

UC Merced

Cognitive Science Capstone Projects

Title

Exploratory Data Analysis: Bias in The Media

Permalink

<https://escholarship.org/uc/item/869055tp>

Author

Perez, Dennis Frank, Jr.

Publication Date

2022-06-16

Data Availability

The data associated with this publication are available upon request.

UNIVERSITY OF CALIFORNIA, MERCED

Exploratory Data Analysis: Bias in The Media

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of
Science

in

Cognitive and Information Sciences

by

Dennis Frank Perez Jr.

Committee in charge:

Professor David Noelle, Chair
Professor Daniel Hicks
Professor Lace Padilla

2022

Copyright

Dennis Frank Perez Jr., 2022

All rights reserved

The Thesis of Dennis Frank Perez Jr. is approved, and it is acceptable in quality and form
for publication on microfilm and electronically:

Daniel Hicks

Lace Padilla

David Noelle, Chair

University of California, Merced

2022

Table of Contents

List of Figures	v
Acknowledgements.....	vi
Abstract	vii
Introduction	1
Methods	3
Findings	7
Discussion	8
Bibliography	9

List of Figures

Figure 1: Bar plot displaying the counts of political party affiliations among participants.....	4
Figure 2: The proportional likelihood of fact-checking social media information per party. Each likelihood adds up to 100 percent for each group, so that likelihoods are relevant to the group itself.....	5
Figure 3: The proportional likelihood of fact-checking online information per party. Each likelihood adds up to 100 percent for each group, so that likelihoods are relevant to the group itself.....	6
Figure 4: A density plot of the time spent taking the survey per political party group. The distributions are also color coded for how likely they are to consume political information online. Participant times below 700 seconds and above 5,000 seconds were omitted as outliers, or people who did not complete the survey properly.....	7

Acknowledgements

First and foremost, I would like to thank my wife, Yuliana, and my son, Benny. They give my life purpose, and nothing would have value without them. Their love and support have saved me from the depths and given me the confidence to strive for the very best life has to offer. I dedicate all of my work to you both.

Next, I would like to thank my family. We have been through the unimaginable, but we have never been broken. I hope my accomplishments serve as tribute to all the love and sacrifices my mother, father, brothers, brothers in arms, sisters, cousins, and grandparents provided.

I would also like to thank my committee members, Dr. David Noelle, Dr. Daniel Hicks, and Dr. Lace Padilla, for their continued support throughout my graduate education. I would also like to thank the Cognitive and Information Science department as a whole. The mentorship and guidance of this program are precisely why I was able to navigate through this difficult time and why I have grown as a researcher. I will continue to grow based on the fundamentals they have instilled in me.

Lastly, I would like to thank my classmates. I would not have succeeded in graduate school if it weren't for our collaborative efforts to succeed.

Abstract

Exploratory Data Analysis: Bias in The Media

by Dennis Frank Perez Jr. for the partial satisfaction of the requirements for the degree of
Master of Science in Cognitive and Information Sciences, University of California,
Merced, 2022
Dr. David Noelle, Chair

The sharing of biased information has become an increasingly pervasive issue. This is quite dangerous considering how the exchange of information can influence perceptions, decision-making, and, most importantly, how well we coexist. Accordingly, as our access to information and interactions grow, we must reestablish control for how information is shared and increase accountability for those sharing information. Unfortunately, this is seemingly impossible given the scale of interactions and the complexity of information passed around. Thus, researchers in an experiment by Westmark et al., suggest that gauging people's ability to detect biased information at the lower group level is where to start to accomplish these initiatives. The present work is an exploratory data analysis of the results from a survey deployed during this experiment, which was used to assess participants' ability to detect bias correctly. The analysis was designed to provide researchers with different perspectives of the original hypothesis to consider. Although no significant relationships were found, comparisons based on gender and party affiliation displayed interesting information about how well these groups deal with the media information they receive and pass on to others.

Introduction

Background

Essentially, the fundamental aspects of our existence are grounded in the information provided to us. We are molded by our upbringing, schooling, and environmental experiences that eventually contribute to establishing beliefs, morality, and behaviors. Eventually, we align our perspectives of the world with like-minded individuals to create groups.

Since we derive from infinitely complex situations, it is natural for groups to differ in opinion. Nevertheless, peaceful coexistence is possible as long as shared information is objective and as sensitive to context as possible. This becomes increasingly more important regarding concepts that may contradict principal foundational beliefs of morality.

The dissemination of biased information is a critical threat to this premise. Entman (2008) notes how bias prevents objectivity and “introduce[s] or raise[s] the salience or apparent importance of certain ideas, activating schemas that encourage target audiences to think, feel, and decide in a particular way”. This type of convoluted information is incredibly dangerous and often a source of tension that has greater ramifications.

To illustrate, consider how some groups use biased information to gain followers and establish social power. Entman (2008) mentions larger possible implications that can exacerbate current issues regarding imbalances of power in a democracy where equity is supposed to be the foundation. A prime example is how large organizations, like news media outlets, take advantage of shared platforms and use biased information to push agendas and influence their audiences. Unfortunately, this has become quite a prevalent problem, especially regarding politically driven topics. For example, recent pandemic and presidential election events have erupted into riots, political movements for racial inequality, and protests against health mandates.

The number of global connections and access to information has exponentiated in the wake of technological advancement and the fast-moving transition from a physical to a virtual world. We now have access to stories, situations, concepts, people, and general information that we have never had before. The quality of sharing information is important, now more than ever.

Accordingly, Westmark et al. (2021) developed an experiment to address this issue. Acknowledging the complexity of bias in larger groups, researchers thought it best to minimize the area of interest to lower-level group analyses. Since controlling bias at the larger group level is a task beyond the scope of a smaller project, researchers felt that an equally important question would be to explore lower-level groups’ ability to detect bias.

Researchers deployed a survey to gauge participants' ability to detect biased information to accomplish this initiative. First, participants were shown various politically charged news articles with a mix of low, medium, and high levels of bias. Afterward, they were asked to determine the level of bias for each article. Finally, when the study was finished, researchers judged each participant's ability to detect bias based on how the research team rated the news articles internally.

Ultimately, there were no final results to report as the experiment was a part of a course project, and it is still being determined whether or not the experiment will continue. In any case, researchers anticipated a need to look beyond the original research question to gain further insights. Thus, they asked if I would be able to provide an engineering perspective to find possible unexplored relationships within the survey response data.

Follow up Analysis

As a first attempt at alternative analysis of the original study, I created a low-level data exploration of the survey results using the R programming language. However, instead of focusing on one's ability to detect bias, this project focused on how well one deals with information. The concept was based on the notion that this was an equally important aspect of quality information exchange and eliminating bias (Perez, 2021).

To explore this phenomenon, I developed three key aspects to consider: First, one should consider where people receive their information. Next, one should consider the quality of the information people share. Lastly, one must observe whom people are sharing their information with. With these three questions answered, one would theoretically understand how well individuals receive and share information.

Accordingly, the data analysis considered variables relevant to the three considerations that were developed to understand how well one deals with information. For example, the amount of time spent on social media was used to explore where participants spend their time and where most of the information is coming from. Whether or not participants fact-checked information they came across was used to gauge the quality of information they were accepting. To determine whom participants share information with, I explored responses for whether or not they discuss current events with family and friends. Finally, the data was subset by gender for group comparisons.

In short, the sample sizes for gender were skewed as there were far more female participants than males. However, general analyses displayed that both male and female participants spent roughly the same amount of time on social media, but female participants were more likely to fact-check the information they came across and were also more likely to discuss current events.

Naturally, there were no significant results to report since the explorations were solely meant to provide researchers with insights into basic relationships within the data. However, researchers hoped that further exploratory data analyses would guide future endeavors. Researchers are currently working on additional sub-analyses to guide the study in future directions.

Current Analysis

The original study's purpose was to gauge participants' ability to detect bias with the hope that improving bias detection would resolve part of the issue of the exchange of biased information. A follow-up analysis was developed to increase the versatility of the original study's research question. The follow-up analysis explored alternative relationships within the participant pool data in search of informative trends of how well one shares information based on gender. To complement the group comparisons of gender, the current analysis is an additional attempt to provide another angle of participant data that will inform the original study's research question.

Namely, the present work will explore how participants' political party affiliations relate to the type of online content they interact with, whether or not they fact-check information, and how engaged they were in the survey. Similar to previous work, the aspects explored here may provide insight into what kind of information a participant receives and how well they deal with it. This analysis also includes information on how long was spent on the survey to inform each participant's engagement, which may also help future research filter out unreliable responses. All data explorations were conducted using the R programming language to produce the observations discussed in this analysis.

Methods

Dataset

The data was extracted from the original study's Qualtrics survey results collected from undergraduate students at the University of California, Merced. The dataset contained each participant's judgment of bias on the news articles presented during the study and additional personal information that researchers would use to contextualize their responses, such as political involvement, social media use, whether they fact-check information, and demographics. When data was extrapolated from Qualtrics, there were only 101 participants in the study.

The original dataset consisted of 85 columns for each question asked in the survey and 101 rows that housed each participant's response. Columns ranged from continuous to discrete values that varied with each question type. Although, most information is categorical since most response options were given on a scale range of "not at all, a little, somewhat, very, and extremely".

Data cleaning

A subset of the data was created for variables relevant to the current research question in the present work. The columns selected were party affiliation, how much time was spent taking the survey, how likely participants were to fact-check information online and on social media, and how likely participants were to click political information. Then, the dataset was cleaned of arbitrary "-99" values which indicated an incomplete response. Lastly, the data was cleaned of rows containing survey question titles and rows with participant ID information. After all the cleaning was finished, the dataset had 72 rows of participant information and five columns for each survey question or variable of interest.

Analysis Visualizations

Given the small size of the dataset, I understood that skewed information could interfere with future data observations. Thus, before conducting any relational analyses, I checked to ensure that the group variable being compared was evenly distributed

throughout the data. The first visual depicted below confirms this intuition.

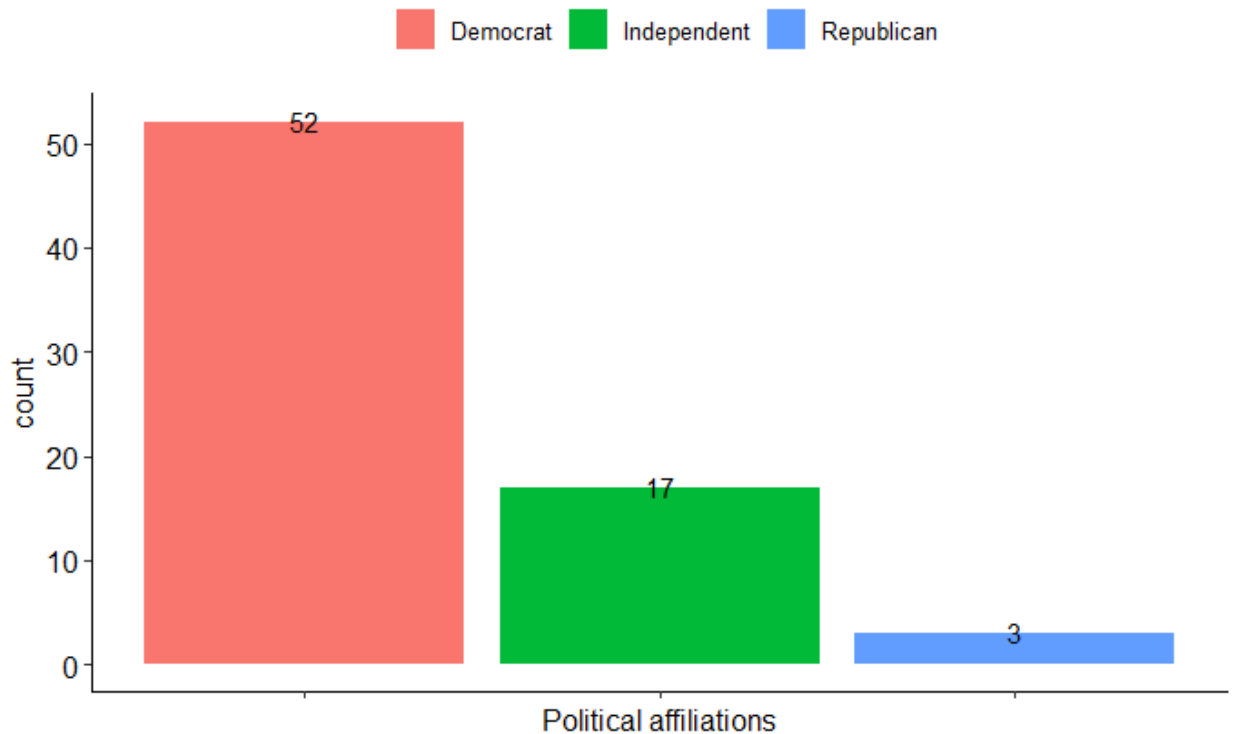


Figure 1: Bar plot displaying the counts of political party affiliations among participants.

As one can see, the distribution of political party affiliations is quite skewed towards the Democratic and Independent parties. Participants who identified as Republican only accounted for about four percent of the total participant pool. Although the sample sizes were quite different, I decided to keep the Republican data purely as an additional observation.

Accordingly, it should be noted that the Republican data will not be a relative comparison to the Independent and Democratic parties since the sample size was so small. To account for the difference in sample sizes in each group, I attempted to make all subsequent analyses as proportional as possible. Even after adjusting for proportions, the Republican data still did not seem to benefit the purpose of this analysis, but it still seemed relevant enough as additional information for visual purposes.

Naturally, the first analyses used proportions to explore the relationship between political party affiliation and how likely the participant was to fact-check information

they come across on social media and the internet in general.

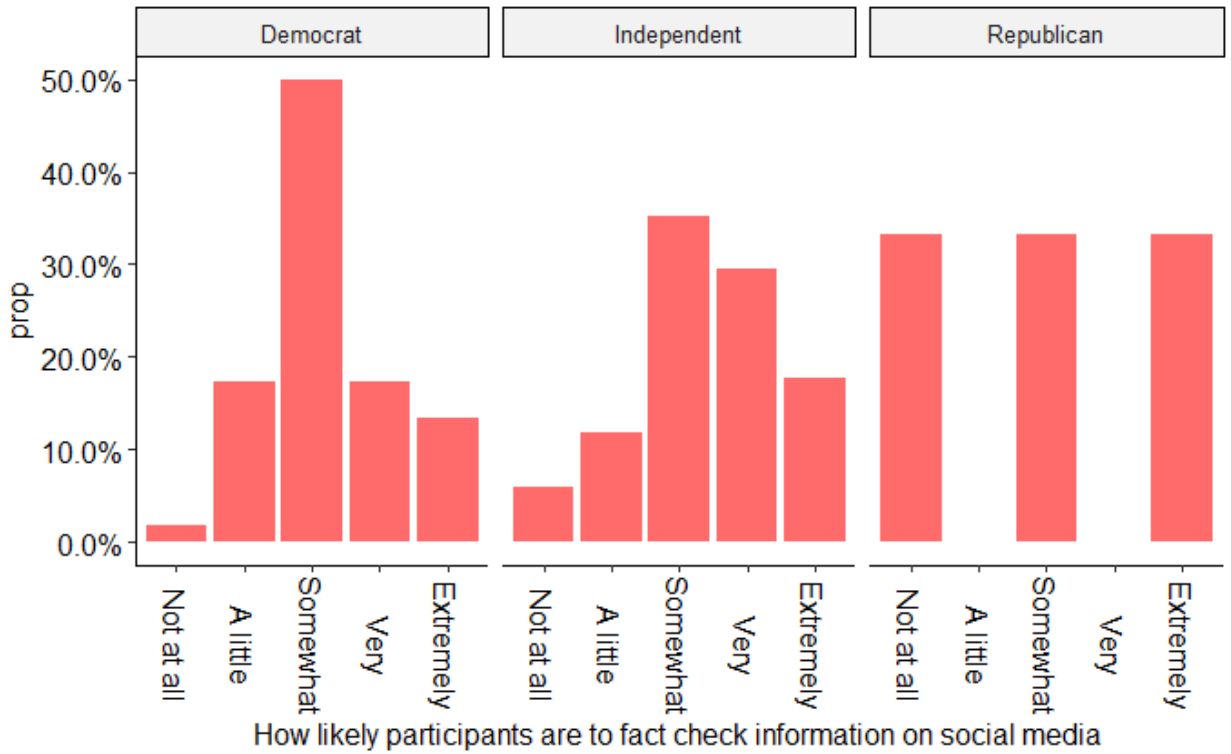


Figure 2: The proportional likelihood of fact-checking social media information per party. Each likelihood adds up to 100 percent for each group, so that likelihoods are relevant to the group itself.

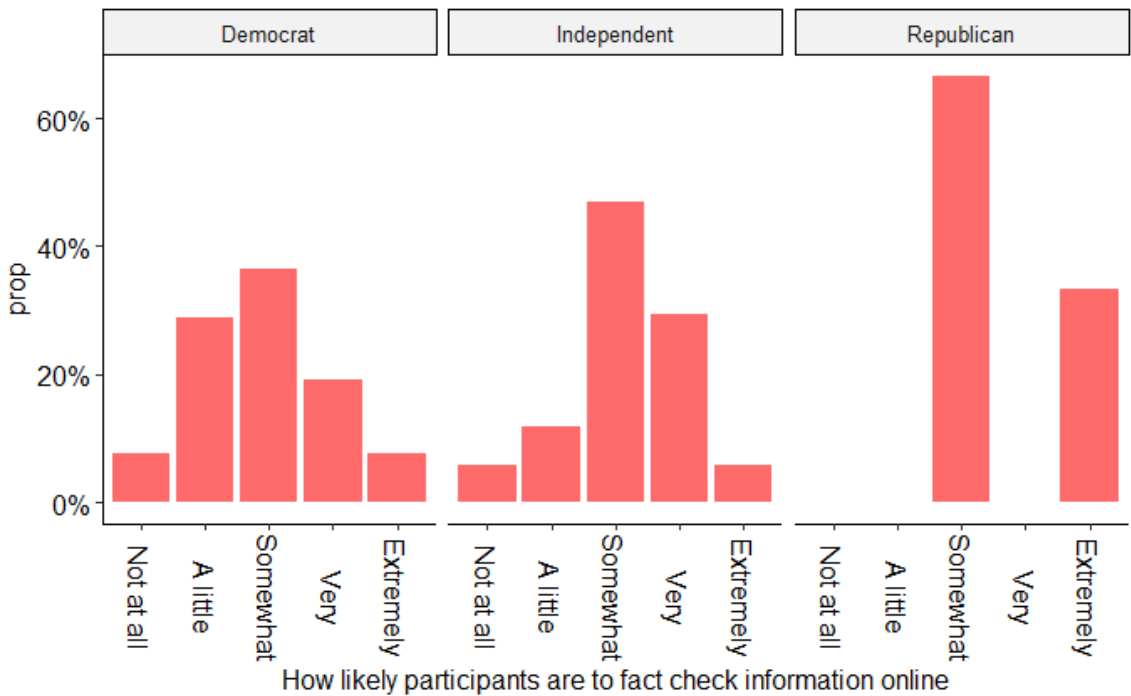


Figure 3: The proportional likelihood of fact-checking online information per party. Each likelihood adds up to 100 percent for each group, so that likelihoods are relevant to the group itself.

Figure 2 demonstrates that the Independent participants had a higher proportion of those who were “extremely” likely to fact-check information on social media. Conversely, Figure 3 shows that the Democratic participants were more “extremely” likely to fact-check general online information. Again, however, Independent participants had a higher portion of fact-checking social media and online information overall. It was unclear what caused the shift in proportions between the figures, so I pressed on to the last analysis.

The final analysis sought to answer two questions: how often each party consumes political information and how much time was spent taking the survey. The two explorations were combined, assuming that more engaged participants would be more likely to provide higher quality responses about how often they consume politically charged information.

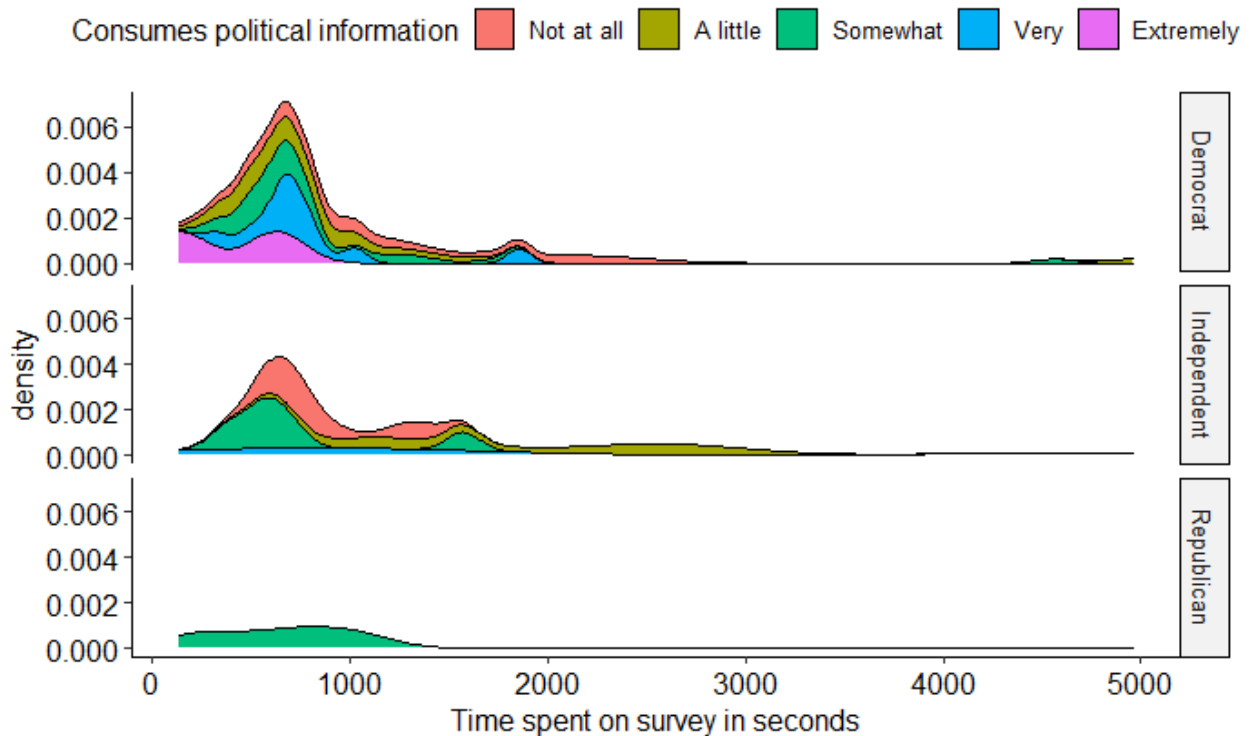


Figure 4: A density plot of the time spent taking the survey per political party group. The distributions are also color coded for how likely they are to consume political information online. Participant times below 700 seconds and above 5,000 seconds were omitted as outliers, or people who did not complete the survey properly.

Figure 4 displays several interesting aspects of the data pertinent to the exploratory goal of this analysis. At first glance it seems that Independent participants who consume “a little” political information spent the most time taking the survey. This may indicate they were the most invested in the survey, and therefore more likely to produce higher quality responses.

However, the average data across participants demonstrates that the survey was completed in under 2,000 seconds or about 33 minutes. Consequently, a more accurate observation may be that Democratic participants who do not consume political information spent the most time on the survey. This may indicate that their responses were more wholesome and possibly more reliable.

Furthermore, the graph illustrates that Democratic participants were more likely to consume political content than Independent participants. Being that it is possible Democratic participants produced more reliable responses based on the notion of time spent on the survey, it is curious to ponder whether or not Independent participants actually provided wholesome responses, and what it could mean if they did not.

Findings

Limitations

Before discussing final observations, it is important to mention the limitations of the present work. First and foremost, the dataset may have been too small and skewed to provide informative answers to the proposed research question. Furthermore, the participant data was made even smaller after outliers, and arbitrary information was removed, which may have contributed to the first issue. Lastly, a power analysis displayed that around 175 participants would be best to test the sort of questions that were being explored in this project, yet there were only 72 participant rows in this analysis.

Results

Concerning the research questions of this project, there are general observations to be noted. Firstly, the bar graphs display that Independent participants were the most likely overall to report that they fact-checked information on social media and online. This information may indicate how well one deals with information or how likely this participant group is to detect bias.

Furthermore, the density graph seems to display that the Democratic group had more individuals respond as “extremely” likely to consume political information online than Independent participants. The density graph may also suggest that the Democratic participants were the most engaged in the survey, which may indicate the quality of their responses. This information is interesting since Democrats were less likely to fact-check information but more likely to consume political content. However, if it were the case that Democrats had more engaged responses, it is interesting to consider how that impacts the validity of the likelihood to fact-check among Independent participants.

Above all, there are no conclusive findings to report. Given the exploratory nature of this project, all observations of the data were mere possibilities of how it could be interpreted. Nevertheless, many future explorations can be made from the analyses of this project.

Discussion

At the very least, the observations of this project may help researchers of the original study explore concepts not considered in the original hypothesis. Political party affiliations compared with how well they deal with information could be an interesting study. In the future, researchers of the original study may need to conclude the initial experiment before data analyses can be helpful.

It may also be beneficial to recruit more participants from each political party, so that comparisons can be more informative. As it stands currently, there are no significant findings to gauge participants' ability to detect bias. Thus, further research is needed to address this issue.

Bibliography

- Entman, R. M. (2008). *Framing bias: Media in the distribution of power*. *Journal of Communication*, 57(1), 163-173. <https://doi.org/10.1111/j1460-2466.2006.00336.x>
- Perez, D. (2021, December 14). *Final Project Analysis Report*. Retrieved from https://github.com/Dennisp7/finalProjectReport/blob/main/docs/Final_Project_Report_.pdf
- Westmark, A., Jones, A., Gamino, S., & Perez, D. (2021). *The Influence of Media Sources: How the Media Affects Judgement*. Unpublished manuscript, University of California, Merced