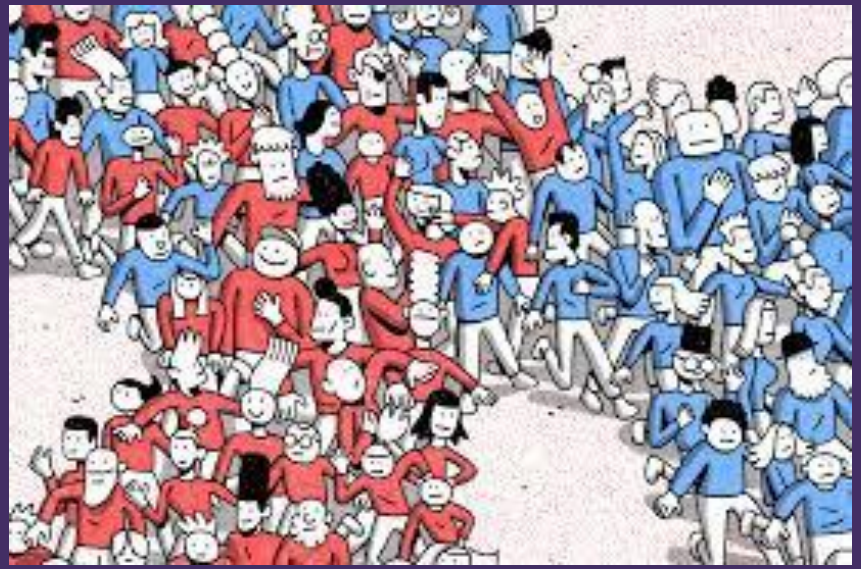
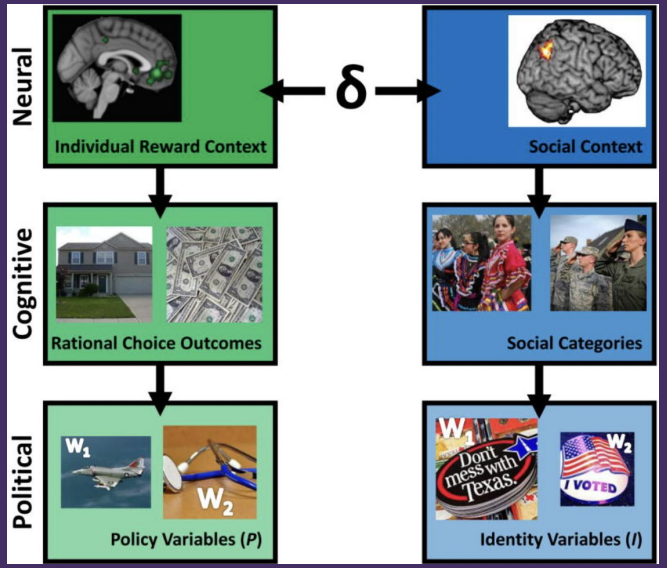
7 Dwarves: Logistic Regression Model to Predict Voter Choice for 2020 Election

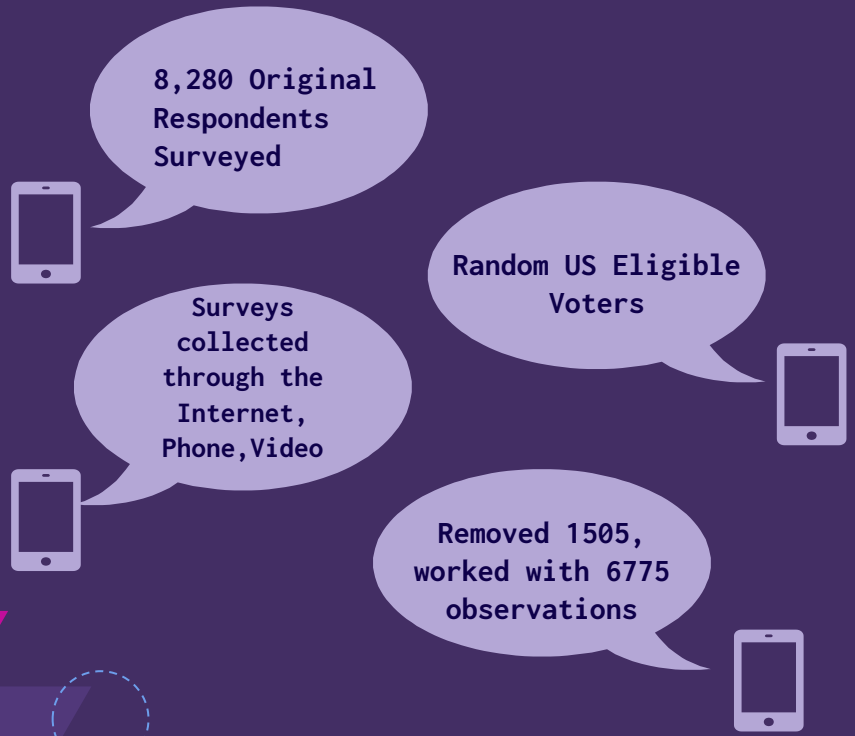
Tali Nam, Morgan Bernstein, Malika Rawal, Lisa Zuo

Topic and Motivation



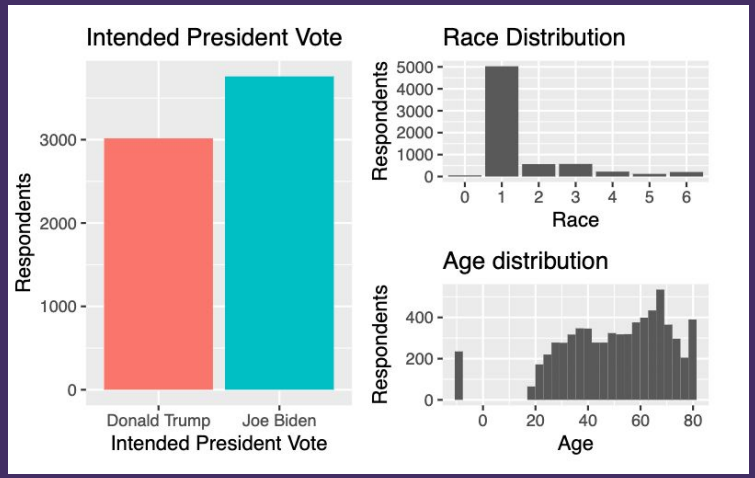


Introducing the Data



- **Source:** American National Election Studies, 2020 National Election
- **Response Variable:** "For whom does the respondent intend to vote for President?"
- We dropped all observations besides those with an answer of "Joe Biden" or "Donald Trump"
- **Predictor Variables:** Favor/Oppose Background Checks for Gun Purchases, Tightening Fed. Budget Spending on Border Security, Self-Identifies Race/Ethnicity, Rural/Urban Area Placement, Respondent Age, Political Alignment

Highlights from the EDA



Response variable:

- Roughly 50/50 intended vote for Trump or Biden, slightly more for Biden

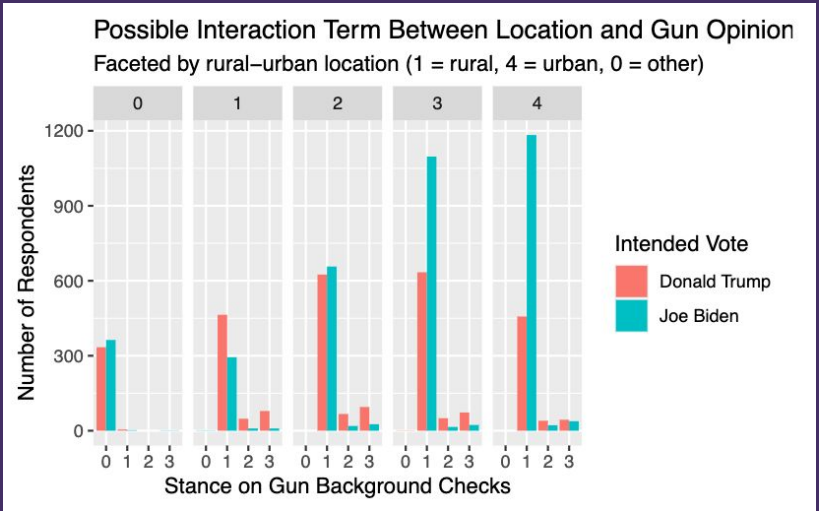
Predictor variables:

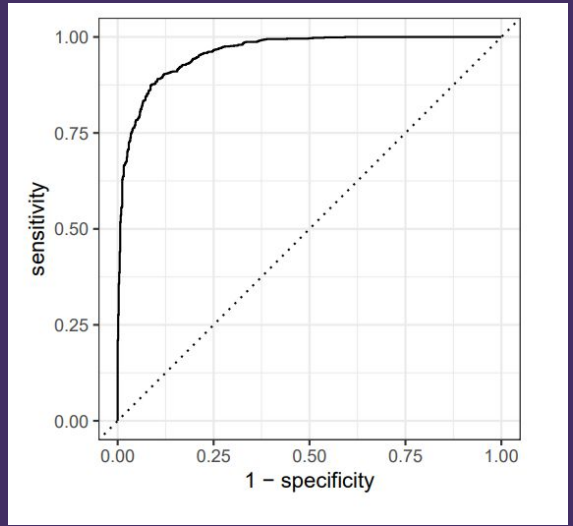
- Majority of respondents were White
- Approximately equal distribution for age, but slightly fewer younger respondents

Possible interaction terms

- Location and stance on gun background checks

Final Model: Interaction Terms + ROC Curve





- Interaction Terms We Considered: Age + Political Alignment, Rural-Urban Identity and Opinion on Gun Background Checks
- Used Speculation / Previous Knowledge about Politics to Infer that these variables could interact

- Area under the ROC Curve: .96 (Very good model fit!)
- Testing Data AUC > Training Data AUC therefore no overfit to the Training Data.
- AIC and BIC were minimized for this model in comparison to models with potential interaction terms.

Interesting findings from the model

- Categories with significant predictors levels: age, political aligment, abortion, federal budget spending on border security, race
- Coefficient of age when predicting which presidential candidate a respondent plans to vote for in the 2020 election

Conclusions + Future Work

CONCLUSIONS

- Both identity and policy opinions were found to be influential when predicting who a person would vote for
- Real-world applications

FUTURE WORK

- More Data
  - Undercoverage and response bias
- Other Elections
  - State and local elections
- Complex Analysis
  - More resources and time

References

- Ali, Shirin. “These Are the Most Difficult States to Vote In.” The Hill. The Hill, September 28, 2022. https://thehill.com/changing-america/respect/accessibility/3665190-these-are-the-hardest-states-to-vote-in/.
- American National Election Studies. 2021. ANES 2020 Time Series Study Full Release \[dataset and documentation\]. July 19, 2021 version. http://www.electionstudies.org/
- Jenke, L., & Huettel, S. A. (2016). Issues or Identity? Cognitive Foundations of Voter Choice. Trends in cognitive sciences, 20(11), 794. https://doi.org/10.1016/j.tics.2016.08.013
- Konnikova, Maria. “Politics and Personality: Most of What You Read Is Malarkey.” The New Yorker, August 23, 2016. https://www.newyorker.com/science/maria-konnikova/politics-and-personality-most-of-what-you-read-is-malarkey.
- “Political Campaign Printing Solutions - Printing Solutions.” Magna IV, March 15, 2021. https://www.magna4.com/printing-solutions/political-campaign/.
- “Premium Photo: HAPPY WINNER! Young Rich Man in Casual Holding Money Dollar Bills with Surprise.” Freepik, July 18, 2021. https://www.freepik.com/premium-photo/happy-winner-young-rich-man-casual-holding-money-dollar-bills-with-surprise_16142404.htm.
- “Running a Political Campaign: What Are the Steps and Operations?” Norwich University Online. Accessed December 10, 2022. https://online.norwich.edu/academic-programs/resources/running-a-political-campaign.
- Snaphunt. “Statistician Job Description.” Snaphunt. Snaphunt, February 25, 2022. https://snaphunt.com/resources/job-descriptions/statistician-job-description.
- See data dictionary