

US General Society Survey Analysis*

Cristina Burca, Yan Mezhiborsky, Sakura Noskor

March 15, 2024

Abstract

Some abstact

Table of contents

Introduction	2
Data	3
Data Used	3
Variables inspected	3
The Destination to Reach with the Data	4
Model 1	4
Model 2	9
Model justification	10
Results	10

*Code and data are available at: [repository](#)

Discussion 10

First discussion point 10

Second discussion point 10

Third discussion point 10

Weaknesses and next steps 10

Appendix 11

Additional data details 11

Model details 11

Posterior predictive check 11

Diagnostics 11

References 12

Introduction

This research examines the voting patterns in the 2016 and 2020 US Presidential elections. We will be looking at data collected by the Cooperative Election Study and accessed through the Harvard University Database **cite**. The anaylsis is based on a representative sample of 61,000 American adults, which provides detailed information about each individuals gender, birth year, race, registered state, employment, education loans, immigration status, dual-citizenship, religion, and 2016 and 2020 Presidential vote. The goal of this study is to use relevant variables from the electoral data to investigate patterns, trend and predictions regrading American electoral preferences from 2016 and 2020.

Data

Data Used

This paper was modeled with the help of R (R Core Team 2023) along with other useful packages like tidyverse (Wickham et al. 2019a) (which includes graphing functions like ggplot2), patchwork (Pedersen 2024). There are parts of the code which were guided by Rohan Alexander’s Telling Stories with fire (Wickham et al. 2019b) chapter 13 section 13.2.2.

Variables inspected

Starting off, we examine the columns ‘votereg’ and ‘voted_for’. They represent the number of persons that registered to vote and which candidate they voted for in 2020, respectively. We filtered out the rows with a ‘votereg’ value of 2, which indicated unregistered voters, to focus exclusively on individuals who were registered to vote. We then focused on the ‘presvote16post’ variable, which reveals the candidates Americans voted for in the 2016 United States Presidential Election. This is an important variable as it enables us to assess whether American citizens were satisfied with the service that the previous government provided. Next we look at ‘gender’ as well as ‘employment’. Both ‘gender’ and ‘employment’ shows us if there is a correlation between certain parties views versus the demographic they represent. ‘Gender’ contains 2 values (male and female) while, employment has 9 values; full time worker, part time worker, laid off, unemployed, retired, permanently disabled individual, Homemaker, Student or Other. We also explore the variable ‘immstat’ which represents the immigration status of the of individual represented by one of the following: immigrant and citizen, immigrant not citizen, born in US, but parent(s) immigrant, parent and I born in US but grandparent(s) immigrant, or all born in US.

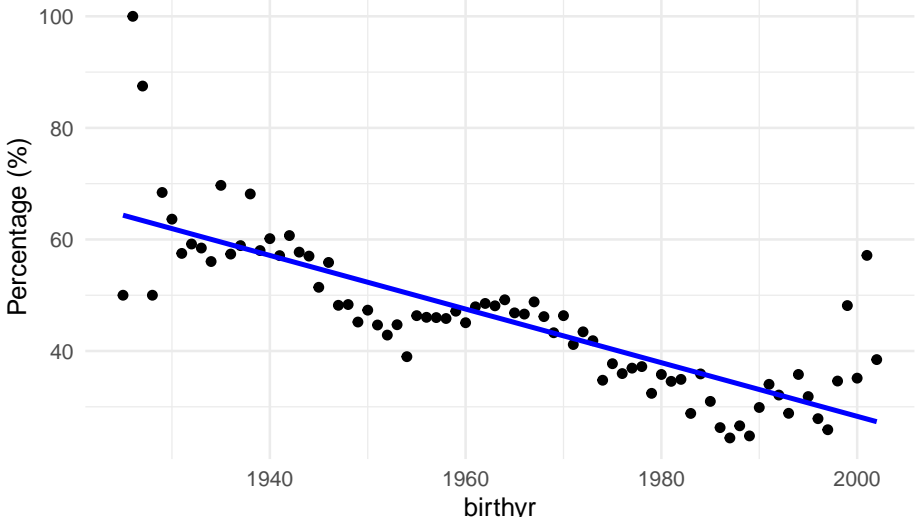
The Destination to Reach with the Data

There could have been many other similar data sets that could have been used for this project for example we could have chosen to look at the census and election data for Canada. However, our group decided that because part of the analysis was done in Wickham et al. (2019b), there were still many other variables that we could explore as we dive further into the 2020 presidential election and try to interpret if there are any correlations between the variables and the result. Our team found it interesting to see all the variables that were collected by the US government and the correlations we saw during the analysis process; where there most definitely was a positive correlation between each variable and the outcome of the votes. Although we are analyzing the 2020 election that has already taken place, the analysis we do in the later sections are believed to apply to the 2024 elections happening this year. This is enough reason for us and the reader to dive into the patterns that exist with this large data set.

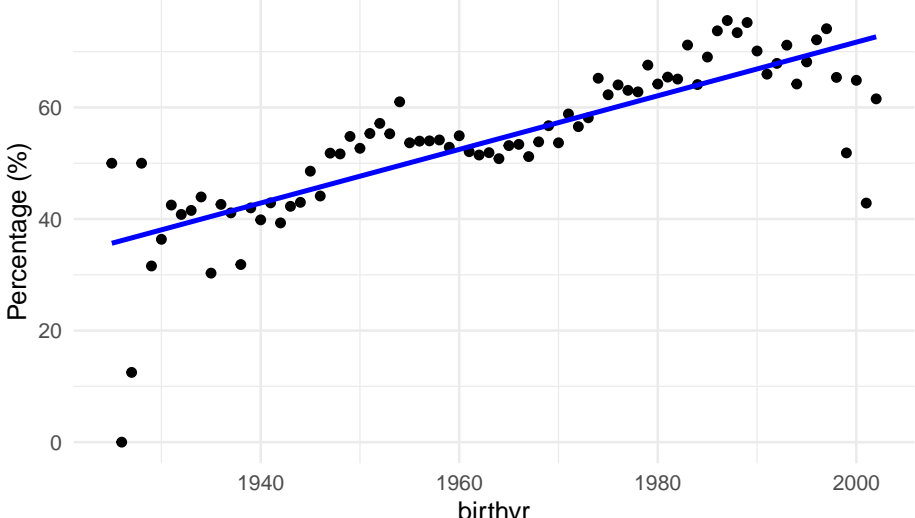
Model 1

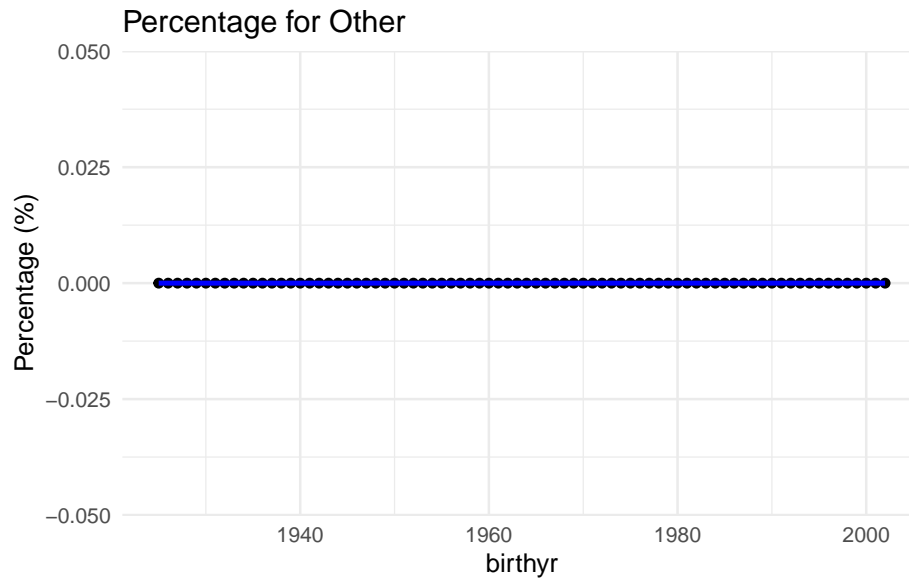
Initially, we filter the data to remove to exclude individuals who did not vote for Hillary Clinton or Donald Trump, due to the minimal votes for Gary Johnson, Jill Stein, Evan McMullin, and all other candidates which are insignificant to this paper. We first analyze the birth year of voters.

Percentage for Hillary Clinton



Percentage for Donald Trump





In this model, we conduct an examination of the relationship between voters' birth year and their gender for the 2016 and 2020 Presidential elections. This analysis is visualized by plotting a histogram that separate female voters on the left and male voters on the right, with voters' birth years measured along the x-axis, which ranges from 1925 to 2002. The y-axis quantifies the voter turnout for the year. For clarity and symbolic representation, the colour blue was chosen to represent the Democratic candidates –Hillary Clinton for the 2016 election, and Joe Biden for the 2020 election, while red was chosen to represent the Republican candidate, Donald Trump, who sought the presidency in both terms. Figure 1 present the distribution of votes in 2016, and Figure 2 presents the data from the 2020 election.

We have modeled the following logistic regression in the graphs:

$$\begin{aligned}
y_i | \pi_i &\sim \text{Bern}(\pi_i) \\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{immigration status}_i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5)
\end{aligned}$$

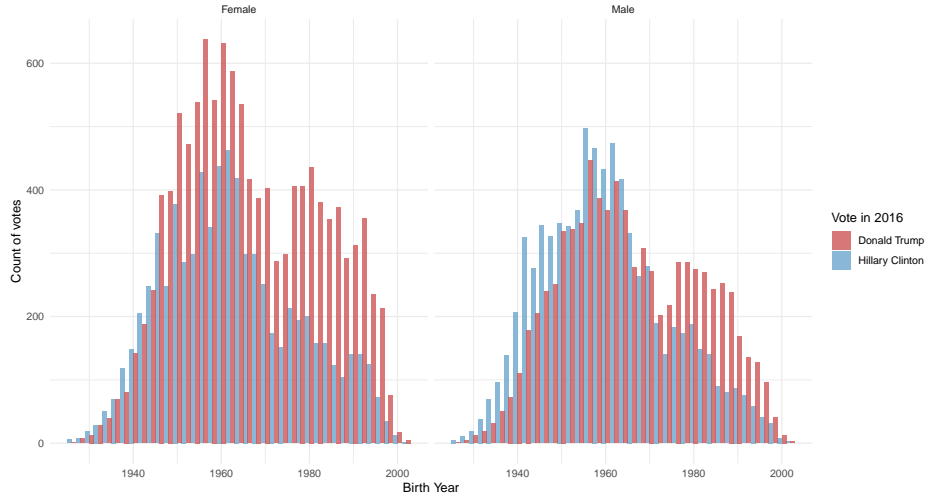


Figure 1: Logistic regression of 2016 US Presidential votes comparing parameters of gender and immigration status

In this model, we conduct an examination of the relationship between voters' birth year and their gender for the 2016 and 2020 Presidential elections. This analysis is visualized by plotting a histogram that separate female voters on the left and male voters on the right, with voters' birth years measured along the x-axis, which ranges from 1925 to 2002. The y-axis quantifies the voter turnout for the year. For clarity and symbolic representation, the colour blue was chosen to represent the Democratic candidates –Hillary Clinton for the 2016 election, and Joe Biden for the 2020

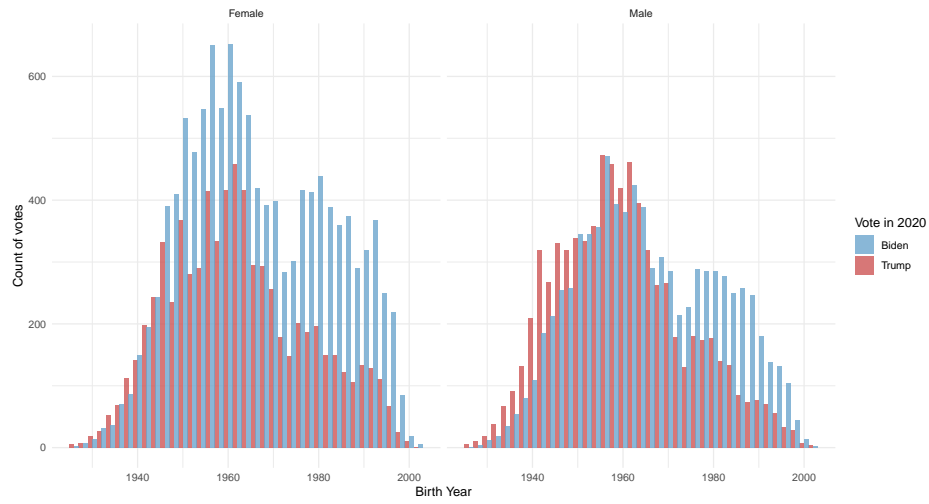
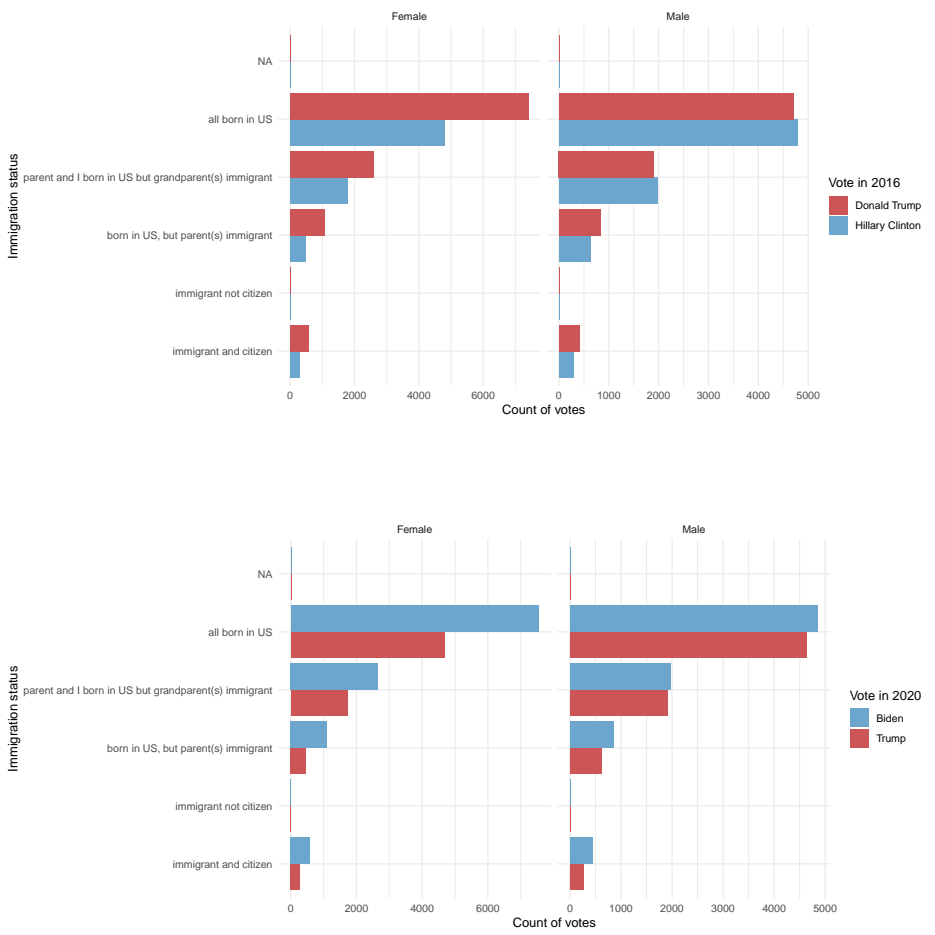


Figure 2: Logistic regression of 2020 US Presidential votes comparing parameters of gender and immigration status

election, while red was chosen to represent the Republican candidate, Donald Trump, who sought the presidency in both terms. Figure 1 presents the distribution of votes in 2016, and (`ifg-genderyr2020?`) presents the data from the 2020 election.

Model 2



Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ .

Results

Discussion

First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Second discussion point

Third discussion point

Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

Additional data details

Model details

Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected by, the data

Diagnostics

Checking the convergence of the MCMC algorithm

References

- Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- , et al. 2019b. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.