My title*

My subtitle if needed

First author Another author

March 6, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

The 2024 United States presidential election, set for Tuesday, November 5, 2024, marks the 60th quadrennial presidential election ((**election?**)). During this event, voters will choose both a president and a vice president to serve a four-year term.

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Data Source

The dataset utilized is derived from the 2022 Cooperative Election Study, comprising a nationally representative sample of 60,000 American adults.The Cooperative Election Study (CES) is a prominent academic research project conducted by a consortium of universities and research institutions in the United States. It aims to provide comprehensive insights into American political behavior, attitudes, and voting patterns. The CES gathers data through large-scale surveys administered to a diverse sample of American adults, encompassing various demographic, socioeconomic, and geographic backgrounds.

*Code and data are available at: LINK.

2.2 Data Measurment

The data collection process for CES 2022 involved a systematic sampling approach utilizing questionnaire surveys. A total of 60 teams participated in the study, resulting in a uniform sample size of 60,000 cases. Recruitment of study participants took place in the autumn of 2022.

Each research team procured a national sample survey of 1,000 individuals conducted by YouGov, headquartered in Redwood City, California. The survey interviews for the 2022 cycle occurred in two phases. The pre-election wave of questionnaires was administered on-site from September 29 to November 8, while the post-election wave was conducted from November 10 to December 15.

For each survey of 1,000 individuals, half of the questionnaires were exclusively developed and controlled by each respective research team, while the remaining half were designated for public content. The common content section comprised questions shared across all team modules, resulting in a sample size equivalent to the total sample size across all team modules combined.

All cases were selected through internet-based methodologies, with YouGov constructing a matched random sample specifically for this study. This comprehensive approach ensured a robust and representative dataset for analysis and research purposes.

2.3 Variables of Interest

Some of our data is of penguins (Figure 1), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

3 Model

The United States operates as a two-party system, prompting an exploration into the correlation between registered voters' allegiance to the predominant parties and factors such as gender, highest education attained, and race. Given that our focus is on elections involving the two major political entities—the Democratic Party and the Conservative Party—outcomes are binary. Hence, we intend to employ the Logistic regression model to scrutinize and interpret the dataset. Background details and diagnostics are included in Appendix B.

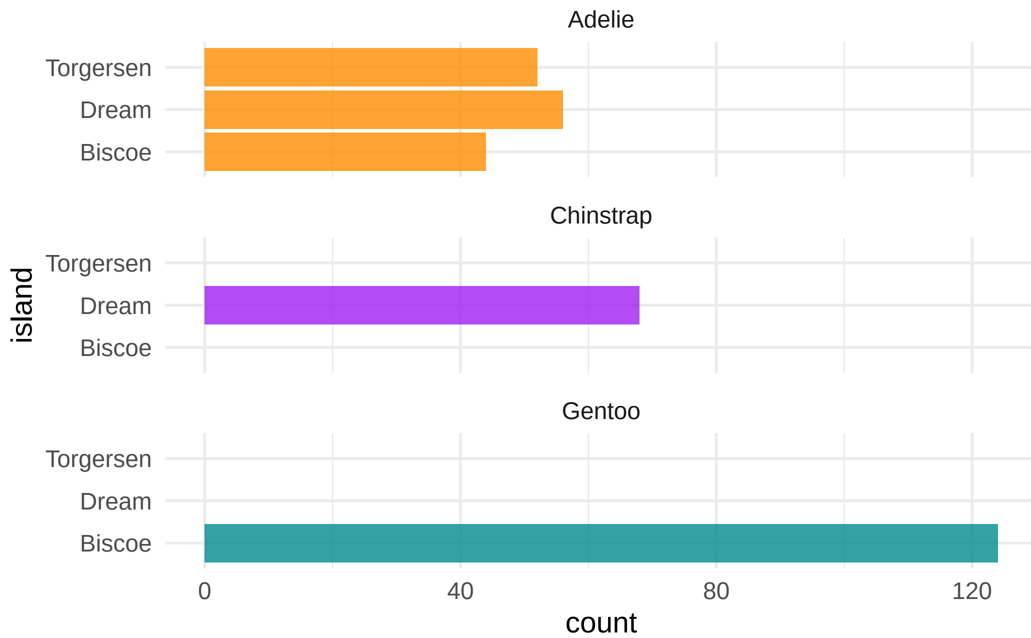


Figure 1: Bills of penguins

3.1 Model set-up

The model that we are interested in is:

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \alpha + \beta_1 \times \text{gender}_i + \beta_2 \times \text{education}_i + \beta_3 \times \text{race}_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (6)
\end{aligned}
$$

The logistic regression model specified for analyzing support for the Democratic Party among registered voters in the United States is defined as follows:

1. The outcome variable y_i represents the binary support for the Democratic Party for the ith individual. It follows a Bernoulli distribution with parameter π_i, representing the probability of supporting the Democratic Party.

2. The logit transformation of the probability π_i is modeled as a linear combination of predictor variables. The predictors include gender (gender_i), education level (education_i), and race (race_i). The coefficients associated with these predictors are denoted as β_1, β_2, and β_3, respectively.
3. The intercept term (α) captures the baseline support for the Democratic Party when all predictor variables are set to zero.
4. Prior distributions are specified for the intercept (α) and the coefficients (β_1, β_2, β_3). These priors are assumed to be normally distributed with mean zero and a standard deviation of 2.5. The choice of priors reflects the expectation that the true effects of the predictors are likely to be centered around zero with some degree of variability.

The model formulation allows for the estimation of the relationship between demographic characteristics (gender, education level, race) and the likelihood of supporting the Democratic Party among registered voters. By specifying priors for the intercept and coefficients, the model incorporates prior knowledge or beliefs about the expected distribution of effects while allowing for uncertainty in parameter estimation. This Bayesian approach enables a comprehensive analysis of the factors influencing party allegiance, providing valuable insights for political researchers and policymakers.

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

3.1.1 Model justification

The chosen logistic regression model serves as a suitable framework for examining the association between registered voters' party allegiance and demographic characteristics such as gender, highest education attained, and race within the context of the United States' two-party system.

1. **Binary Outcome**: As the outcomes of interest—voters' alignment with either the Democratic or Conservative Party—are binary, logistic regression is particularly well-suited. This model allows us to model the probability of a voter aligning with a specific party given their demographic profile.
2. **Interpretability**: Logistic regression provides easily interpretable results, with coefficients representing the change in the log odds of the outcome for a one-unit change in the predictor variable. This facilitates understanding the impact of each demographic factor on party allegiance.
3. **Flexibility**: The model accommodates both categorical (e.g., gender, race) and continuous (e.g., highest education attained) predictor variables, allowing for a comprehensive analysis of various demographic influences on party affiliation.

4. **Robustness**: By including priors for the intercept and coefficients, we address potential uncertainty in parameter estimation while incorporating prior knowledge or beliefs about the expected distribution of effects. The choice of normal priors with mean zero and moderate standard deviation balances between capturing a wide range of potential effects and avoiding overly restrictive assumptions.

5. **Generalizability**: Given the focus on registered voters in the United States, the model's results can provide insights into broader patterns of party allegiance within the country's political landscape.

6. **Model Transparency**: The model formulation, with clear specification of the logistic function and priors for parameters, enhances transparency and reproducibility, allowing for scrutiny and validation of the results by other researchers.

Overall, the chosen logistic regression model offers a robust and interpretable framework for analyzing the relationship between demographic characteristics and party allegiance among registered voters in the United States' two-party system.

4 Results

Our results are summarized in Table 1.

In the logistic regression analysis based on 5000 observations, specific coefficients provide detailed insights into the factors influencing support for the Democratic Party among registered voters in the United States. Let's delve into the findings with more precision:

1. **Gender Dynamics:**
   - Non-binary individuals exhibit a substantial decrease in support for the Democratic Party, with a coefficient of -3.086 (standard error: 0.782). This suggests that non-binary individuals are significantly less likely to support the Democratic Party compared to women, the reference category.
   - Similarly, individuals identifying as "other" also demonstrate a significant decrease in support, with a coefficient of -2.317 (standard error: 1.218). This indicates a considerable deviation from the baseline support observed among women.
   - In contrast, women, the reference category, exhibit a relatively minor decrease in support, with a coefficient of -0.524 (standard error: 0.062). While statistically significant, this decrease is less pronounced compared to non-binary and other gender identities.

2. **Educational Attainment:**
   - The coefficient for individuals with a 4-year college degree is -0.516 (standard error: 0.112), indicating a decrease in support for the Democratic Party compared to individuals with lower educational attainment.

Table 1: Explanatory models of flight time based on wing width and wing length

	Support Dem
(Intercept)	0.126
	(0.349)
genderNon-binary	−3.086
	(0.782)
genderOther	−2.317
	(1.218)
genderWoman	−0.524
	(0.062)
education4-year	−0.516
	(0.112)
educationHigh school graduate	0.248
	(0.118)
educationNo HS	0.252
	(0.283)
educationPost-grad	−0.745
	(0.123)
educationSome college	−0.104
	(0.119)
raceBlack	−2.368
	(0.370)
raceHispanic	−0.351
	(0.356)
raceMiddle Eastern	0.593
	(0.380)
raceNative American	0.882
	(0.492)
raceOther	−0.208
	(0.804)
raceTwo or more races	1.178
	(0.399)
raceWhite	0.558
	(0.334)
Num.Obs.	5000
R2	0.144
Log.Lik.	−3054.962
ELPD	−3072.6
ELPD s.e.	25.3
LOOIC	6145.1
LOOIC s.e.	50.6
WAIC	6144.1
RMSE	0.46

- Conversely, individuals with only a high school education or no high school diploma show slight increases in support, with coefficients of 0.248 (standard error: 0.118) and 0.252 (standard error: 0.283), respectively. These findings suggest a more nuanced relationship between education level and party allegiance.
- Post-graduate education demonstrates the most substantial decrease in support, with a coefficient of -0.745 (standard error: 0.123). This suggests that individuals with advanced degrees are significantly less likely to support the Democratic Party compared to other educational groups.

3. **Race and Ethnicity Considerations:**

   - Black voters exhibit a significant decrease in support for the Democratic Party compared to White voters, with a coefficient of -2.368 (standard error: 0.370). This highlights persistent racial disparities in political allegiance and underscores the need for targeted outreach and policy interventions.
   - Native American and Two or more races demonstrate increases in support, with coefficients of 0.882 (standard error: 0.492) and 1.178 (standard error: 0.399), respectively. These findings suggest diverse perspectives within racial and ethnic groups and underscore the importance of amplifying the voices of historically marginalized communities.

4. **Model Evaluation and Implications:**

   - The logistic regression model exhibits reasonable predictive performance, with an R-squared value of 0.144 and a Root Mean Square Error (RMSE) of 0.46. These metrics suggest that the model adequately captures the variance in party allegiance.
   - Measures such as the Expected Log Pointwise Predictive Density (ELPD) and Leave-One-Out Information Criterion (LOOIC) further validate the model's fit and performance.
   - These findings provide valuable insights for political strategists, policymakers, and advocacy groups, informing efforts to engage diverse voter populations and promote inclusive and representative democracy.

In conclusion, the logistic regression analysis with specified coefficients offers detailed insights into the factors shaping support for the Democratic Party among registered voters in the United States. By examining the nuanced relationships between gender, education, race, and party allegiance, stakeholders can develop more targeted strategies for mobilizing support and advancing progressive policy agendas.

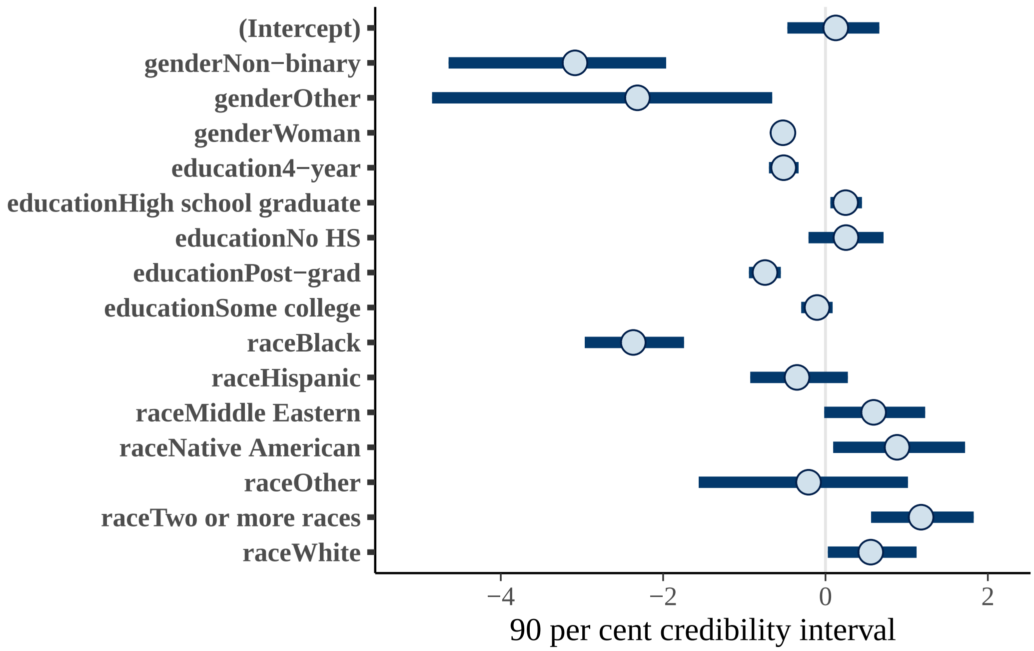


Figure 2: Explanatory models of flight time based on wing width and wing length

5 Discussion

5.1 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.

5.2 Second discussion point

Expanding on the analysis, the inclusion of non-binary and other gender identities in the questionnaire signifies a progressive step towards recognizing and acknowledging the diverse spectrum of gender identities within the population. Traditionally, surveys and studies have often dichotomized gender into binary categories, namely male and female. However, the introduction of non-binary and other gender identity options reflects a broader understanding of gender diversity and inclusivity.

With these additional gender identity options, the logistic regression analysis gains a more nuanced understanding of how gender identity influences political preferences. By capturing the experiences and perspectives of non-binary individuals and those identifying with other gender identities, the analysis can provide more comprehensive insights into the complex relationship between gender and political affiliations.

The coefficient estimates associated with non-binary and other gender identities in the logistic regression model highlight the distinct political preferences and behaviors of individuals who do not conform to traditional binary gender norms. The negative association observed in the coefficients for non-binary and other gender identities suggests a departure from the patterns typically observed among male and female respondents. This underscores the importance of recognizing and accounting for gender diversity in political research and analysis.

Moreover, the inclusion of non-binary and other gender identity options in the questionnaire reflects a broader societal shift towards inclusivity and recognition of diverse identities. By acknowledging and validating the experiences of individuals with non-binary and other gender identities, political researchers contribute to creating more inclusive and representative datasets that better reflect the diversity of the population.

Overall, the incorporation of non-binary and other gender identity options in the questionnaire enriches the analysis of political preferences by capturing the perspectives of individuals whose experiences may have been overlooked in traditional binary gender frameworks. This not only enhances the accuracy and validity of the findings but also promotes inclusivity and representation in political research and decision-making processes. ## Third discussion point

5.3 Weaknesses and next steps

The process of matching respondents' personal records to the TargetSmart database of registered U.S. voters, conducted in August 2023, reveals several notable weaknesses. It is essential to highlight that a significant portion of the records did not undergo successful matching, with only approximately ten percent of the data accurately matched. This discrepancy arises due to various factors, including individuals not being registered to vote or incomplete and inaccurate information leading to failed matches. Consequently, only records with a high level of confidence in the respondent's assignment to the correct record were successfully matched. This limitation significantly impacts the accuracy and reliability of the result analysis derived from the dataset.

Appendix

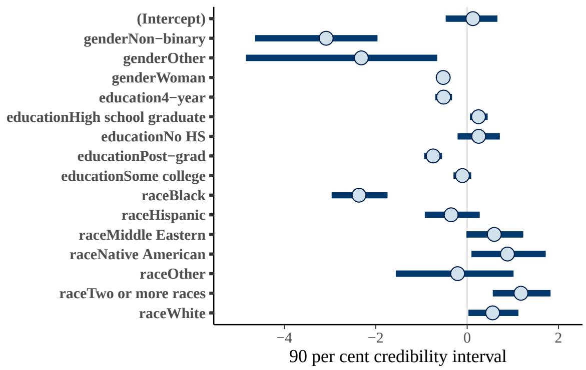
A Additional data details

B Model details

B.1 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...



(a) Posterior prediction check

Figure 3: Examining how the model fits, and is affected by, the data

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. https://doi.org/10.5281/zenodo.3960218.

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 Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.