Predicting Political Support in the United States*

The Effects of Culture and Key Political Choices

Janssen Myer Rambaud Timothius Prajogi

March 14, 2024

Using data gathered from the 2022 Cooperative Election Study (CES), we navigate through the various possible factors that influence a US citizen's voting decision. By examining a representative sample of 60 000 Americans, we can discern patterns in beliefs and personal backgrounds—ranging from gun control and student loan stances to race, religion, and educational background. This paper discusses the impact these individual factors may have towards predicting a voter's alignment, as well as how multiple factors can coalign to portray a more substantial impact. We aim to highlight these correlations and their potential implications with future elections.

Table of contents

*Code and data are available at: https://github.com/prajogt/predicting_political_support_us.git .

1 Introduction

Voting for the leader of your country is one of the biggest decisions an American citizen can make. It decides many of the decisions that will be made throughout that four year term as well as how their life will change in that period of time. There may be changes that this elected government makes that they believe to be a great decision, but it may also be seen as terrible one in the eyes of others. This is why it is important to understand there are many factors that influence a voter's decision. They vote for the candidate who they believe will make the best decisions for them and their country. In this paper, we hope to show the impact of various factors on a voter's decision, and how these factors can be used to predict a voter's alignment. From personal stances and religious beliefs to political stances, or to biological factors, we hope to show how these factors can be used to predict a voter's alignment.

The topic of concern is the 2020 United States presidential election held on November 3rd, 2020, where we saw Democratic Party leader Joe Biden win against incumbent president Donald Trump (Gambino 2020). Using the 2022 Cooperative Election Study (CES) survey data, we aim to predict who a voter supported in the 2020 election based on their cultural background, political stances, and other factors. We will be using a representative sample of 60000 Americans to discern patterns in these voters' responses and how their personal backgrounds may have influenced their voting decision. We will be examining the impact of various factors such as: gender, race, educational background, religon, and personal stances on key political issues such as gun control and student loans. Our main focus is to compare the correlations in the 2022 CES survey data to the 2020 CES survey data, and to see if our estimand, the effect of a voter's cultural background and political stances on their voting decision, has influenced their voting decision in the 2020 election. Using these understandings, we hope to highlight the potential implications these correlations may have for this upcoming 2024 election.

With this context and information in mind, we developed this paper to be structured as follows. We will first present the data and the models we used to analyze the data. We will then present the results of our analysis, and discuss the implications of these results. Finally, we will conclude with a summary of our findings and the potential implications of these findings for the upcoming 2024 election. To go in more detail, after downloading and cleaning the 2022 CES survey data, we used the `rstanarm` (Goodrich et al. 2024) package to create a generalized linear model to predict who a voter supported in the 2020 election based on their cultural background, political stances, and other factors, as listed above. We then used the `modelsummary` (Arel-Bundock 2022) package to summarize the results of our model. We found that there were strong correlations, with a strong example of this being the voter's personal stances on student loan forgiveness and gun ownership. With the two main parties, Democrats and Republicans, having opposing stances on these issues, we found that there was a strong correlation between these stances and who a voter supported. As Biden was in favour of student loan forgiveness, but opposed to gun ownership, while Trump was opposed to student loan forgiveness, but in favour of gun ownership, we discovered a correlation between a voter's ties to these stances because they directly oppose each other. This is just one example of the many correlations we found, and we will discuss these in more detail in the results section. These findings are important because it shows how a leading presidential party's stance on a certain topic has the potential to sway a voter's decision in the event that they align with that stance. This is important because it shows how a leading presidential party's stance on a certain topic has the chance to adjust and adapt to certain factors to possibly sway some more votes in their favour. We will also discuss the implications of these findings, and how they may be used to predict a voter's alignment in the upcoming 2024 election.

TODO: cite https://www.theguardian.com/us-news/2020/nov/07/joe-biden-wins-us-election-donald-trump-loses-final-result-2020

2 Data

These delay statistics were downloaded, cleaned, parsed, analyzed, and visualized using R (R Core Team 2023), a statistical programming language, with package support from `tidyverse` (Wickham et al. 2019), a collection of libraries which included the following packages that were utilized:

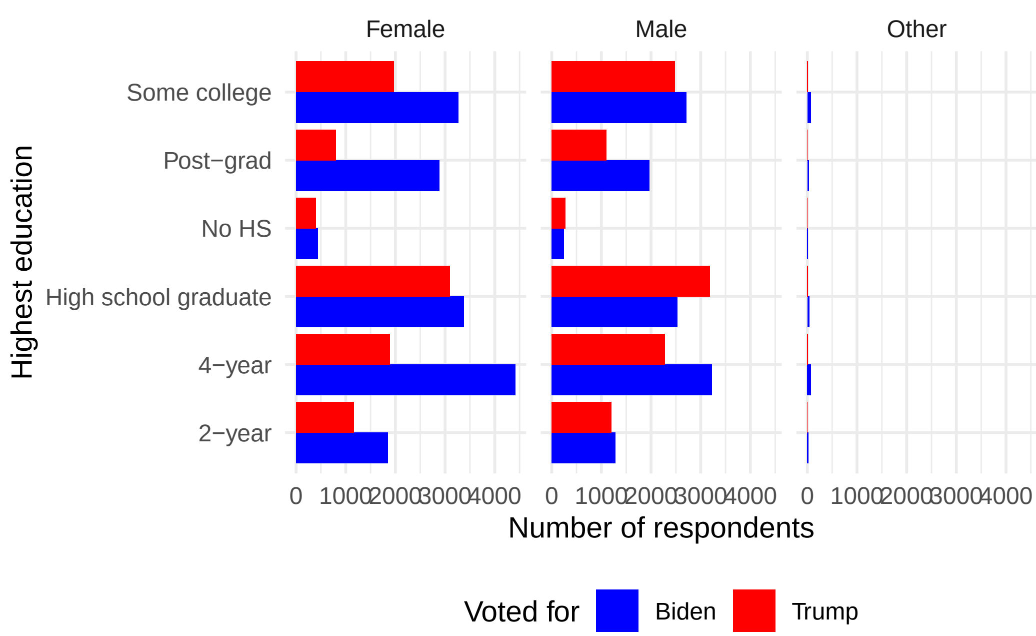
- `ggplot2` (Wickham 2016)
- `dplyr` (Wickham et al. 2023)
- `readr` (Wickham, Hester, and Bryan 2023)
- `tibble` (Müller and Wickham 2023)

The data was retrieved from Harvard's CCES Dataverse database, using the `dataverse` (Kuriwaki, Beasley, and Leeper 2023) package. In this report we consider data collected from the 2022 CES survey, stored by Dataverse as *Cooperative Election Study Common Content* (Schaffner, Ansolabehere, and Shih 2023). There was an option to use the 2020 CES dataset, but we opted to use the newer, 60000 observation, 2022 CES dataset for a more updated and accurate representation of the current political climate in the United States after two years of Joe Biden being in office.

Possible put in appendix? The variables we used in our analysis were: - `presvote20post`, renamed to `voted_for` - This data was represented in the form of a numerical variable corresponding to the presidential candidate that voted for in the 2020 eleciton. This was cleaned to limit the options to either "Biden" or "Trump", the two leading candidates in the 2020 and 2024 election. - `birthyr`, which was used to calculate the age of the voter, and later grouped into age buckets and renamed to `age`. - This data was represented in the form of a numerical variable corresponding to the year the voter was born. This was cleaned to calculate the age of the voter, and then grouped into the age buckets: "18-29", "30-44", "45-64", "65+". - `gender4`, renamed to `gender`, - This data was represented in the form of a numerical variable corresponding to the gender of the voter. This was cleaned to limit the options to "Man", "Woman", and "Other". - `educ`, renamed to education, - This data was reprented in the form of a numerical variable corresponding to the highest level of education the voter has completed. This was cleaned to limit the options to "No HS" "High School graduate", "Some College", "2-year", "4-year", "Post-grad". - `race`, - This data was represented in the form of a numerical variable corresponding to the ethnicity of the voter. This was cleaned to limit the options to "White", "Black", "Hispanic", "Asian", "Middle Eastern", "Native American", "Multiracial", and "Other". - `religpew`, renamed to `religion`, - This data was represented in the form of a numerical variable corresponding to the religion of the voter. This was cleaned to limit the options to "Protestant", "Roman Catholic", "Jewish", "Atheist", "Agnostic", "Non-Religious", and "Other". There were various other religions that were not included in the dataset, but represented a small portion of the sample, and were thus grouped into "Other". - `gun_own`, renamed to `gun_ownership`, - This data was represented in the form of a numerical variable corresponding to whether the voter or anyone in their household owned a gun. This was cleaned to limit the options to "Yes", "No", and "Unsure". - `edloan`, renamed to `student_loan`, - This data was represented in the form of a numerical variable corresponding to whether the voter had student loans. This was cleaned to limit the options to "Yes", "No".

The reason that we chose a large amount of variables/factors for our paper is so that we can get a bigger representation of the political climate. Should this be limited to a single model, then it may not reflect other correlations that may be present in the data. By using these variables, we take the most significant factors that may influence a voter's decision, and use them to predict who a voter supported in the 2020 election `voted_for`. It was important to get the voter's age, `age`, because it is important to understand the generational differences and how their experiences and personal interests may influence their voting decision. In our dataset, it was clear that 'Female' respondents were in more support towards the Democrats at 52% as opposed to 43% for 'Males'. In our given age buckets, it was discovered that 65% of 18-29 year olds supported the Democrats, whereas only 40% of 65+ year olds supported the Democrats, however, 18-29 year olds only made up 12% of the sample, whereas 65+ year olds made up 30% of the sample. It was also important to get the voter's gender, `gender`, and race, `race`, because it is important to understand the biological differences and how their experiences and personal interests may influence their voting decision. In our dataset, it was discovered that 84% and 71% of Black and Asian respondents, respectively, supported the Democrats, whereas only 41% of White respondents supported the Democrats. Incoorporating the voter's religion, `religion`, was important because it is important to understand that religious beliefs and differences can have a strong influence on a person's personal beliefs and therefore impact their personal stances on topics such as gun ownership. To put a contrasting example out there, 30% of Protestant/Christian respondents support the Democrats as opposed to a whopping 72% by Atheists. This could be attributed to certain parties aligning more to certain religious beliefs. A voter's education, `education`, was important because there may be certain understandings they arrive at during the completion of their education. Perhaps they viewed the world in one way at first, but after becoming more 'educated' in a topic, feel they have a better understanding of the world and therefore have a different stance on certain topics. This goes hand in hand with the voter's stance on student loans, `student_loan`, as it is important to understand how a voter's personal experiences with student loans may influence their voting decision. There was a time of serious consideration during Biden's tenure where he was considering forgiving student loans, and this intent alone may have swayed some support in his favour.

```
ces2022_1 |>
  ggplot(aes(x = education, fill = voted_for)) +
  stat_count(position = "dodge") +
  facet_wrap(facets = vars(gender)) +
  theme_minimal() +
  labs(
    x = "Highest education",
    y = "Number of respondents",
    fill = "Voted for"
  ) +
  coord_flip() +
  scale_fill_manual(values = c("Biden" = "blue", "Trump" = "red")) +
  theme(legend.position = "bottom")
```



3 Model

To create the generalized linear model we made use of the `rstanarm` (Goodrich et al. 2024) package.

In particular the models we had created are logistic regression, all predicting who the participant voted for during the 2020 US Presidential Election.

```
# Load in models
political_preferences_cultural <-
  readRDS(file = "models/political_preferences_cultural.rds")

political_preferences_race <-
  readRDS(file = "models/political_preferences_race.rds")

political_preferences_importances <-
  readRDS(file = "models/political_preferences_importances.rds")
```

The first model that was created was intended to discover whether we can forecast who a respondent was likely to vote for, knowing their cultural background, including their age, gender, race, and religion.

We expect that due to various differences in experiences due to this background, whether it be because of a generational difference, a difference in culture between countries, or a difference in beliefs in religions, a voter who identifies in those demographics would vote differently, voting for the candidate who most undestands their experiences.

The second model only forecasts who the respondent was likely to vote for based on race, to extract how much of the intercept in the previous model can be explained by a different in race.

Finally, the last model aims to forecast a respondents vote knowing their prominent political ideals including education, gun ownership status, and student loan status. These are common political standpoints in the U.S.. Gun policy and student loan forgiveness is often a key political issue that separates the candidates. For the 2020 election, the candidates we are predicting for, Joe Biden and Donald Trump had opposing positions on student loan forgiveness. Whilst Trump was opposed to student loan forgiveness, backing incone-based repayment plans instead ()(**citetrumploanstance?**), Biden proposed to forgive a large amount of undergraduate tuition-related student debt ()(**citeballotpedia?**). They also opposed each other on gun laws, with Trump advocating for the right to bear arms and Biden advocating gun safety policies aimed at holding manufacturers accountable ()(**citeballotpedia?**).

As such, we expect that there would be a strong correlation with the respondents gun ownership and student loan status to the candidate they supported in the 2020 election.

TODO add to refs

https://www.washingtonpost.com/politics/interactive/2023/presidential-candidates-2024-policies-issues/donald-trump-student-loans-education/

https://ballotpedia.org/2020_presidential_candidates_on_student_loan_debt https://www.rand.org/research policy/key-findings/gun-policy-in-america.html

```
modelsummary(
  list(
    "Support Biden" = political_preferences_importances
  ),
  statistic = "mad"
  )
```

3.1 Results

3.2 Discussion

Although there seems to be a strong correlation between if a voter identifies as an other gender to supporting Biden, and a strong correlation between if a voter identifies as an other race to supporting Biden, but these must be considered carefully as within the sample (with seed 302), only 5 and 1 voters identified in those categories respectively. To get a more accurate reading in regards to these features, it would be necessary to train a model which includes much more of these demographics.

3.3 Appendix

```
cm <- c('genderMale' = 'Male',
        'raceWhite' = 'White',
        'raceNative American' = 'Native American',
        'raceMultiracial' = 'Multiracial',
        'raceMiddle Eastern' = 'Middle Eastern',
        'raceHispanic' = 'Hispanic',
        'raceBlack' = 'Black',
        'raceAsian' = 'Asian',
        'religionNone' = 'Non-Religious',
        'religionProtestant' = 'Protestant',
        'religionAgnostic' = 'Agnostic',
        'religionAtheist' = 'Atheist',
        'religionRoman Catholic' = 'Roman Catholic',
```

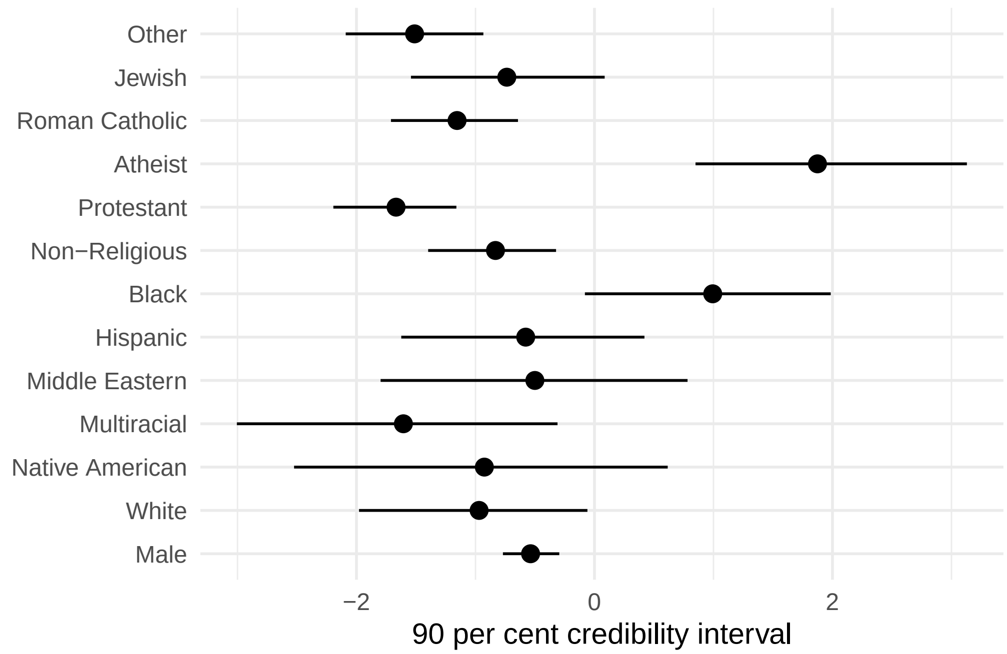
Table 1: Explaining whether a voter supported Biden or Trump in 2020 given their gender, age, race, and religion

	Support Biden
(Intercept)	2.286
	(0.669)
genderMale	−0.537
	(0.149)
genderOther	23.740
	(20.653)
age_group30-44	0.634
	(0.273)
age_group45-65	0.045
	(0.254)
age_group65+	0.110
	(0.259)
raceBlack	0.994
	(0.633)
raceHispanic	−0.578
	(0.613)
raceMiddle Eastern	−0.501
	(0.800)
raceMultiracial	−1.606
	(0.805)
raceNative American	−0.925
	(0.948)
raceOther	47.330
	(44.616)
raceWhite	−0.970
	(0.582)
religionAtheist	1.874
	(0.673)
religionJewish	−0.737
	(0.491)
religionNone	−0.832
	(0.321)
religionOther	−1.512
	(0.348)
religionProtestant	−1.668
	(0.307)
religionRoman Catholic	−1.155
	(0.310)
Num.Obs.	1000
R2	0.177
Log.Lik.	−585.616
ELPD	−604.2
ELPD s.e.	13.7
LOOIC	1208.4
LOOIC s.e.	27.4
WAIC	1208.1
RMSE	0.45

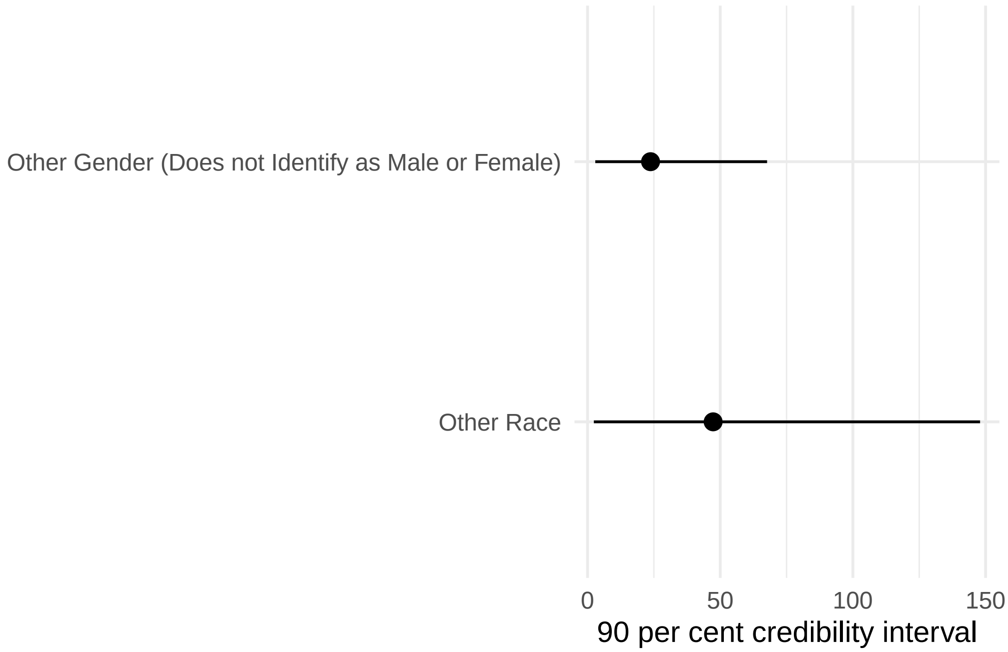
	Support Biden
(Intercept)	0.594
	(0.213)
education4-year	0.267
	(0.246)
educationHigh school graduate	0.005
	(0.251)
educationNo HS	0.450
	(0.658)
educationPost-grad	0.400
	(0.269)
educationSome college	0.255
	(0.252)
household_gun_ownershipYes	−1.152
	(0.147)
student_loan_statusYes	0.587
	(0.188)
Num.Obs.	948
R2	0.097
Log.Lik.	−593.812
ELPD	−602.1
ELPD s.e.	11.0
LOOIC	1204.2
LOOIC s.e.	22.0
WAIC	1204.2
RMSE	0.47

```
        'religionJewish' = 'Jewish',
        'religionOther' = 'Other'
        )

modelplot(political_preferences_cultural, conf_level = 0.9, coef_map = cm) +
  labs(x = "90 per cent credibility interval")
```




```
modelplot(political_preferences_cultural,
          conf_level = 0.9,
          coef_map = c("raceOther" = "Other Race", "genderOther" = "Other Gender (Does not
  labs(x = "90 per cent credibility interval")
```


Other Gender (Does not Identify as Male or Female)
Other Race
0
50
100
150
90 per cent credibility interval

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. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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

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 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.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.