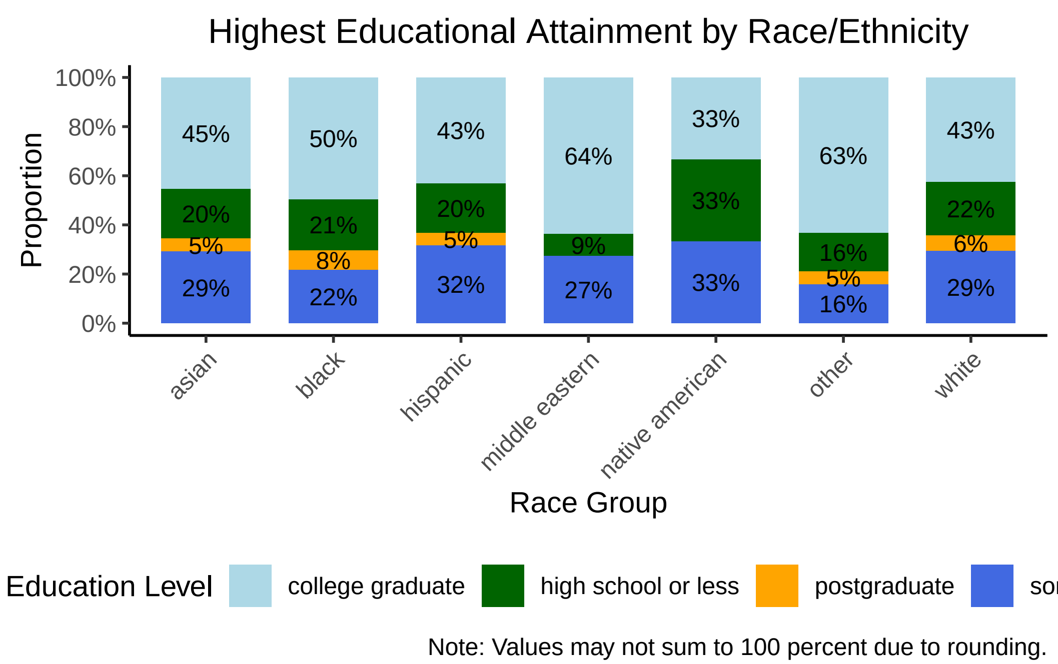

Figure 3: Distribution of Respondent Age: The age distribution shows a peak around middle-aged respondents, with a relatively symmetric spread.

voting behavior. For instance, research like (**enos2018?**) emphasizes that education's political effects are context-dependent, moderated by local demographics and historical inequalities, further supporting the need to explore this interaction in understanding voting trends.

3 Model

Our modeling approach aims to quantify the relationship between demographic, contextual, and behavioral factors and individuals' voting choices. For this analysis, we employ a multinomial logistic regression (MLR) model to predict the probability of selecting each voting category (`republican`, `democrat`, or `other`) based on a set of predictors. The model is implemented using the `multinom` function in R, which estimates the log-odds of each outcome relative to a reference category.

In this analysis, we use predictors that capture a broad range of individual characteristics and contextual factors influencing voting behavior. Specifically, we include variables such as `gender`, representing the respondent's self-reported gender; `education`, capturing the highest level of education attained; and `race_group`, representing the respondent's racial or ethnic background. Additional predictors include `urbancity`, reflecting the type of community (urban, suburban, or rural) where the respondent resides; `gunown`, indicating firearm ownership; and `edloan`, indicating the presence of education loans. Finally, we incorporate `age` as a continuous variable.

To account for interactions between education and race, an extended version of the model includes an interaction term (`education:race_group`). This term allows us to assess how the relationship between education and voting choice varies across racial groups, offering a more nuanced understanding of these predictors' combined effects.

The multinomial logistic regression assumes that the relative log-odds of any two outcomes are linearly related to the predictors, and the independence of irrelevant alternatives (IIA) assumption holds. This allows us to model the factors influencing voting preferences in a consistent and interpretable framework.

3.1 Model Parameters

The model predicts the probability of voting for each category (`republican`, `democrat`, or `other`) by including the following predictor variables:

- **Gender (`gender`)**: Categorical variable indicating the respondent's self-reported gender (`male` or `female`).
- **Education (`education`)**: The highest level of education attained, categorized as `high school or less`, `some college`, `college graduate`, or `postgraduate`.

- **Race Group (race_group)**: Categorical variable representing the respondent's racial or ethnic background (White, Black, Hispanic, Asian, or Other).
- **Urbancity (urbancity)**: The type of community where the respondent resides (urban, suburban, or rural).
- **Gun Ownership (gunown)**: Binary variable indicating whether the respondent owns a firearm (Yes or No).
- **Education Loan (edloan)**: Binary variable indicating whether the respondent has outstanding education loans (Yes or No).
- **Age (age)**: Continuous variable representing the respondent's age.

In the second model, an **interaction term** between education and race_group is included to assess the combined effect of these predictors on voting preferences:

- **Interaction Term (education:race_group)**: Captures how the effect of education on voting choice varies across racial groups.

We include variables like urbancity, gunown, and edloan as controls to account for contextual and behavioral factors that could influence voting behavior. By modeling these relationships, we aim to provide a balanced and detailed assessment of how demographic and behavioral factors shape voting preferences.Background details and diagnostics are included in Appendix B.

3.2 Model Set-Up

In our analysis, we employ two multinomial logistic regression models to investigate voting patterns. The first model includes primary predictor variables such as gender, education, race group, urbanicity, gun ownership, education loan status, and age. These variables capture key demographic and socioeconomic factors influencing voting choices. The second model extends the first by introducing an interaction term between education and race group to account for potential joint effects of these variables on voting preferences.

Define y_i as the voting choice for individual i, categorized as "Republican," "Democrat," or "Other." The probability $P(y_i = k)$ of an individual i choosing category k is modeled as:

$$
\begin{aligned}
y_i|\eta_{ik}, \sigma &\sim \text{Multinomial}(P(y_i = k)) \\
P(y_i = k) &= \frac{\exp(\eta_{ik})}{\sum_{j=1}^{K} \exp(\eta_{ij})}
\end{aligned}
$$

$$
\begin{aligned}
\text{First Model} : \eta_{ik} &= \beta_{0k} + \beta_{1k} \cdot \text{Gender}_i + \beta_{2k} \cdot \text{Education}_i + \beta_{3k} \cdot \text{Race Group}_i \\
&+ \beta_{4k} \cdot \text{Urbancity}_i + \beta_{5k} \cdot \text{Gun Ownership}_i + \beta_{6k} \cdot \text{Education Loan}_i + \beta_{7k} \cdot \text{Age}_i
\end{aligned}
$$

$$
\begin{aligned}
\text{Second Model} : \eta_{ik} &= \beta_{0k} + \beta_{1k} \cdot \text{Gender}_i + \beta_{2k} \cdot \text{Education}_i + \beta_{3k} \cdot \text{Race Group}_i \\
&+ \beta_{4k} \cdot \text{Urbancity}_i + \beta_{5k} \cdot \text{Gun Ownership}_i + \beta_{6k} \cdot \text{Education Loan}_i + \beta_{7k} \cdot \text{Age}_i \\
&+ \beta_{8k} \cdot (\text{Education}_i \times \text{Race Group}_i)
\end{aligned}
$$

The model is executed in R using the `multinom` function from the **`nnet`** package, which fits the multinomial logistic regression model.These models allow us to examine both the independent and combined effects of education and race on voting patterns while controlling for other demographic and socioeconomic factors.

3.2.1 Model justification

Scholars in political science have established that demographic factors such as gender, education, race, and urbanicity, along with contextual variables like firearm ownership and education loans, play significant roles in shaping voting behavior. Furthermore, the interaction between `education` and `race_group` is well-supported in the literature, as studies have shown that the impact of education on voting choice often varies by race due to differences in socioeconomic context and access to opportunities. Including this interaction term allows us to better capture these nuanced relationships and improve the interpretability of the model.

We fit a frequentist multinomial logistic regression model, which is appropriate for categorical outcomes with more than two levels. Coefficients are estimated using maximum likelihood estimation, offering interpretable results without relying on prior assumptions. Unlike Bayesian methods, no priors are applied. Additionally, we do not apply weights to the observations, assuming the dataset accurately reflects the underlying population. Future analyses, however, could incorporate weighting schemes to address demographic imbalances or over- or under-representation of specific voting categories.

This model framework is well-suited for understanding the effects of multiple predictors and their interactions on voting choice. Each coefficient represents the change in the log-odds of selecting a given voting category relative to the reference category (`republican`). By including main effects and interaction terms, the model provides a theoretically informed and methodologically sound approach to analyzing voting preferences. Diagnostics such as multicollinearity checks and assessments of the independence of irrelevant alternatives (IIA) assumption further ensure the reliability of the results.

4 Results

4.1 Model Results

The results from the two Bayesian models are summarized in **?@tbl-modelresults**. The interaction term between after_injunction and abortion_illegal is positive and statistically significant (0.111 in first model and 0.280 in second model), indicating a change in infant mortality rates associated with the combination of abortion bans and the post-injunction period.

The results from the multinomial logistic regression models are summarized in **?@tbl-modelresults**. The first model examines the relationship between demographic and socioeconomic factors and vote choice, while the second model includes interaction terms between education and race group to explore their combined effects.

For the main effects, gender shows significant associations with vote choice. Non-binary individuals are more likely to vote Democrat compared to Republicans in both models, with a coefficient of 0.895 in the first model. Education also demonstrates varying impacts; individuals with postgraduate education are less likely to vote Democrat compared to those with high school education in Model 1, but this trend changes when interacting with race group in Model 2. For instance, Black voters with a high school education are more inclined to vote Democrat, highlighting the nuanced influence of race and education on voting behavior.

Race group plays a critical role in shaping vote choice. Hispanic voters are positively associated with voting for "other" candidates, with a coefficient of 0.555 in the first model, reflecting a potential preference for non-mainstream parties. Similarly, urbanicity shows a strong correlation with vote choice; rural voters are less likely to vote Democrat, while suburban and urban voters lean more toward Democrats.

The second model's interaction terms reveal significant insights into how education and race group intersect. For example, the interaction between postgraduate education and Black voters suggests a greater likelihood of voting Democrat, while Hispanic voters with high school education are more inclined to vote for other candidates.

The models performed well, with the residual deviance decreasing from 1825.802 in the first model to 1795.585 in the second, indicating improved fit with the inclusion of interaction terms. The AIC values further confirm this trend, decreasing from 1897.802 to 1931.585. These results underscore the complex interplay of demographic and socioeconomic factors in influencing vote choice, while highlighting the importance of considering intersectional effects for a more nuanced understanding.

5 Discussion

5.1 Key Findings and Implications

The analysis reveals notable patterns in infant mortality rates influenced by policy and demographic factors. For states where abortion remained legal, the post-injunction period (after_injunction1) is associated with reductions in infant mortality rates, with a decrease of 0.525 deaths per 1,000 live births in the first model and 0.947 in the second model, compared to the pre-injunction period. In contrast, states where abortion became illegal (abortion_illegal1) experienced significantly higher infant mortality rates, with increases of 3.365 deaths in the first model and 1.401 in the second model, relative to states where abortion remained legal.

The Difference-in-Differences (DID) interaction term (after_injunction1 × abortion_illegal1) indicates that abortion bans during the post-injunction period resulted in an additional increase of 0.111 deaths per 1,000 live births in the first model and 0.280 in the second model. These effects highlight the compounded impact of restrictive abortion policies on infant health outcomes.

Demographic factors also play a significant role. Compared to mothers aged 15–19 years (reference level), older maternal age groups (20–24, 25–29, and 30–34 years) are associated with progressively lower infant mortality rates, with reductions of 3.571, 5.732, and 6.155 deaths per 1,000 live births in the second model, respectively. Regarding maternal race, Black or African American mothers have a significantly higher infant mortality rate, with an increase of 6.604 deaths per 1,000 live births compared to Asian mothers (reference level). In contrast, White mothers exhibit a lower infant mortality rate, with a decrease of 2.562 deaths per 1,000 live births compared to Asian mothers.

These findings underscore the interplay of policy and demographic factors, emphasizing the disproportionate burden of restrictive abortion policies and systemic inequities on infant mortality rates.

5.2 Racial and Socioeconomic Disparities

The results reveal stark racial disparities in infant mortality rates, with Black mothers experiencing significantly higher rates compared to other racial groups. These disparities are exacerbated by restrictive abortion policies, which disproportionately affect already vulnerable populations. The intersection of systemic healthcare inequities, socioeconomic challenges, and restrictive legislation creates compounded risks for marginalized communities, leading to worse health outcomes. Addressing these inequities requires targeted policies that ensure equitable access to maternal and infant healthcare, particularly in communities most impacted by these disparities. Such policies should prioritize investments in healthcare infrastructure, culturally competent care, and programs that support at-risk populations.

5.3 Addressing Public Health and Long-Term Implications of Restrictive Abortion Laws

Addressing Public Health and Long-Term Implications of Restrictive Abortion Laws

Restrictive abortion policies significantly impact maternal and infant health, as evidenced by an increase of 0.285 infant deaths per 1,000 live births in states enforcing abortion bans during the post-injunction period. These policies exacerbate systemic issues such as inadequate prenatal care, gaps in healthcare access, and socioeconomic inequities, disproportionately affecting marginalized communities. Over time, the consequences may include higher rates of untreated maternal mental health issues, delayed or insufficient prenatal care, and elevated infant mortality, creating lasting health challenges for families and communities. Addressing these disparities requires targeted interventions, such as expanding maternal mental health services, improving prenatal and postnatal care in underserved areas, ensuring culturally competent healthcare, and advocating for equitable policies. Investments in community-based support systems, enhanced healthcare delivery, and robust data infrastructure are essential to monitor and mitigate the long-term impacts of these policies, fostering sustainable improvements in maternal and infant health outcomes.

5.4 Data and Temporal Limitations

A key limitation of our analysis is the temporal and demographic scope of the dataset. The data includes only infant deaths under 1 year of age occurring within the United States to U.S. residents during 2021–2022. This timeframe does not capture the full year of infant mortality outcomes following the Dobbs decision, as infants who died in 2023 are not included. Furthermore, the dataset's focus on infants under 1 year of age may exclude cases that reflect longer-term health impacts. These limitations could lead to an underestimation of the full effects of abortion bans. As no publicly available dataset currently provides complete 2023 data or longer-term health outcomes, future research should incorporate more detailed and updated datasets to provide a clearer and more complete analysis of the long-term trends and implications of restrictive abortion policies.

5.5 Weaknesses and Future Directions

While this study provides meaningful findings, several limitations should be addressed in future research. The analysis relies on aggregate state-level data, which may obscure localized effects and intra-state disparities. Unmeasured confounders, such as economic conditions, healthcare infrastructure differences, or state-specific policies, could also influence the results. Moreover, the short timeframe of the dataset limits the ability to assess the long-term effects of abortion policies. Future research should integrate more granular data, such as county-level analyses, and extend the study period to capture longer-term impacts. Qualitative studies focusing on the lived experiences of affected populations can complement quantitative analyses, offering a deeper understanding of the broader implications of restrictive abortion laws. These efforts will support the development of evidence-based policies aimed at reducing disparities and improving maternal and infant health outcomes.

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 raw dataset, consisting of U.S. infant mortality records from 2021 and 2022, was cleaned and pre-processed to ensure its suitability for analysis. The dataset was first imported from a tab-separated file, with irregular lines handled appropriately. Variables were standardized using `janitor::clean_names()` to ensure consistent naming conventions. Key variables such as state, maternal age, maternal race, year of death, month, and infant death rates were selected for analysis, and missing values were removed.

Additional variables were created to facilitate analysis: `after_injunction`, indicating whether the data correspond to the period after June 2022 (post-Dobbs decision), and `abortion_illegal`, denoting whether the state had restrictive abortion laws. Death rate reliability was assessed, and unreliable rates were flagged for transparency. The death rate column was cleaned to remove annotations and converted to numeric format for analysis. A date variable was created by combining the year and month columns, providing a timeline for analysis. Finally, a date variable was created by combining the year and month variables. The cleaned dataset was saved in CSV and Parquet formats to ensure compatibility with analytical workflows.

B Model details

B.1 Diagnostics

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.

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

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

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

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

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.

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.