Predicting Political Support in the United States*

The Effects of Culture and Key Political Choices

Janssen Myer Rambaud Timothius Prajogi

March 12, 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

include Estimand in here

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:

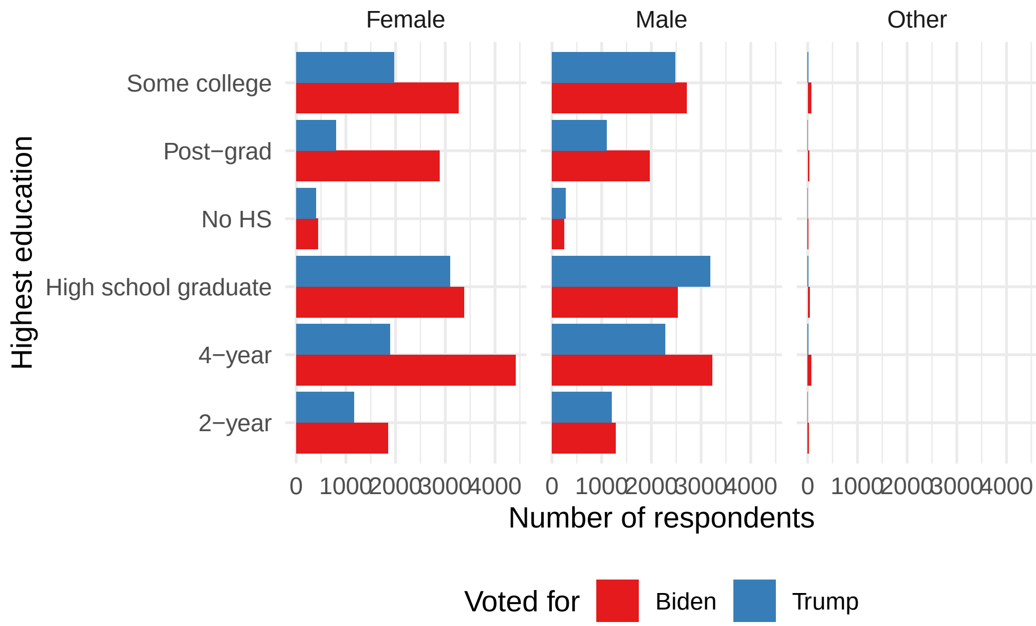
- `ggplot2` (Wickham 2016)
- `dplyr` (Wickham et al. 2023)
- `readr` (Wickham, Hester, and Bryan 2023)
- `tibble` (Müller and Wickham 2023)

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).

needs measurement here

```
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_brewer(palette = "Set1") +
  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 prediciting 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

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

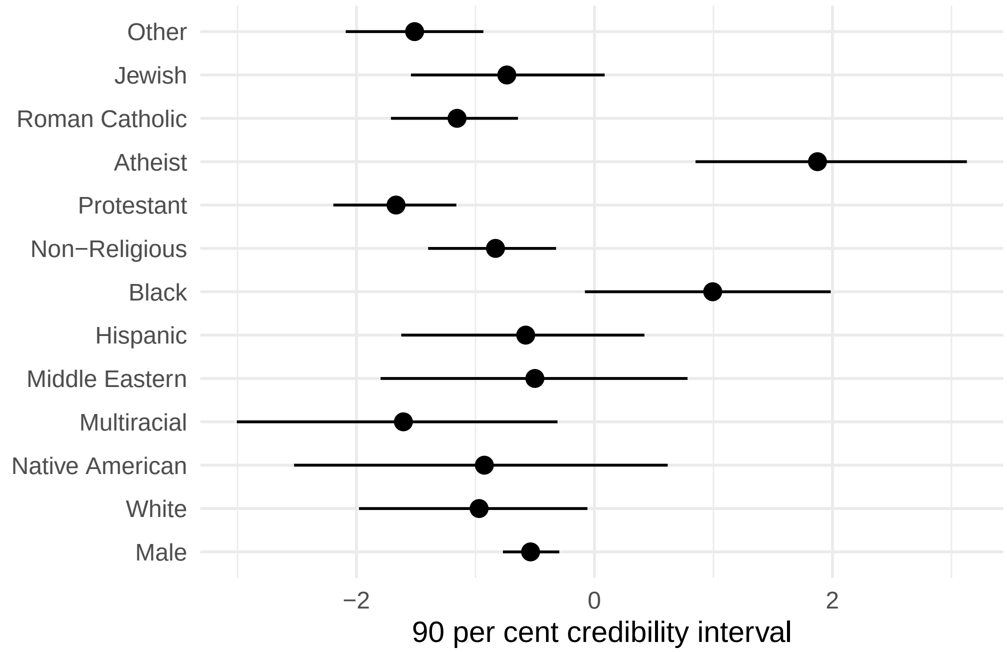
	Support Biden
(Intercept)	0.594
	(0.213)
education4-year	0.267
	(0.246)
educationHigh school graduate	0.005
	(0.251)
educationNo HS	0.450
	(0.658)
educationPost-grad	0.400
	(0.269)
educationSome college	0.255
	(0.252)
household_gun_ownershipYes	−1.152
	(0.147)
student_loan_statusYes	0.587
	(0.188)
Num.Obs.	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

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.

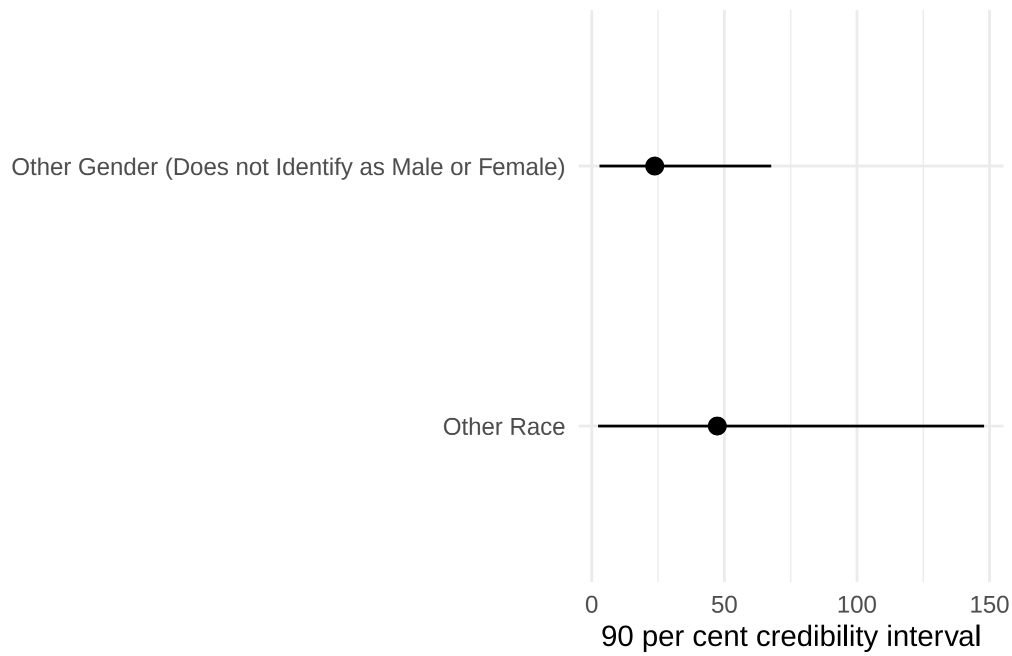
3.3 Appendix

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

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




```r
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

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.