Dynamics of 2016 United States Election persist in 2022 poll*

Allen Uy

March 18, 2024

This study examines the factors shaping support for the Republican and Democratic parties of the United States. By analyzing demographics in a 2022 political poll by the Pew Research Center, we discovered that the Democratic Party is favoured by minorities and higher education, a trend continuing from 2016.

1 Introduction

A poll published by The New York Times in 2016 revealed a significant disparity in support for the two candidates, Donald Trump and Hillary Clinton, based on ethnicity in the state of Florida (Cohn 2016). With the 2024 United States election on the horizon, we use the latest publicly available data from the Pew Research Center to compare whether the same voters are likely to vote Republican or Democrat.

We find that women and individuals with higher levels of education are more likely to vote Democrat. Non-binary and black individuals stand out as the main factors for voting Democrat over Republican.

The subsequent sections follow a structured format. Section 2 outlines the source and variables of interest for our analysis. Section 3 details the construction and methodology of the statistical models used. Section 4 presents the key findings of our analysis, while Section 5 critically reviews the content, addresses the implications of the results, acknowledges model limitations, and suggests potential research directions.

*Code and data are available at: https://github.com/varygx/US2024PollEthnicity.

Table 1: Sample of cleaned dataset

gender	education	ethnicity	party
A woman	Postgraduate	White non-Hispanic	Dem/Lean Dem
A woman	Associate's degree	White non-Hispanic	Rep/Lean Rep
A woman	Associate's degree	White non-Hispanic	Rep/Lean Rep
A man	Some college, no degree	Hispanic	Rep/Lean Rep
A woman	College graduate/some post grad	White non-Hispanic	Rep/Lean Rep

2 Data

The dataset used in this paper was gathered by the Pew Research Center from 5,098 panelists in 2022 ("American Trends Panel Wave 116" 2024) and analyzed using R (R Core Team 2023b) with help from `tidyverse` (Wickham et al. 2019), `arrow` (Richardson et al. 2024), `rstanarm` (Goodrich et al. 2022), `modelsummary` (Arel-Bundock 2022), `testthat` (Wickham 2011), `here` (Müller 2020), `foreign` (R Core Team 2023a), `knitr` (Xie 2023), and `kableExtra` (Zhu 2021). Panelists were surveyed questions on the politics and topics relevant to the time. The cleaned dataset has 4 variables of interest: gender, education, ethnicity, and party. All of these values are self-reported through the survey the panelists completed. Table 1 shows a preview of the cleaned dataset.

3 Model

We investigate one model that might explain political support. A logistic regression model using gender, education, and ethnicity as predictor variables. Logistic regression is well-suited for this problem due to the United States' binary party system. We can estimate the probability of political support based on our predictor variables of gender, education, and ethnicity.

3.1 Model set-up

Define y_i as the political preference of the respondent and equal to 1 if Democrat and 0 if Republican. Then $gender_i$, $education_i$, $ethnicity_i$ are the respective answers of the panelist.

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{education}_i + \beta_3 \times \text{ethnicity}_i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5)
\end{aligned}
$$

We run the model in R (R Core Team 2023b) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

4 Results

Our results are summarized in Table 2.

`genderIn some other way`, which describes non-binary individuals, and `ethnicityBlack non-Hispanic` stand out as they have the largest coefficients of 1.74 and 2.22. Our model predicts that the majority of these individuals are likely to vote for the Democrat party. College graduates and postgraduates are more likely to vote Democrat compared to lower levels of education. Other visible minorities are also predicted to increase the chance of voting for the Democrat party.

5 Discussion

5.1 Conclusion

Through analysis of a 2022 political poll done by the Pew Research Center we were able to create a logistic regression model to predict support for the Republican and Democrat of the United States. Our model indicates that those with higher education levels and visible minorities are more likely to vote Democrat. In particular, black and non-binary individuals are much more likely to vote Democrat.

Table 2: Explanatory model of party affiliation based on gender, education, and ethnicity

	Poll model
(Intercept)	−0.70
	(0.16)
genderA woman	0.41
	(0.06)
genderIn some other way	1.74
	(0.45)
educationHigh school graduate	−0.06
	(0.17)
educationSome college, no degree	0.09
	(0.17)
educationAssociate's degree	−0.03
	(0.17)
educationCollege graduate/some post grad	0.41
	(0.16)
educationPostgraduate	0.64
	(0.16)
ethnicityBlack non-Hispanic	2.22
	(0.12)
ethnicityHispanic	0.81
	(0.08)
ethnicityOther	0.46
	(0.19)
ethnicityAsian non-Hispanic	0.77
	(0.12)
Num.Obs.	4860
R2	0.116
Log.Lik.	−3049.463
ELPD	−3061.6
ELPD s.e.	23.7
LOOIC	6123.1
LOOIC s.e.	47.5
WAIC	6123.1
RMSE	0.47

5.2 Model Limitations

Our model has an R2 of 0.116, meaning approximately 11.6% of variance in the outcome variable is explained by our chosen predictor variables. This value is adequate in the context of social science, indicating that the chosen factors can explain political support.

Logistic regression assumes a linear relationship between our predictor variables and the log-odds of our outcome variable. This assumption may not hold up, especially when our variables are correlated. For example, there is a correlation between ethnicity and age in the sense that equal opportunity is a concern.

5.3 Next Steps

Our analysis should be repeated on a state level to see whether these trends apply nationally or vary from state to state.

References

"American Trends Panel Wave 116." 2024. Pew Research Center. https://www.pewresearch.org/politics/dataset/american-trends-panel-wave-116/.

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.

Cohn, Nate. 2016. https://www.nytimes.com/interactive/2016/09/19/upshot/florida-poll-clinton-trump.html.

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://CRAN.R-project.org/package=here.

R Core Team. 2023a. *Foreign: Read Data Stored by 'Minitab', 's', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...* https://CRAN.R-project.org/package=foreign.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

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/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.