Exploring the Influence of Gender and Age on Voting Behavior in the 2020 U.S. Presidential Election*

A Statistical Analysis of Socio-Demographic Factors Shaping Electoral Preferences

Ping-Jen (Emily) Su

March 16, 2024

This study examines the influence of gender and age group on voting behavior in the 2020 U.S. presidential election using Logistic Regression and Bayesian analysis. Contrary to traditional assumptions, males show a lower likelihood of voting for Trump. Additionally, distinct voting patterns emerge across age groups, with younger voters displaying a higher propensity to support Biden. These findings highlight the complex dynamics of political participation and suggest avenues for further research into demographic influences on electoral outcomes.

Table of contents

*Code and data are available at: https://github.com/emisu36/Political-Support-in-US-2020.

1 Introduction

A fundamental component of modern civilizations is the democratic process of choosing leaders, which represents the will of the people in determining how they will be governed. According to this concept, US presidential elections are significant occasions that are characterized by heated discussions, passionate campaigns, and important choices. The complex interactions between socioeconomic, cultural, and demographic elements that affect voting behaviour and influence the political landscape can be examined via the prism of these elections.

Following the 2020 U.S. presidential election, which was marked by previously unseen difficulties and increased political divisiveness, it is important to analyze the key factors that shaped the results. The paper conducts a thorough investigation, concentrating on the relationship between voting behaviour, age, and gender. By examining these factors, we hope to shed light on the larger patterns influencing American democracy and clarify the complex web of factors that inform people's voting decisions, such as recent effects of gender on vote choice (Karami et al. 2022).

Our investigation is based on thorough empirical analysis and uses information from credible sources and academic studies. By utilizing sophisticated statistical techniques, such as logistic regression, we aim to reveal complex relationships between political tendencies and demographic characteristics. By doing this, we hope to further our understanding of the factors that influence electoral behaviour and how they affect democratic governance.

Furthermore, this study is driven by a recognition of the transformative potential of electoral research. We hope to contribute to scholarly discourse by clarifying the influence of age and gender on voting patterns, as well as to offer practical advice for electoral strategists, political analysts, and policymakers. Our findings are intended to equip stakeholders with the knowledge necessary to successfully navigate future election landscapes while negotiating the complicated terrain of American politics. By working together, we hope to make a significant contribution to the current discussion about democracy, voting, and the quest of a more representative and inclusive political system. This study was done possible by R Core Team (2023) and Wickham, Averick, et al. (2019).

2 Data

For this study, we utilized data sourced from the 2020 Cooperative Congressional Election Study (CCES) (Schaffner, Ansolabehere, and Luks 2021), which we processed using R, a programming language widely employed for statistical analysis (R Core Team 2023). To manage the dataset efficiently, we leveraged tools from the `tidyverse` suite (Wickham et al. 2019), including `ggplot2` for visualization (Wickham 2016), `dplyr` for data manipulation (Wickham et al. 2023), `readr` for data importing (Wickham, Hester, and Bryan 2024), and `tibble` for data formatting (Müller and Wickham 2023). Model summaries were generated using the `modelsummary` package (Arel-Bundock 2022). Data retrieval was facilitated by the `dataverse` package (Kuriwaki, Beasley, and Leeper 2023), while the reliability of our data processing and analysis was verified using the `testthat` package (Wickham 2011) The `here` package (Müller 2020) aided in maintaining file organization and ensuring the reproducibility of our analysis.

2.1 Data Discussion

The study focuses on examining the relationship between gender, age group, and voting preference for the 2020 U.S. presidential election candidates. The dataset comprises information on the number of individuals who voted for either Trump or Biden across different gender and age groups. Summing up all participants, we have 43,554 responses across different gender and age groups.

- For the gender variable, we included only male and female which are choices more of sex instead of gender. We have made the decision to remove N/A responses as it may be difficult to the real proportion of male and female for the N/A responses without causing estimation and ethical concerns.
- For the age group variable, we aligned our age groups with the breakdown given in the CCES guide (Schaffner, Ansolabehere, and Luks 2021). We also obtained the age of the participants by calculating the difference between 2020 and the birth year variable (birthyr) given by the dataset. Although there may be concerns of birthday not past yet for the participants, calculating full age would be the most sensible option as birthdays are not provided in the dataset.
- For the voted for variable, we have decided to compare only Trump and Biden as they are the representatives of the two largest parties in the United States. This choice also helps the data fit into the requirement of the logistics regression model we hope to perform.

2.2 Data Summary

The data is structured as follows:

- voted_for: Categorical variable indicating the candidate voted for (Trump or Biden).
- gender: Categorical variable representing the gender of the voter (Female or Male).
- age_group: Categorical variable denoting the age group of the voter (18-29, 30-44, 45-64, or 65 and over).
- n: Number of individuals in each category who voted for the respective candidate.

Figure 1 provides counts of individuals who voted for Trump and Biden, segmented by gender and age group. For instance, in the "Trump" category, the number of females and males who voted for Trump is provided across different age groups. Similarly, the "Biden" category presents counts of individuals who voted for Biden across gender and age groups.

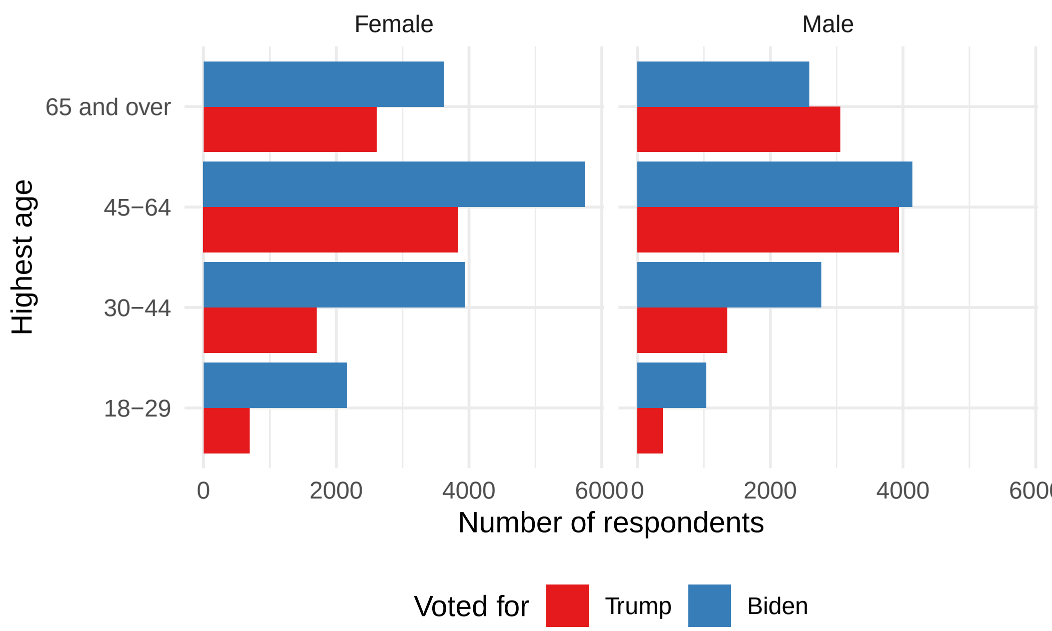

Figure 1: Data Distribution Across Vote, Age Group, Gender

Some interactions between the variables can be noticed in Figure 1.

- Between gender and age group, there is a trend of increasing support for Biden with age, with the highest proportion of Biden supporters in the 45-64 and 65 and over age groups among females. At the same time for males, there is a trend of increasing support for Biden with age, although less pronounced compared to females.

- Among gender and candidate voted for, there is a higher proportion of Biden supporters across all age groups for females compared to males.Among males, there is a higher proportion of Trump supporters across all age groups compared to females.
- For age group and candidate voted for, there is a higher proportion of Biden supporters among females compared to males, across all age groups. For both genders, the proportion of Biden supporters tends to increase with age, while the proportion of Trump supporters decreases with age, although there are some fluctuations.
- Between all three variables,the relationship between age group and the candidate voted for may differ based on gender. For instance, among females, the proportion of Biden supporters increases with age, while among males, the proportion of Trump supporters decreases with age.There may also be variations in voting preferences within each age group based on gender, although the exact patterns may vary.

Overall, these patterns suggest complex interactions between gender, age group, and the candidate voted for, highlighting the importance of considering multiple demographic factors when analyzing voting behaviour.

3 Model

Our study implements logistic regression to examine the relationship between gender, age group, and voting preference for the 2020 U.S. presidential election candidates. Logistic regression is suitable for analyzing binary outcomes, making it an appropriate choice for investigating whether individuals voted for Biden based on their gender and age group.

While Bernoulli distributions represent the binary outcome in logistic regression, Bayesian analysis extends this framework by incorporating prior beliefs and updating them using observed data to obtain a posterior distribution of model parameters. The Bayesian model is shown in Appendix A.

3.1 Model set-up

In this logistic regression model, we aim to predict the probability of a binary outcome y_i (voting for Biden) for each individual i based on their gender and age group. We denote y_i as the binary outcome variable for individual i, indicating whether they voted for Biden (1) or not (0).π_i is the probability of individual i voting for Biden, given by the logistic function. $\text{logit}(\pi_i)$ is the logit transformation of the probability π_i, which is modeled as a linear combination of predictors.

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \alpha + \beta_1 \times \text{gender}_i + \beta_2 \times \text{age group}_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 5.02) && (4)\\
\beta_2 &\sim \text{Normal}(0, 6.34) && (5)
\end{aligned}
$$

The intercept term, a, in the logistic regression equation. It represents the log-odds of voting for Biden when both gender and age group are zero. The coefficient, β_1, associated with the gender predictor. It represents the change in the log-odds of voting for Biden for a one-unit change in the gender variable, holding age group constant. β_2 is the coefficient associated with the age group predictor. It represents the change in the log-odds of voting for Biden for a one-unit change in the age group variable, holding gender constant. The coefficients a, β_1, and β_2 are assigned prior distributions to reflect our uncertainty about their values before observing the data

The model will estimate the posterior distributions of these parameters based on the observed data, allowing us to make inferences about the relationship between gender, age group, and the likelihood of voting for Biden.

The model is built using R (R Core Team 2023) and the `rstanarm` package of Goodrich et al. (2022).

3.1.1 Model justification

We anticipate that both gender and age group will influence the likelihood of individuals voting for Biden in the 2020 US presidential election. To be specific, we hypothesize that gender may play a significant role in shaping voting behavior, as gender identities and associated social roles can influence political preferences. Additionally, age group is expected to be a relevant predictor, as age often correlates with differing life experiences, values, and policy priorities, which in turn may influence voting decisions.

For gender, we expect that male respondents may exhibit different voting preferences compared to female respondents. Regarding age group, we anticipate that older individuals may demonstrate different voting behaviors compared to younger age groups. Older individuals, particularly those in the 65 and over category, may have different policy priorities compared to younger age groups, which could show in their voting choices. Conversely, younger age groups, such as those in the 18-29 category, may have unique perspectives or concerns that influence their electoral preferences.

4 Results

The logistic regression analysis revealed important predictors of voting for Biden in the 2020 US presidential election. The intercept term, representing the expected voting for Biden when all other predictors are zero, was estimated to be 1.76 (SE = 0.35), indicating a positive baseline probability of voting for Biden.

Gender emerged as a significant predictor of voting behavior, with males exhibiting lower log-odds of voting for Biden compared to females. The coefficient for gender (Male) was estimated to be -0.38 (SE = 0.13), suggesting that males were less likely to vote for Biden than females after controlling for other variables.

In terms of age group, distinct patterns were observed across different age categories. Compared to the reference category (18-29), individuals in the 30-44 age group showed a decrease of voting for Biden by -0.56 (SE = 0.39). Similarly, individuals in the 45-64 age group displayed a larger decrease in the log-odds of voting for Biden, with a coefficient of -1.36 (SE = 0.37). The largest decrease was observed among individuals aged 65 and over, with a coefficient of -1.55 (SE = 0.38), indicating a substantial reduction in the likelihood of voting for Biden compared to the youngest age group.

The model fit statistics indicated moderate model performance, with an R-squared value of 0.056, suggesting that approximately 5.6% of the variance in voting behavior for Biden was explained by the predictors included in the model. Overall, the results suggest that both gender and age group are significant predictors of voting behavior for Biden in the 2020 US presidential election, with males and older individuals demonstrating decreased likelihoods of supporting Biden compared to females and younger age groups, respectively.

The results are summarized in Table 1.

Table 1: Logistic Regression Coefficients Estimating the Effect of Gender and Age Group on Voting for Biden

	First model
(Intercept)	1.76
	(0.35)
genderMale	−0.38
	(0.13)
age_group30-44	−0.56
	(0.39)
age_group45-64	−1.36
	(0.37)
age_group65 and over	−1.55
	(0.38)
Num.Obs.	1000
R2	0.056
Log.Lik.	−648.192
ELPD	−653.2
ELPD s.e.	9.2
LOOIC	1306.4
LOOIC s.e.	18.4
WAIC	1306.4
RMSE	0.48

5 Discussion

This study looks into the influence of gender and age group on voting behavior in the context of the 2018 U.S. election. Using Logistic Regression and Bayesian analysis, we explore how these demographic factors shape electoral decisions, providing valuable insights into the dynamics of political participation.

5.1 Age Group Influence on Voting Patterns

As we shift our focus to age groups, we find that various demographic cohorts exhibit diverse voting behaviours. Voters under the age of thirty-nine, in particular, were more likely than older age groups to support Biden. This generational gap emphasizes the differing viewpoints held by people from various generations and emphasizes the significance of age in determining political preferences. A more sophisticated knowledge of electoral dynamics can be achieved by comprehending these age-based voting patterns, which offer insightful information on the changing political beliefs and behaviours of different age groups.

5.2 Gender Influence on Voting Patterns

We find interesting trends about voting preferences and gender in our analysis. Male voters were less likely than female voters to support Trump, which went against the established gender-based voting conventions. This research points to an emerging political environment where reasons other than economic conservatism are important, implying that gender dynamics in political behaviour may be more nuanced than previously thought. The gender-based voting trends that have been discovered provide insight into how political identity and candidate assessment are evolving.

5.3 Weaknesses and next steps

There are a few drawbacks to the logistic regression model, despite the fact that it offers insightful information on the connection between voting behaviour, age group, and gender. Initially, unmeasured confounding variables like education level, political ideology, and regional variations may limit the model's forecast accuracy. If these factors are not included in the study, the results may have skewed estimations and be less broadly applicable. Secondly, the cross-sectional nature of the data utilized in this study hinders our ability to examine changes in voting behaviour over time or demonstrate causality. By using a wider range of factors and using longitudinal data analysis to capture temporal patterns in voting preferences, future study could solve these limitations. Additionally, the sample size and composition of the dataset may not fully represent the diverse population of voters in the United States, potentially limiting the external validity of the findings. Future studies could benefit from larger and more representative samples to enhance the generalizability of the results.

5.4 Future Research

Future research in this area could explore several avenues to build upon the findings of this study. First, incorporating additional demographic, socioeconomic, and political variables into the analysis could provide a more comprehensive understanding of the factors driving voting behavior. To represent the complex dynamics of electoral decision-making, variables including education level, political ideology, party affiliation, and geographic region could be added. Second, long-term research on people's voting patterns may provide insight into the consistency and changes in political inclinations among various demographic groupings. Researchers could look at how societal trends, political events, and changes in life circumstances affect voting decisions by using longitudinal data analysis. Finally, by shedding light on the underlying attitudes and reasons influencing voters' decisions, qualitative research techniques like focus groups and interviews could supplement quantitative analysis. Future study could produce a more thorough and nuanced knowledge of the numerous elements influencing electoral outcomes by utilizing a multidisciplinary approach that integrates quantitative and qualitative approaches.

5.5 Conclusion

In conclusion, this study examined the association between gender, age group, and voting behaviour in the 2020 U.S. presidential election using logistic regression analysis. The results showed that people's likelihood of voting for Biden was highly influenced by their age and gender. In particular, compared to females and younger age groups, males and older age groups were less inclined to favour Biden. Even though the model shed light on the demographic factors that influence voting behaviour, there are a few things to keep in mind. These include the cross-sectional character of the data and the absence of certain significant confounding variables. Despite these limitations, the study contributes to our understanding of the complex interplay between demographic characteristics and political preferences, highlighting the need for further research to elucidate the underlying mechanisms driving electoral outcomes.

Appendix

A Model details

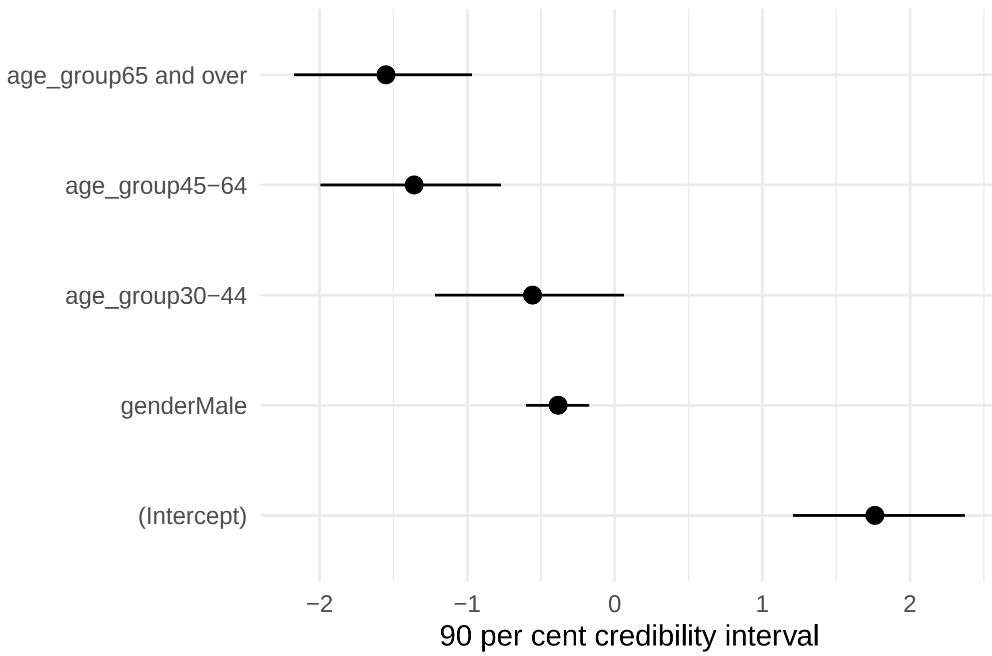

Figure 2: Logistic Regression Model

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.

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

Karami, Amir, Spring B. Clark, Anderson Mackenzie, Dorathea Lee, Michael Zhu, Hannah R. Boyajieff, and Bailey Goldschmidt. 2022. "2020 u.s. Presidential Election in Swing States: Gender Differences in Twitter Conversations." *International Journal of Information Management Data Insights* 2 (2): 100097. https://doi.org/https://doi.org/10.1016/j.jjimei.2022.100097.

Kuriwaki, Shiro, Will Beasley, and Thomas J. Leeper. 2023. *dataverse: R Client for Dataverse 4+ Repositories*.

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

Müller, Kirill, and Hadley Wickham. 2023. *tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

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

Schaffner, Brian, Stephen Ansolabehere, and Sam Luks. 2021. "Coorperative Election Study Common Content, 2020." Harvard Dataverse. https://doi.org/10.7910/DVN/E9N6PH.

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

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

Wickham, Hadley et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

——— et al. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

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.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.