US Political Support in 2016 Presidential Election: Preferences Based on Gender and Employment Status*

Daisy Huo

March 11, 2024

This paper uses the methodology of the logistic regression model to analyze the effect of gender and employment status on political preferences in the context of the 2016 US presidential election. For actual data, we will look at the 2016 Cooperative Congressional Election Study (CCES), which contains a nationally representative sample of 64,000 respondents. The findings reveal that women, full-time workers and students demonstrate a higher tendency to vote for Donald Trump over Hillary Clinton. This study serves to provide solid evidence for further analysis, aiming to improve the accuracy of political forecasting and the governance decision-making process.

1 Introduction

Given that this year marks the 60th quadrennial United States presidential election, understanding political support and voter behaviour has never been more important. Former US president Donald Trump recently announced his candidacy for another non-consecutive presidential term. However, the New Yorker once harshly commented on Trump's victory to the presidency in 2016 as "an American tragedy" (Remnick 2016). Taking all of these diverse public opinions into consideration, contemporary political scientists look forward to using the regression model of past elections to forecast the next election cycle. This is known as political forecasting. Continued onward in this paper, we will explore the exact method of the logistic regression in an effort to identify some factors that influenced US political support.

Numerous socio-demographic characteristics, including gender, education, race, employment, marital status, religion, family income, and citizenship, all had a substantial impact on the outcomes of the pre- and post-election polls as well as a majority of online surveys (Ansolabehere,

*Code and data are available at: https://github.com/dai929/US_Political_Support_with_Generalized_Linear_Model.git

Schaffner, and Luks 2017). Moreover, Formplus Blog (2021) emphasized the significance of including profile questions in surveys in order to have a better understanding of sample observations, categorize respondents into different sub-groups and, more importantly, ensure a fairly represented collection process. For instance, we all know that gender and race are sensitive topics, which are used to clarify one's identity and cultural background, while education and employment status are also the two crucial indicators of respondents' socio-economic levels, which may play an essential role on their perceptions of political issues. Hence, by examining the relationship between political support and socio-demographic factors, we are able to provide a more sophisticated understanding of voter behaviour. Specifically, the intersection of gender and employment status has not yet received as much attention as it should and is worth exploring.

Aiming to fill this gap, we carefully reviewed and analyzed the 2016 Cooperative Congressional Election Study (CCES) (Ansolabehere and Schaffner 2017) that was gathered throughout the 2016 US presidential campaign. Through the measurement of our estimand, the average effect of gender and employment status on political support, and the use of the logistic regression model, considerable patterns on presidential preference for Trump were uncovered. The paper proceeds by first outlining all variables, introducing the purpose and aspects of data cleaning and presenting summary statistics between predictors under Section 2. The following Section 3 specifies a generalized linear regression model to estimate the parameters, applying the Bayesian approach with the help of `rstanarm` package (Brilleman et al. 2018). Subsequently, the results of the model and the main findings are discussed under Section 4 and Section 5, utilizing summary statistics and a variety of visual representations of numerical data. Overall, within the historical context of the 2016 US Presidential Election, this analysis maintains a focus on exploring the effect of gender and employment status on political preferences, which contributes to a deeper understanding of the socio-demographic factors shaping electoral behaviour in contemporary political science.

2 Data

The dataset utilized in this paper is the 2016 Cooperative Congressional Election Study (CCES) Common Content (Ansolabehere and Schaffner 2017), taken from Harvard Dataverse. Data was collected, cleaned, and analyzed using the statistical programming software R (R Core Team 2023), with additional support from R packages `arrow` (Richardson et al. 2023), `dataverse` (Kuriwaki, Beasley, and Leeper 2023), `here` (Müller 2020), `janitor` (Firke 2023), `knitr` (Xie 2014), `marginaleffects` (Arel-Bundock 2024), `modelsummary` (Arel-Bundock 2022), `readr` (Wickham, Hester, and Bryan 2024), `rstanarm` (Brilleman et al. 2018), `tidybayes` (Kay 2023), and `tidyverse` (Wickham et al. 2019).

Table 1: Respondents categorized by presidential preference

voted_for	n	proportion
Trump	4892	0.49
Clinton	5172	0.51

2.1 Introduction to Dataset

The 2016 Cooperative Congressional Election Study (CCES) (Ansolabehere and Schaffner 2017), a well-established annual survey of US political choices since 2005, recorded a critical nationally representative sample of 64,000 American adults, whose socio-demographic characteristics and political opinions were captured before and after the election. Besides the dataset, a full guide and codebook to the common content (Ansolabehere, Schaffner, and Luks 2017) was attached as well. There are five components to the 2016 CCES common content: sample identifiers, profile questions, pre-election questions, post-election questions, and contextual data. In addition, the sampling methodology, detailed in Ansolabehere, Schaffner, and Luks (2017), uses YouGov's matched random sample methodology, and the goal is to find the closest matching respondent to the selected member of the target sample.

Choosing the CCES dataset for the year 2016 holds particular significance due to its alignment with the 58th quadrennial United States presidential election, which marks a pivotal moment in US politics. Along with the rest of the globe, the United States witnessed a highly heated, competitive, and dramatic presidential campaign that year. A thorough examination of the factors influencing political support may help to explain the astonishing victory of a celebrity businessman winning over the former US Secretary of State.

2.2 Variables of Interest

To begin with, once `dataverse` (Kuriwaki, Beasley, and Leeper 2023) has been installed and loaded, we then apply the function `get_dataframe_by_name()` to access the 2016 CCES dataset. Variables that are of interest to us are selected using `select()` from `dplyr` (Wickham et al. 2023) which was loaded as part of the `tidyverse` (Wickham et al. 2019).

Using the codebook to guide our data cleaning process, we only want respondents who are registered to vote. These are observations in which the variable "votereg" is 1. For now, we are only interested in those who voted for either Trump or Clinton. The respondent voted for Trump when the variable "CC16_410a" is 1 and voted for Clinton when it is 2. The dataset provides information on gender, and when the variable "gender" is 1, it indicates "male", and when it is 2, it indicates "female". Finally, according to the codebook, the variable "employ" is a variable from 1 to 8, in decreasing level of stability of employment status.

Table 2: Respondents categorized by gender

gender	n	proportion
Female	5477	0.54
Male	4587	0.46

Table 3: Respondents categorized by employment status

employment	n	proportion
Full-time	3825	0.38
Part-time	949	0.09
Temporarily laid off	45	0.00
Unemployed	383	0.04
Retired	3526	0.35
Permanently disabled	670	0.07
Homemaker	533	0.05
Student	133	0.01

Each of Tables 1, 2 and 3 presents counts and proportion of respondents categorized by the three variables, "voted_for", "gender", and "employment". Table 1 displays the voter preference distribution and reveals that the percentage of people who supported Hillary Clinton (51%) and Donald Trump (49%) were nearly equally split. Table 2 presents voter demographics by gender, which is noteworthy that female voters constitute a smaller majority (54%) in comparison to male voters (46%). Table 3 offers information on the employment status of respondents. The data indicates that the majority is made up of full-time workers (38%), followed by retirees (35%). Part-time workers, the jobless, and homemakers comprise comparatively smaller groups.

2.3 Relationship between Variables

The association between voter preferences in the 2016 Cooperative Congressional Election Study (CCES) dataset and the two socio-demographic factors, gender and employment status,

Table 4: The frequency distribution of gender over respondents voted in the 2016 CCES

	voted_for	
gender	Trump	Clinton
Female	48% (2,369)	60% (3,108)
Male	52% (2,523)	40% (2,064)

Table 5: The frequency distribution of employment status over respondents voted in the 2016 CCES

	voted_for	
employment	Trump	Clinton
Full-time	36% (1,757)	40% (2,068)
Part-time	9% (459)	9% (490)
Temporarily laid off	0% (19)	1% (26)
Unemployed	4% (198)	4% (185)
Retired	38% (1,842)	33% (1,684)
Permanently disabled	6% (313)	7% (357)
Homemaker	5% (269)	5% (264)
Student	1% (35)	2% (98)

is depicted in the following two frequency distribution graphs. Table 4 shows a significant gender gap in the voters' choices for president, with males favoring Trump (52%), and women backing Clinton (60%) in the polls. Regarding job status, however, Table 5 reveals that of the two largest groups, full-time employees (40%) and retirees (38%) favour Clinton.

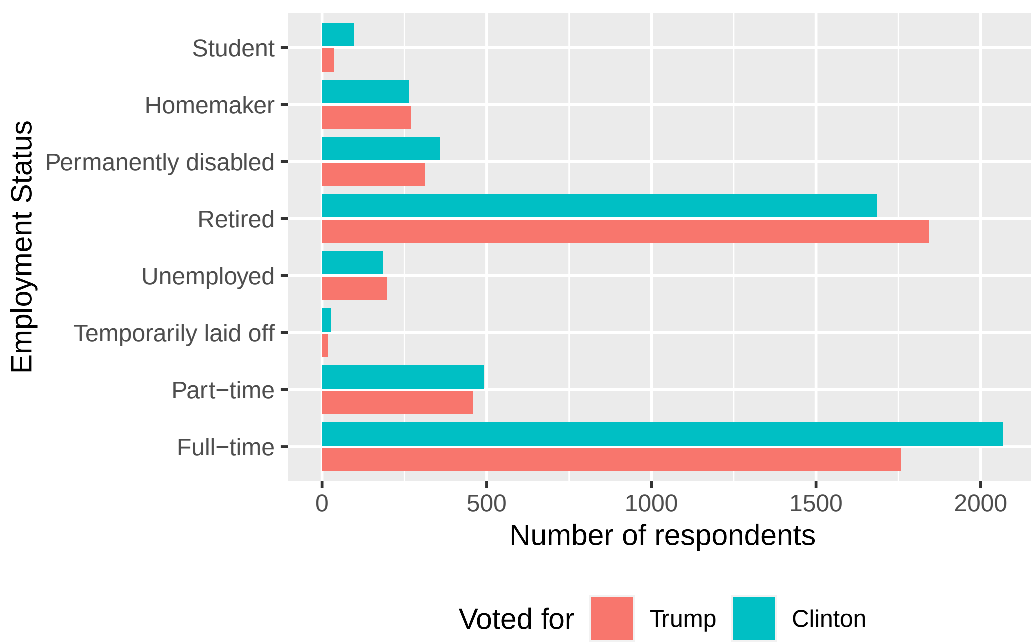


Figure 1: Distribution of employment status, and vote preference, in the 2016 CCES dataset

Figure 1 illustrates a bar chart depicting the distribution of survey respondents in the 2016 presidential election, grouped by their employment status and political allegiance to either

Trump (depicted in red) or Clinton (shown in blue). The data reveals that the majority of full-time workers tend to favour Clinton while a significant portion of retirees, towards Trump. Apart from homemakers and those unemployed individuals, other groups display a stronger preference for Clinton over Trump.

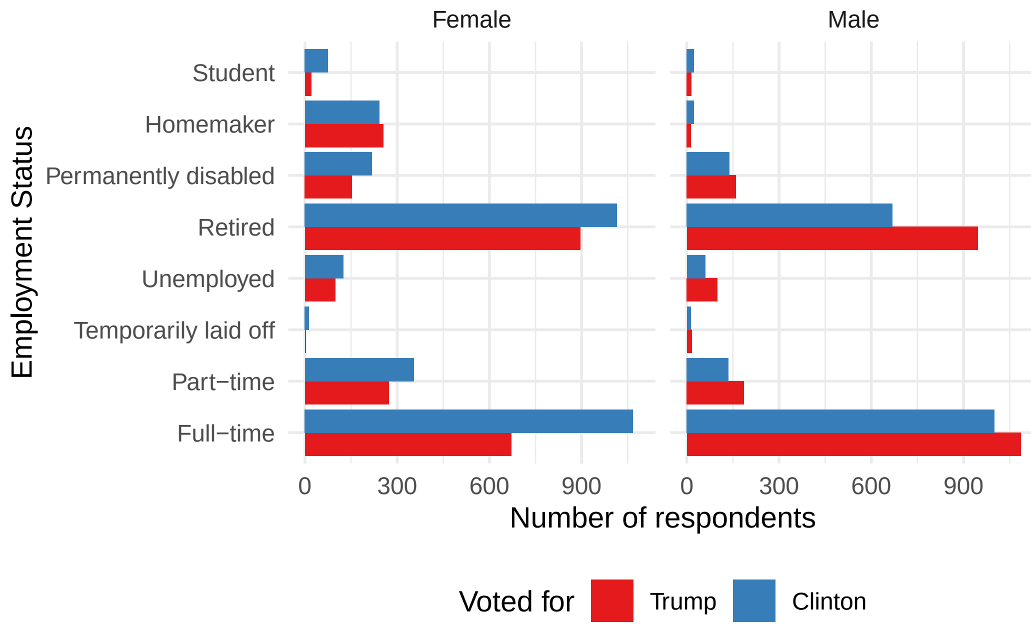


Figure 2: The distribution of presidential preferences, by gender and employment status

Lastly, the distribution of presidential preferences by gender and employment status can be seen in Figure 2. For each level of employment status category, it shows the proportion of respondents who supported Trump (red) and Clinton (blue), broken down into two bar charts by gender. Full-time workers are the biggest group of respondents for both genders, with females preferring Clinton and males preferring Trump. Among females, only homemakers indicated a small percentage of support for Trump, while among males, the population that was permanently disabled, retired, unemployed and part-time and full-time employees all showed a stronger preference for Trump.

3 Model

The model that we are interested in is:

$$y_i|\pi_i \sim Bern(\pi_i)$$

$$logit(\pi_i) = \beta_0 + \beta_1 \times gender_i + \beta_2 \times employment_i$$

$$\beta_0 \sim Normal(0, 2.5)$$

$$\beta_1 \sim Normal(0, 2.5)$$

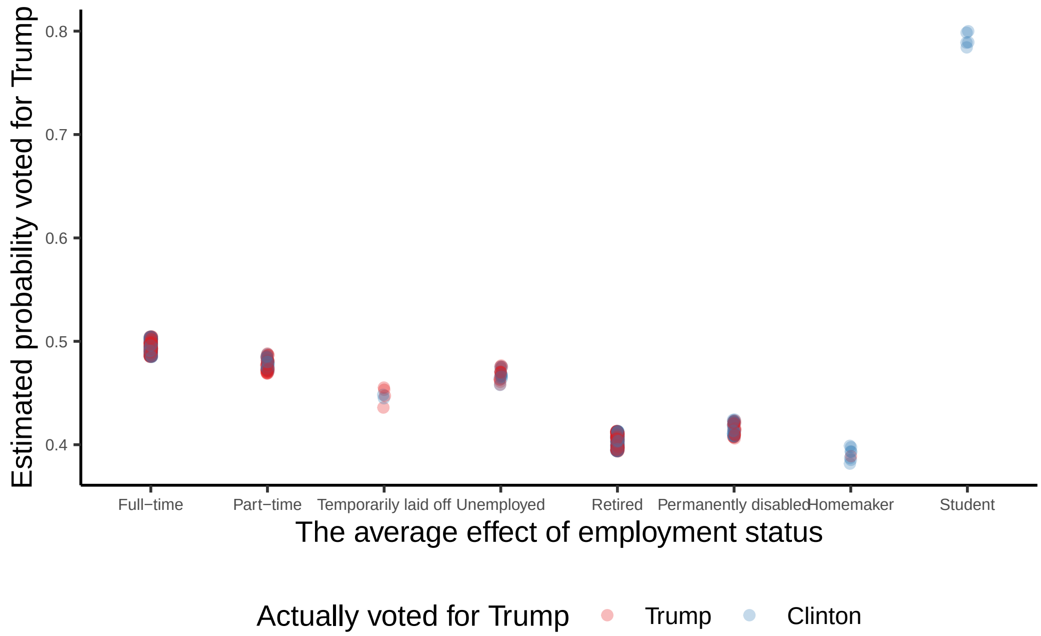
$$\beta_2 \sim Normal(0, 2.5)$$

where y_i is the political preference of the respondent and is equal to 1 if Trump and 0 if Clinton, $gender_i$ is the gender of the respondent, and $employment_i$ is the education of the respondent. We will build this logistic regression model using `stan_glm()` from `rstanarm` (Brilleman et al. 2018). Rather than using the entire dataset, we will fit the model on a random sample of 2,000 observations in order to save runtime.

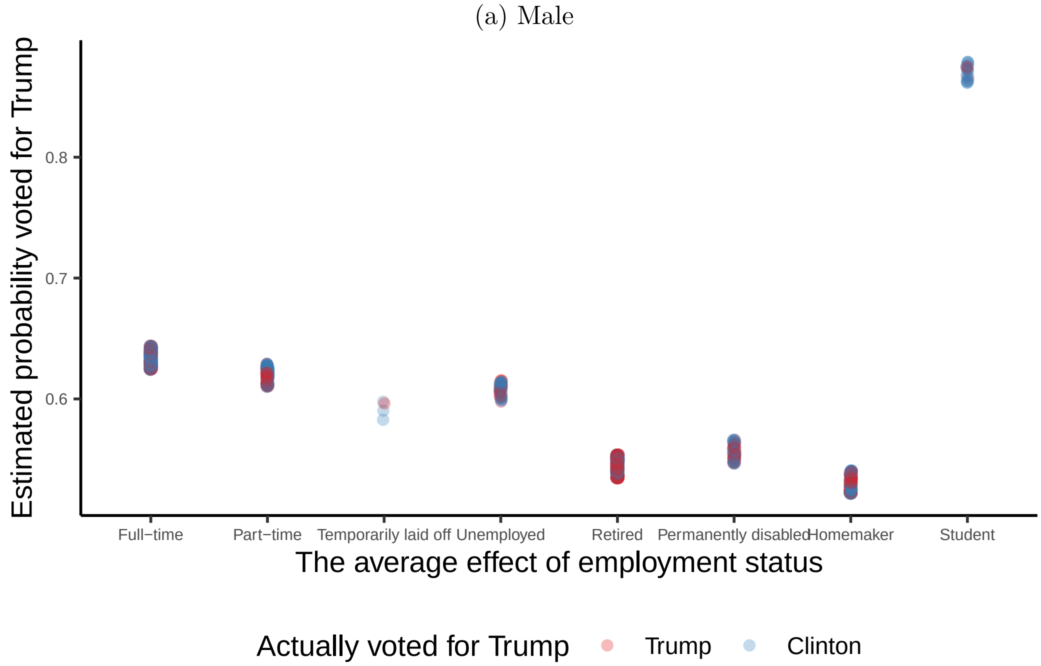
The estimated mean on the gender "Males" is -0.6, which is the average change in the log-odds of respondents who voted for Trump with observing one extra man. This coefficient is negative, meaning a decrease of 0.6. Similar interpretations apply to the rest predictors. Additionally, we translate our estimate into the probability of voting for Trump. We can then add the implied probability for each observation using `predictions()` from `marginaleffects` (Arel-Bundock 2024). The probability that our model implies is graphed in Figure 3 using a scatterplot.

4 Results

Table 6, the model summary, presented below depicts the likelihood of a respondent voting for Trump based on their gender and employment status. The intercept, inferred as a female and full-time employee, suggests a baseline likelihood of 0.550 for the reference group. The predictor of gender "Male" is related to a lower probability of voting for Trump, indicated by a coefficient of -0.570. Except for students showing a strong positive association of 1.361, all other predictors of employment status show a negative association with presidential support for Trump.



(a) Male



(b) Female

Figure 3: Logistic regression probability of whether a respondent voted for Trump or Clinton based on employment status

Table 6: Whether a respondent is likely to vote for Trump based on their gender and employment status

	Support Trump
(Intercept)	0.550
	(0.092)
genderMale	−0.570
	(0.094)
employmentPart-time	−0.064
	(0.164)
employmentTemporarily laid off	−0.200
	(0.628)
employmentUnemployed	−0.109
	(0.245)
employmentRetired	−0.371
	(0.105)
employmentPermanently disabled	−0.321
	(0.206)
employmentHomemaker	−0.421
	(0.211)
employmentStudent	1.361
	(0.561)
Num.Obs.	2000
R2	0.033
Log.Lik.	−1352.747
ELPD	−1362.1
ELPD s.e.	8.3
LOOIC	2724.1
LOOIC s.e.	16.5
WAIC	2724.0
RMSE	0.49

5 Discussion

Based on a thorough data cleaning and examination of the 2016 Cooperative Congressional Election Study (CCES), this paper employs logistic regression model to explore how gender and employment status impact political preferences during the 2016 US presidential election. The findings uncover some notable patterns in voter behaviour. Table 6 reveals that, surprisingly, females, full-time workers and students exhibit a higher probability of voting for Trump, while Figure 2 displays a relatively opposite and varied result in which Trump actually received a larger proportion of votes from males, homemakers, and those retired or unemployed population. Despite some slight confusion, the study contributes to a better understanding of the factors that potentially shape election outcomes and offers some degree of valuable insights for future studies on historical electoral trends, which aim to increase the precision of political forecasting.

One notable takeaway from this paper is the gender gap evident in the 2016 presidential election. Our discovery highlights how gender can be a powerful factor impacting political views and opinions. This stresses the necessity of taking gender dynamics into account when investigating political support and election results. Such gender gap and preferences between women and men appears to become the largest ever since the past 11 presidential contests (Chaturvedi 2016). There are numerate collective reasons causing this phenomenon, including party affiliation, different attitudes toward political issues and candidate appealing. Together, these factors are worth investigating for politicians, political scientists and governance decision-makers, who wish to summarize and predict voter profiles and modify their tactics accordingly.

Another interesting takeaway is the correlation between employment status and presidential preferences, where certain demographic groups demonstrate a stronger tendency toward Trump. Generally, full-time employees and retirees make up the largest two population among the electorate and therefore have greater voice during elections. Reasons behind how various employment status groups behave differently include opinions on the government policies regarding social welfare, perceptions of different parties' ideologies on employment-related solutions.

Undoubtedly, there are some shortcomings in this paper that we need to consider. Firstly, we are missing some observations in our dataset. In particular, we omitted respondents who did not registered to vote and those who did not vote for Trump or Clinton. When building the logistic regression model, we only chose to use 2,000 observations in order to decrease running time. This could potentially create an issue of biased result. In addition, our study typically concentrates on the 2016 US presidential election, so there can be limitations. Whether our results and findings are applicable to the presidential election in different years, or all election scenarios as whole, still requires further studies.

There are a number of directions that future research may pursue. For instance, political scientists may pay attention to the fundamental causes of the gender gap in presidential preferences that we have noted before. Analysis investigating the relationship between employment status and other socio-demographic factors such as family income may help to explain respondents' views on socio-economic issues and their political choices. Furthermore, the logistic regression model utilized in this paper could be improved. All things considered, future studies has the potential to improve the efficiency and precision of regression model on political forecasting and deepen our comprehension of the socio-demographic factors impacting political support.

References

Ansolabehere, Stephen, and Brian F. Schaffner. 2017. "CCES Common Content, 2016." Harvard Dataverse. https://doi.org/10.7910/DVN/GDF6Z.

Ansolabehere, Stephen, Brian Schaffner, and Sam Luks. 2017. *Guide to the 2016 Cooperative Congressional Election Survey*. Harvard Dataverse. https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/GDF6Z0/RK0ONG&version=4.0.

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.

———. 2024. *Marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. https://CRAN.R-project.org/package=marginaleffects.

Blog, Formplus. 2021. "Socio-Demographic: Definition & Examples in Surveys." *Formplus*. Formplus. https://www.formpl.us/blog/socio-demographics#:~:text=Researchers%20depend%20on%20socio%2Ddemographic.

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. "Joint Longitudinal and Time-to-Event Models via Stan." https://github.com/stan-dev/stancon_talks/.

Chaturvedi, Richa. 2016. "A Closer Look at the Gender Gap in Presidential Voting." *Pew Research Center*. https://www.pewresearch.org/short-reads/2016/07/28/a-closer-look-at-the-gender-gap-in-presidential-voting/.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Kay, Matthew. 2023. *tidybayes: Tidy Data and Geoms for Bayesian Models*. https://doi.org/10.5281/zenodo.1308151.

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.

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

Remnick, David. 2016. "An American Tragedy." The New Yorker. https://www.newyorker.com/news/news-desk/an-american-tragedy-remnick-trump-wins-presidency-2016.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2023. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

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. 2024. *Readr: Read Rectangular Text Data*.

https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.