Paper 2*

Gavin Crooks Samarth Rajani

February 15, 2024

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

Table of contents

1 Introduction

Household income is defined as the gross income earned by all members in a household above 15 years of age (SCOTT 2024). Over the years, it has been debated whether household incomes at all affect one's affiliation towards a political school of thought. It is a reasonable hypothesis to assume a sort of relationship between income and voting either Democrat or Republican, as both parties have different economic outlooks thereby affecting incomes differently. Maybe higher income inequality polarizes political leaning further. Therefore, it is in our best interests to study whether the poor vote to improve their quality of life.

In 'Income Inequality and Partisan Voting in the United States', Andrew Gelman, Lane Kenworthy and Yu-Sung Su (Gelman, Kenworthy, and Su 2010) make a case for higher earning Americans voting Republican, whereas Jeff Madrick (**article?**) argues how working-class Americans voted against their interests in voting Republican. Conflicting theories have emerged, and we intend on tackling this issue at hand of whether different income brackets tend to vote differently.

Our data has been obtained from 'The Welfare Effects of Social Media' (**paper?**) . Our code is supported by the following packages (R Core Team 2022) (Wickham et al. 2019) (**her?**) (Xie 2023)

*Code and data are available at: LINK.https://github.com/Crooksyyy/The-Effects-of-Social-Media , Original data available https://www.openicpsr.org/openicpsr/project/112081/version/V1/view

2 Data

2.1 Data Introduction

The data used in this paper is from (cite og paper). The data used in the paper is extremely complicated as it combines numerous data sets to complete their analysis. In this paper, we wanted to simplify the data to determine if their paper had underlying biases within the data. To do this we focused on one of their eleven data sets, baseline dataset as it was the most comprehensive. This data was collected through a facebook ad. The respondents answered a number of questions including questions about income, ethnicity, family, political beliefs and political following. THis is a very useful data set outside the scope of the original paper as it can be analyzed to answer a number of questions. The data set was cleaned to focus on respondents income, race and how closely they follow politics. This provided a data set of approximately 6000 complete responses after removing unfinished responses. This was a substantial decrease from the orignal 24000 responses in the data.

2.2 Income Data

The variable within the data was the income variable. The questionnaire included a categorical value for the the household income of respondents. The data can be visualized in (**figure1?**). This graph illustrates that the least number of households make greater than 100,000 USD, below 20,000USD or preferred not to answer. This graph also shows what the categorical option were for the respondents to the questionnaire. (**figure1?**) closely resembles the expected distribution of USA household income. As expected in any income distribution the majority of responses fall within the average income ranges of the USA, between 20,000USD and 100,000USD. These factors indicate that the data set has an accurate representation of household income.

2.3 Ethnicity Data

The second question of the data that we have included in our analysis is the ethnicity of the respondents. This again is a categorical variable, that the individual self identifies their own ethnicity. The response options included Asian or Pacific Islander, White / Caucasian, Hispanic, Black or African American and other. (**race_dist?**) shows the percentage of respondents in each with the overwhelming majority of responses being Caucasian at nearly 70%. This is actually less than the most recent estimates by the United States government which estimate over 75% of the population is Caucasian (cite us gov). The data also has an over representation of Asian and Native Americans. This results in an under representation of Hispanic and African American populations.

https://www.census.gov/quickfacts/fact/table/US/PST045222

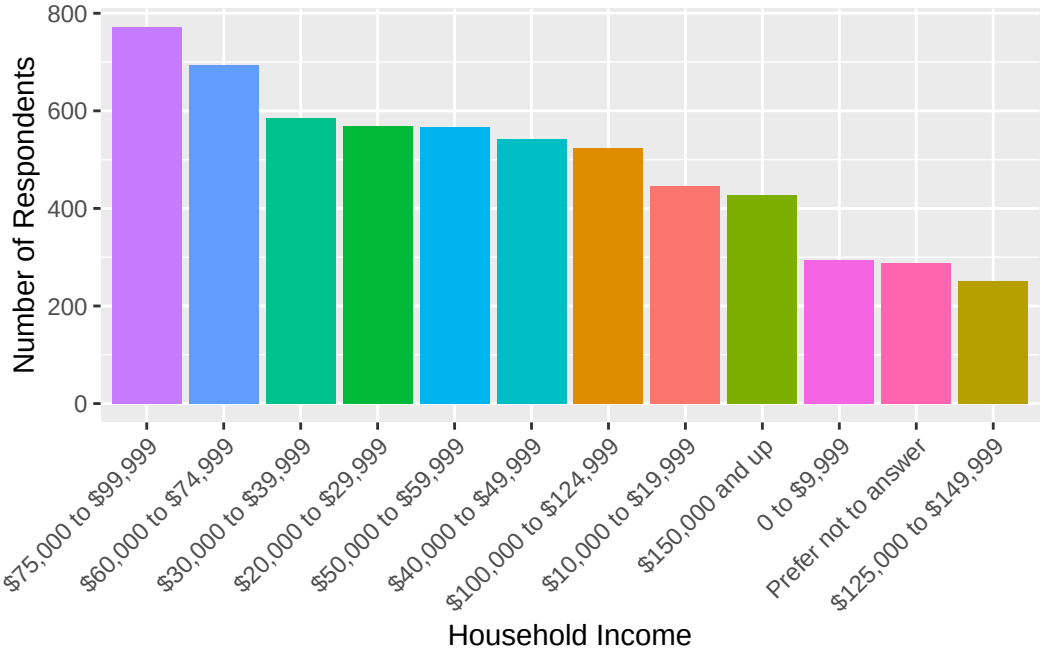

Figure 1: Distribution of Income from Responses in a facebook ad

Table 1: Percentage of each Ethnicity from Responses in a facebook ad

Ethnicity	Percentage of Responses
American Indian or Alaskan Native	0.7554138
Asian or Pacific Islander	13.5806614
Black or African American	6.0936713
Hispanic	8.0577472
Other (please specify)	2.5851939
White / Caucasian	68.9273124

2.4 Politics Data

The third variable in the data that is included in our analysis is a variable of respondents self identifying how closely they follow politics. This is another categorical variable measured as Not at all closely, Somewhat closely, Rather closely and Very closely. This variable faces many problems as this categorical scale is not consistent across respondents. To be specific we mean someone who identifies as someone who does follows Not at all closely can be following politics more than someone who identifies as Somewhat closely. This is a measurement issue within to the questions asked in the survey and all self identifying variables in general. (**figure2?**)

shows the quantity of respondents in each group. The most common response is that they follow somewhat closely and the other responses are relatively even.

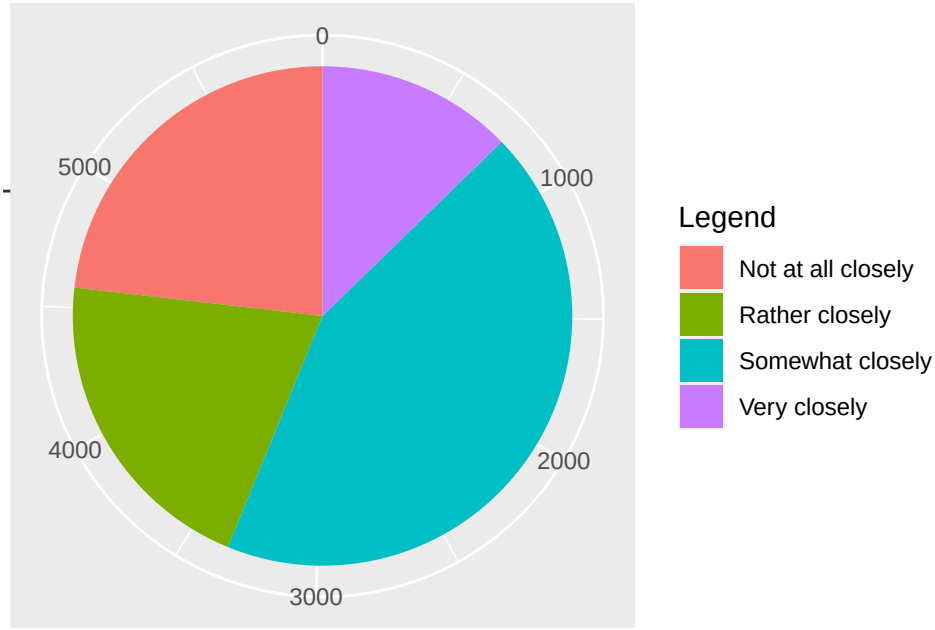

Figure 2: How Closely People Follow Mr.Donald Trump from Responses in a facebook ad

2.5 Missing Data

3 Results

```
# Create a table of proportions
income_follow_table <- table(cleaned_data$follow_trump, cleaned_data$income_group)
income_proportions <- prop.table(income_follow_table, margin = 1) * 100

table1 <- tibble(income_proportions, colnames(c("Income Group", "Follows Trump" )))
show(table1)
```

```
# A tibble: 4 x 1
  income_proportions[,"$0 to $19,9~1 [,"$100,000 and up"] [,"$20,000 to $49,99~2
                               <dbl>                <dbl>                  <dbl>
1                               9.12                 16.9                   28.1
```

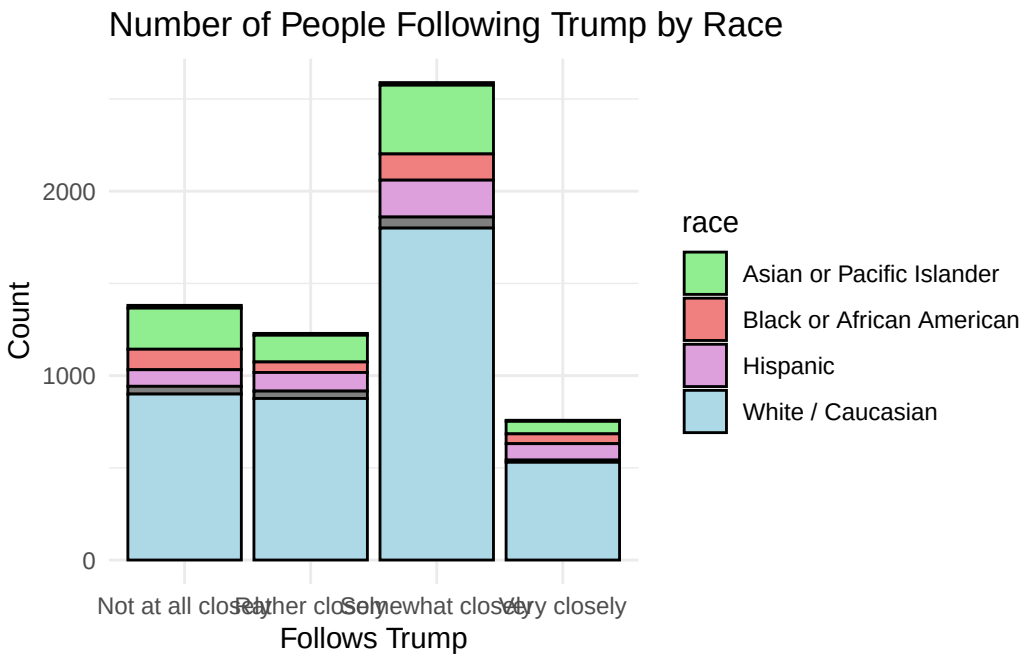

Figure 3: Number of Respondents who follow Donald Trump at different levels by Race

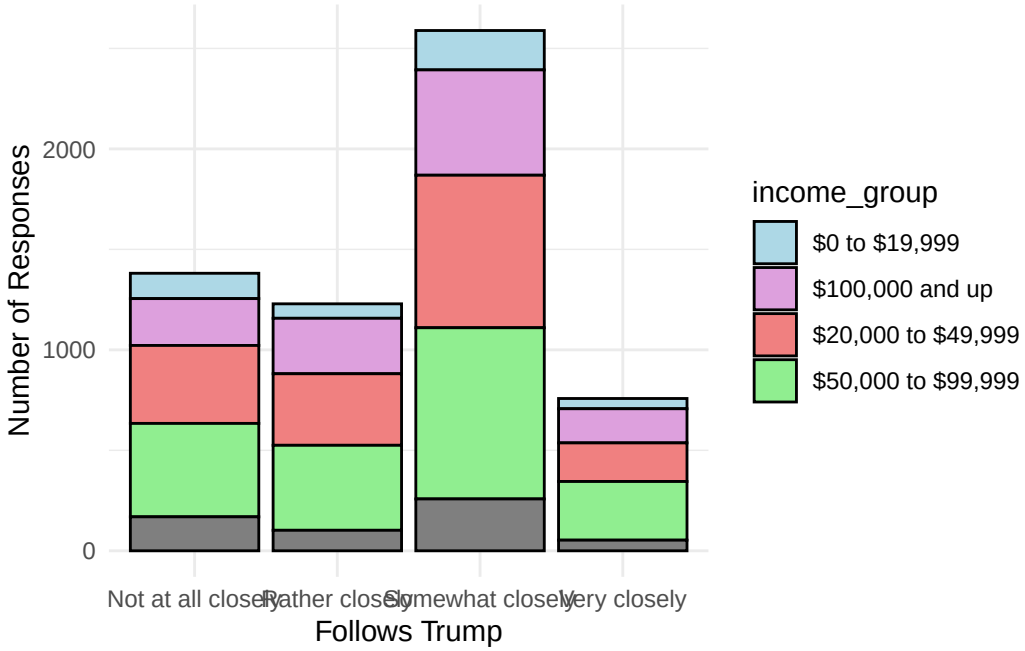

Figure 4: Number of Respondents who follow Donald Trump at different levels by Household Income

```
2                              5.86                 22.5                 29.0
3                              7.57                 20.2                 29.3
4                              6.73                 22.4                 25.3
# i abbreviated names: 1: income_proportions[,"$0 to $19,999"],
#   2: [,"$20,000 to $49,999"]
# i 1 more variable: income_proportions[4:5] <dbl>
```

This papers goal was to identify if lower income household voted against their own interest. To understand this relationship we used the variable for how closely a respondent follows Mr. Donald Trump as our measurement of subscription to Republican ideas. Using this measurement, we organized the proportion of individuals by income class to how closely they follow Mr. Donald Trump in Figure **??**. This graph shows that a proportional amount of each income class follows Mr. Donald Trump at similar levels across all income levels. Specifically, we mean the percentage of individuals follow Mr. Donald Trump at different levels is the same no matter the income class. This means we can not conclude that income class impacts how closely individuals follow Mr. Donald Trump, and therefor, cannot conclude that different income levels subscribe to republican ideas more than the other.

To further our analysis we computed the same graph however organized by race not income Figure **??**. This resulted in a similar result as race is proportional between all levels of following Mr. Donald Trump. Therefore, consistent across races at to following republican rhetoric. Again, this means the same percentage of people that follow Mr. Donald Trump at different levels is the same for each race. Obviously, this is more difficult to conclude for minorities as their representation within the data set is so small as stated in Section **??**

Overall, the results of this analysis were inconclusive to measure how income impacts individuals propensity to follow republican rhetoric. There are many reasons this could be true and as stated in the (**intro?**) there are multiple schools of thought previously studies on the topic.

4 Discussion

like to incluse variables like state etc data set too small/ incomplete for that ## First discussion point {#sec-first-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.

4.1 Second discussion point

4.2 Third discussion point

4.3 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

References

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review*. https://doi.org/10.1257/aer.20190658.

Gelman, Andrew, Lane Kenworthy, and Yu-Sung Su. 2010. "Income Inequality and Partisan Voting in the United States." *Social Science Quarterly*. [University of Texas Press, Wiley]. http://www.jstor.org/stable/42956457.

Madrick, Jeff. 2020. "Why the Working Class Votes Against Its Economic Interests." *The New York Times*, July. https://www.nytimes.com/2020/07/31/books/review/the-system-robert-reich-break-em-up-zephyr-teachout.html.

Müller, Kirill. 2020. "Here: A Simpler Way to Find Your Files." https://CRAN.R-project.org/package=here.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

SCOTT, MICHELLE P. 2024. "Household Income." investopedia. https://www.investopedia.com/terms/h/household_income.asp.

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*. 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/.