An Analysis of Voter Factors in the 2020 US General Election*

Yuchen Chen

March 13, 2024

This article uses the 2020 Collaborative Election Study (CES) to investigate the relationship between gender, race, age, census area, and presidential preferences among voters in the 2020 US presidential election. Research has found that the black population is more inclined to support Biden, younger voters are more inclined to support Biden, and the male population is more inclined to support Trump. These insights can provide valuable insights for the presidential candidate team to develop effective policies to increase support.

1 Introduction

The quadrennial presidential election is a core component of American politics, reflecting the changes in the political, economic, and social ecology of the United States. The core element of a presidential election is the voters. In 1960, American scholars such as Angus Campbell proposed "three voter voting tendencies: political party identification, candidate evaluation, and issue voting" (Angus Campbell (1960)). From a practical perspective, there are differences among voters in terms of ethnicity, education level, religious beliefs, gender and age, as well as their perceptions of values and political consciousness, and their attitudes towards exercising their political rights, They will have an impact on the general election through their voting behavior.

In this article, we investigated the relationship between individual characteristics and behavior of voters, which mainly include gender, age, race, and census area. Behavior is represented by the preferences of voters towards candidates, that is, whether they support Trump or Biden. Different groups have different concerns and interests, and we attempt to predict who they may vote for using the gender, age, race, and voting area of voters, that is, the preferences of different characteristic groups towards two candidates.

*Code and data are available at: https://github.com/Victor1114/Political-support-in-the-United-States

The data used in this article is sourced from the 2020 Collaborative Election Study (CES) (Schaffner, Ansolabehere, and Luks (2021)). Through research, we found that personal characteristics have a significant impact on voter behavior. Specifically, men are more willing to support Trump, black people are more inclined to support Biden, and younger voters have a more pronounced tendency towards Biden. This study can provide valuable insights for the presidential candidate team to develop effective policies to attract voters and increase support.

In Section 1, we will introduce the research objectives and data used in this article. The second section will provide a detailed introduction and presentation of the data, and conduct visual analysis of the data, focusing on the relationship between four personal characteristics and behavior. In Section 3, a logistic regression model will be established to demonstrate and predict the relationship between individual characteristics and behavior of voters in Russia. Section 4 presents the model results and analyzes them. Section 5 analyzes and introduces the research results and the weaknesses of this article, and explains the next research plan.

2 Data

2.1 Data Description and Methodology

The data used in this article is from the 2020 Collaborative Election Study (CES) and can be publicly obtained through the Harvard University Data Center (Schaffner, Ansolabehere, and Luks (2021)). Cooperative Election Study (CES) is a long-standing annual survey of American political opinion that primarily examines how Americans perceive Congress, election experiences, voting behavior, and more. In 2020, a total of 60 teams participated in the cooperative election research. Each team purchased a national sample survey of 1000 people, and a total of 61000 respondents completed the post election survey. The samples were selected through Internet sampling, and the method used was mainly based on matching. Ansolabehre, Schaffner and Luks (Ansolabehere and Luks (2021)) introduced this in detail.

This article will use 5 data variables, namely voting behavior, gender, age, race, and census area. In the actual use process, the voters who voted for Trump and Biden were first selected and the original data was processed.

Table 1: A summary table of cleaned data

voted_for	gender	race	age	region
Trump	Male	White	30-44	Midwest
Biden	Female	White	> 64	South
Biden	Female	White	30-44	Midwest
Trump	Male	White	30-44	South
Trump	Female	White	> 64	South

voted_for	gender	race	age	region
Trump	Female	White	30-44	South
Biden	Female	White	30-44	Midwest
Biden	Female	White	30-44	West
Biden	Female	White	30-44	West
Biden	Male	White	30-44	South

Table 1 presents the cleaned dataset, which includes 5 variables and 43553 observations. The meanings of variables in the dataset are as follows:

- Voted for: support Trump or support Biden.
- Gender: Male or female.
- Race: Divided into four categories: White, Black, Hispanic, Asian, and Other.
- Birthyr: The age of voters in 2020, divided into four categories: 18-29, 30-44, 45-64, and>64.
- Region: Census reagion divided into Northeast, Midwest, South, and West.

This article mainly uses programming software R (R Core Team (2022)) and other tools for data processing, visualization, and modeling analysis, such as "knit" (Xie 2023), "here" (Müller 2020), "modelsummary" (V 2022), "tidyverse" (Wickham et al. 2019) "rstanarm"(Goodrich et al. (2022)).

2.2 Data Visualization

2.2.1 Presidential preferences

Figure 1 shows the overall voting behavior of 43553 surveyed voters, with approximately 40% supporting Trump and 60% supporting Biden. Overall, Biden's approval rate is higher.

2.2.2 Presidential preferences and gender

Figure 2 shows the relationship between presidential preferences and gender. The image shows that the proportion of males and females is almost equal among those who support Trump, but among those who support Biden, the number of females is about 1.5 times that of males. This result suggests that women may have a certain bias towards supporting Biden, and gender can provide valuable insights for predicting voter behavior.

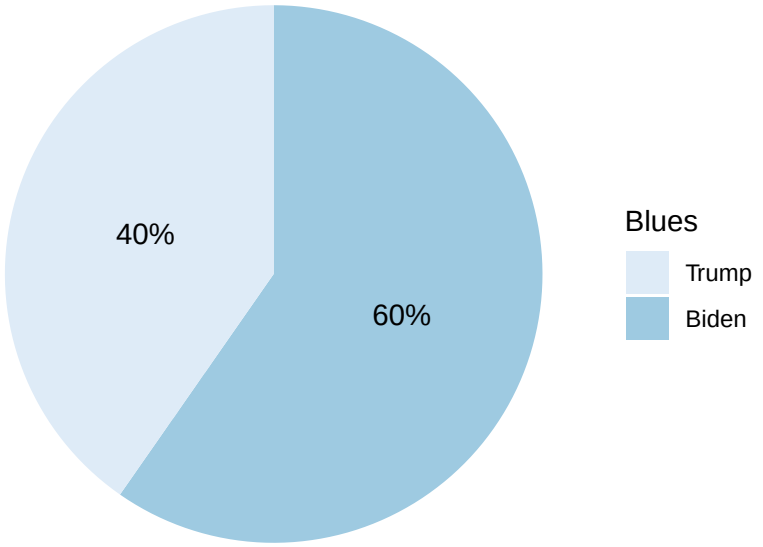

Figure 1: The distribution of presidential preferences

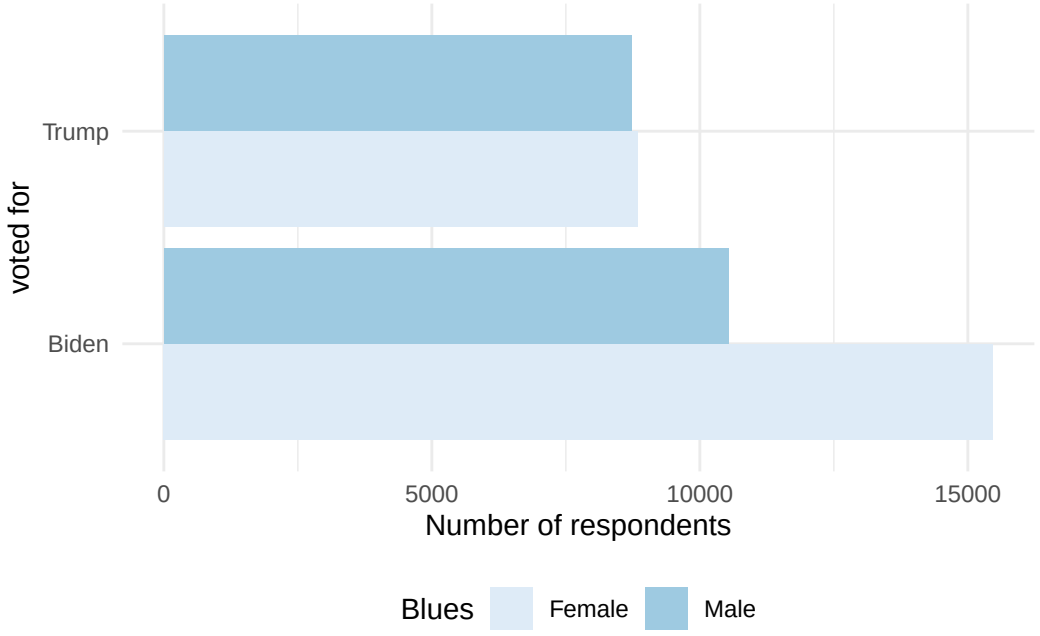

Figure 2: The distribution of presidential preferences by gender

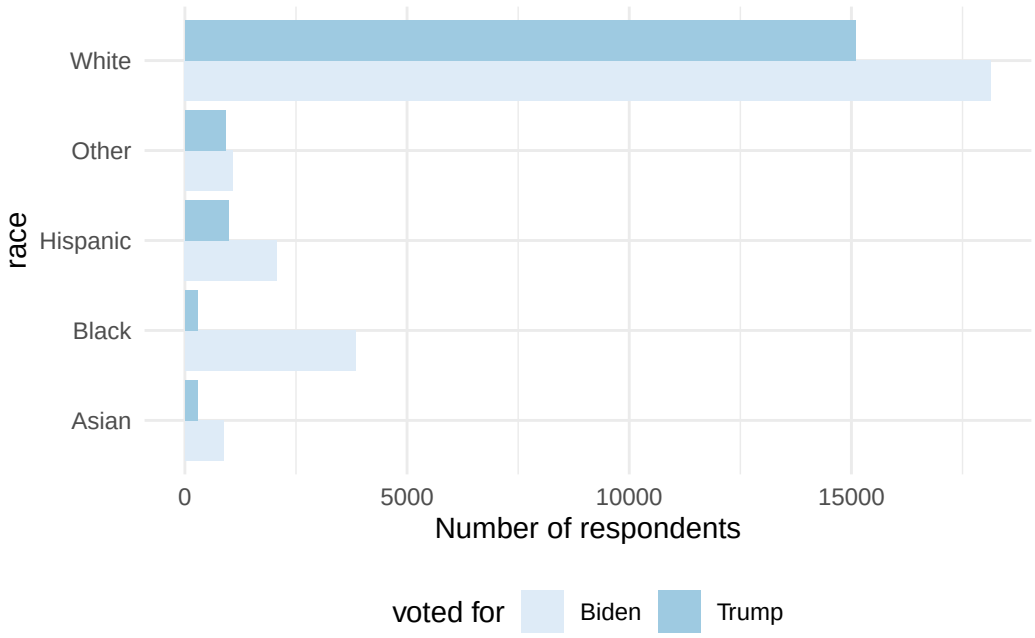

Figure 3: The distribution of presidential preferences by race

2.2.3 Presidential preferences and race

Figure 3 shows the relationship between presidential preferences and race. The image shows that the proportion of black people supporting Biden is much higher than that of Trump, while the proportion of Hispanic people supporting Biden is about twice that of Trump. The difference in the proportion of white people supporting the two is relatively small. This result indicates that there are certain differences in presidential bias among people of different races, and analyzing race can provide valuable insights for predicting voter behavior.

2.2.4 Presidential preferences and age

Figure 4 shows the relationship between presidential preferences and age. The figure shows that the proportion of people aged 18-29 who support Biden is much higher than Trump, while the proportion of people aged 30-44 who support Biden is about twice that of Trump. The difference in the proportion of people aged 64 and above who support the two is relatively small. This result indicates that there are certain differences in presidential bias among different age groups, and analyzing age can provide valuable insights for predicting voter behavior.

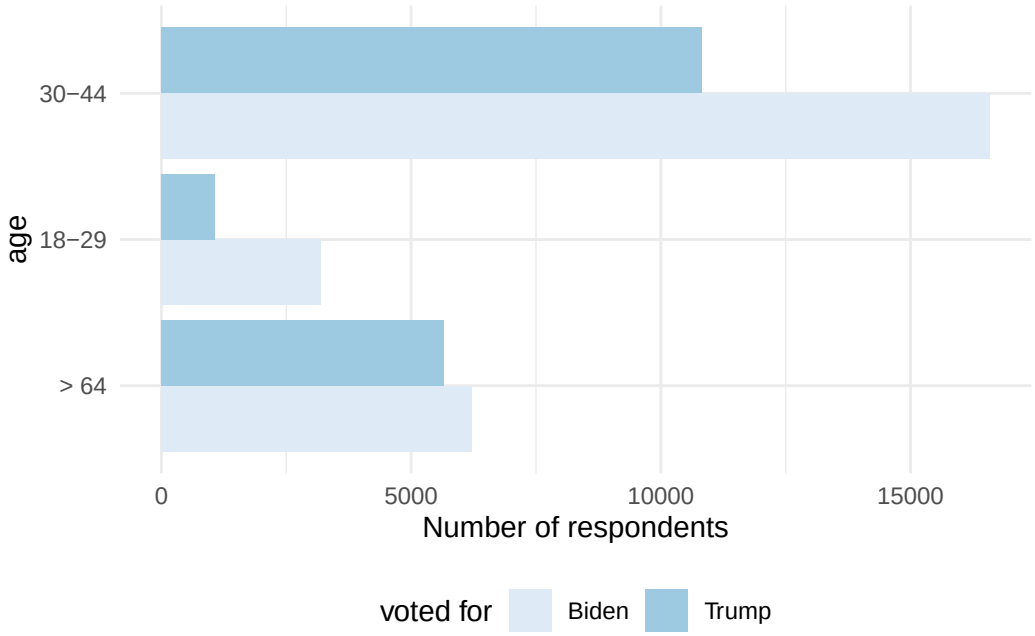

Figure 4: The distribution of presidential preferences by age

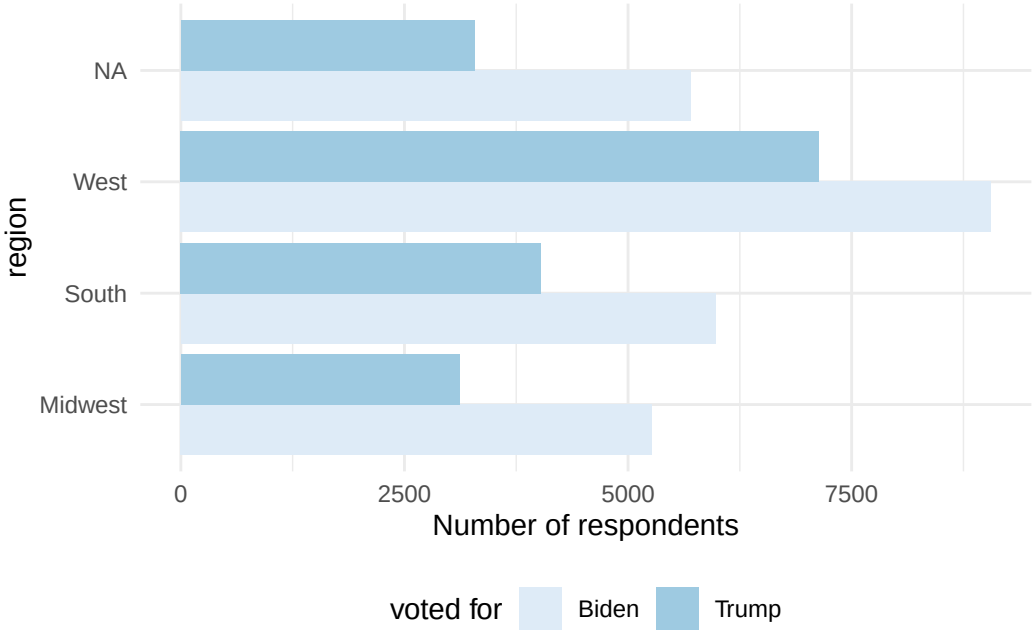

Figure 5: The distribution of presidential preferences by region

2.2.5 Presidential preferences and region

Figure 5 shows the relationship between presidential preferences and census regions. The image shows that regardless of which region, the proportion of people supporting Biden is higher than Trump. However, the largest difference in this proportion is in the west, and the smallest difference is in the south. Therefore, it can be considered that different census regions have different biases towards presidents. Analyzing the census regions where voters are located can provide valuable insights for predicting voter behavior.

3 model

After visual exploratory analysis of the data, we found that gender, age, race, and census area all have a certain impact on voter voting behavior, indicating a potential relationship between these four variables and voter behavior. To further analyze the impact of these variables on voter behavior, a logistic regression model was established, with the following model form:

$$y_i|\pi_i \sim Bern(\pi_i)$$
$$logit(\pi_i) = \beta_0 + \beta_1 * gender_i + \beta_2 * race_i + \beta_3 * birthyr_i + \beta_4 * region_i$$
$$\beta_i \sim Normal(0, 2.5) \quad (i = 0, 1, 2, 3, 4)$$

where:

- y_i is the presidential preference of voter i, supporting either Trump or Biden.
- π_i is the probability that voter i will vote for Trump.
- β_0 is intercept.
- β_i $(i = 1, 2, 3, 4)$ is the slope coefficient.

The independent variables of the model, gender, race, birth, and region, are all categorical variables, while the dependent variable is the presidential preference of voters, that is, whether they support Trump or Biden, which is also a categorical variable.

4 Results

Table 2: Whether a respondent is likely to vote for Biden based on their gender, race, age and census region.

	Support Trump
(Intercept)	-0.747
	(0.076)
genderMale	0.275
	(0.021)
raceBlack	-1.843
	(0.094)
raceHispanic	0.185
	(0.080)
raceOther	0.719
	(0.085)
raceWhite	0.636
	(0.071)
age18-29	-0.921
	(0.041)
age30-44	-0.663
	(0.030)
age45-64	-0.102
	(0.024)
regionNortheast	-0.131
	(0.032)
regionSouth	0.298
	(0.027)
regionWest	-0.14
	(0.031)
Num.Obs.	43554
R2	0.092
Log.Lik.	-27096.052
ELPD	-27108
ELPD s.e.	70.3
LOOIC	54216
LOOIC s.e.	140.6
WAIC	54216
RMSE	0.47

Table 2 presents the predictive variable coefficients of the logistic regression model, and these variables in the model have p-values less than 0.1, indicating a significant impact on voters' presidential preferences. The results show that men are more likely to support Trump. From a racial perspective, black people are more inclined to support Biden, while people of other

races are more inclined to support Trump. From an age perspective, people of all age groups are more inclined to support Biden, but the older they get, the weaker this tendency becomes. From the census area, people in the South region tend to support Trump, while others are more inclined to support Biden.

5 Discussion

5.1 Male voters are more supportive of Trump.

Due to differences in social division of labor, there are differences in policy preferences between male and female voters, which leads to differences in their presidential preferences. Our research results show that male voters are more inclined to support Trump, while female voters are more inclined to support Biden.

Since 1980, there has been a persistent difference in party orientation between male and female voters, where female voters always have a higher support rate for Democratic candidates than their vote share among all voters, while male voters always have a higher support rate for Republican candidates than their vote share among all voters. Meanwhile, in the past 50 years, economic power has been redistributed from men to women, and Trump, with his masculinity, has a stronger appeal to male voters.

5.2 Black people are an important supporter of Biden

Our research findings indicate that the black population is more likely to support Biden than Trump, and Trump's advantage lies in the white population. The reason for this phenomenon may be the continuous fermentation of racial discrimination. The Democratic Party's tolerance towards non white groups won Biden more votes. Trump, like in the 2016 election, chose to promote "white supremacy" and gained the support of the largest group of white people, but this also pushed black people towards Biden.

5.3 Young people are more supportive of Biden

From an age perspective, younger voters are more likely to support Biden and have a more pronounced bias. As the younger generation, the millennials and Generation Z are very concerned about topics such as climate change, racial discrimination, and economic inequality, and hope that the government can come up with more solutions. The agenda proposed by the Democratic Party can address the concerns of the younger generation, especially the economic needs. Trump's anti immigration remarks and support for white supremacism further push young voters to lean towards the Democratic Party.

Trump's campaign slogan "Let America be great again" caters to the elderly population that has been around since the 1950s, and he has repeatedly promised not to touch the money of social security and healthcare, which is the most concerning issue for the elderly. Therefore, in terms of age structure, the younger group is more inclined to support Biden, while the older group has relatively higher support for Trump.

5.4 Weaknesses and Next Steps

This article analyzes the relationship between four personal characteristics of voters: gender, age, race, census area, and presidential preferences. Although some connections were discovered through data visualization and modeling analysis, providing valuable clues, there are also certain limitations.

On the one hand, the data used in this study was sampled and selected through the Internet, which means that the sample does not include people who cannot access the internet. This group of people may have a certain impact on the analysis results. Another convenience is that the research in this article is to some extent limited by the time of data collection, and the preferences of voters will change with the times and policies. At the same time, 2020 is during the epidemic period, which is a relatively special historical background. Therefore, the results of this study have certain time limitations. In future research, it is planned to integrate multiple institutional survey datasets and expand the timeline of election related data to improve data diversity and enhance the validity of the study.

Appendix

A.1 Sketches

Sketches of the main data and graphs are available in the GitHub Repository associated withthis report.

A.2 Simulation

A script containing data simulation is available in the CitHub Repository associated with thisreport. This script was generated using the programming software R (R Core Team (2022)) and the tidyverse (Wickham et al. 2019) packages.

A.3 Tests

We tested simulated data using R programming language. The GitHub repository associated with this report provides scripts containing test code.

A.4 model

We use R programming language(R Core Team (2022)) to establish a logistic regression model for real data, estimating the parameters using stan_glm(). The GitHub repository associated with this report provides scripts for modeling.

References

Angus Campbell, etc. 1960. "The American Voter." *Chicago:The University of Chicago Press*.

Ansolabehere, Brian Schaffner, Stephen, and Sam Luks. 2021. "Guide to the 2020 Cooperative Election Study." 1https://doi.org/10.7910/DVN/E9N6PH.

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

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

R Core Team. 2022. *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. "Cooperative Election Study Common Content, 2020." Harvard Dataverse. https://doi.org/10.7910/DVN/E9N6PH.

V, Arel-Bundock. 2022. "Modelsummary: Data and Model Summaries in r." *Journal of Open Source Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

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

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