Do people watch the news for the news? Or for their confirmation bias?

Chien-Che Hung

12/22/2020

Abstract

In this report, we would be discussing the Taiwanese's political behavior with the influence of the Taiwanese television news outlets. In Taiwan, most new outlets would report news in favor of the major political parties, the Democratic Progressive Party (DPP, Green Party) and the Chinese Nationalist Party (KMT, Blue Party). Specifically, we want to determine whether people watch the news for political information or their confirmation bias. We utilize the data from Taiwan Communication Survey (TCS) to conduct this analysis. It is a survey aiming to capture the people's political behaviors on different forms of media. We decide to focus on news from televisions because the respondents believe that television news is more trustworthy than the internet's information. Due to the reason that the response variable that we are using has five categories, "Tsai Ing-wen from the DPP" (GreenParty), "Eric Chu from the KMT" (Blue Party), "James Soong from the PFP" (Neutral Party), "Don'tKnow," and "You would not cast a vote," we would be using Multinomial Logistic Regression to find out the relationship between the odds of voting for the candidates and the media behavior. In our analysis, regardless of the party, the respondents initially supported, watching the news from the opposing party would change their support parties. We also find out that people who support the Blue Party (KMT) are more likely to alternate their choice based on the new channels they watch. We also find out that people who acknowledge themselves as neutral in the political environment are more likely to get influenced by the type of news outlets they usually watch. Thus, we can conclude that even though people support their parties outside the elections, watching the news from different television sources could increase their chances of changing their support to other parties.

Keywords: Taiwanese Media, Election, Multinomial Logistic Regression

Introduction

In Taiwan, television news channels are relatively competitive compared to other channels, such as entertainment channels. Including paid TV channels and free-to-air TV channels, there are approximately 19 news channels. In the Taiwanese News market, it is believed that while a small number of channels report news in an objective and neutral sense, the majority of the news channel report news in favor of either one of the significant Taiwanese political parties, Democratic Progressive Party (DPP) and Chinese Nationalist Party (KMT). The ideology for DPP is Social Liberalism, whereas KMT's ideologies are Conservatism and Chinese Nationalism. For example, the news Sanlih E-Television News (SET) and Formosa Television (FTV) reports typically are slanted towards DPP, and the information from Television Broadcast Satellite (TVBS) is usually slanted towards KMT.

With the apparent stances that these news channels have, this report explores people's behavior of watching the news. For people who support either DPP or KMT, we want to find out the odds of changing the parties they support in relation to the news outlets they watch during the 2016 presidential election by analyzing the Taiwan Communication Survey data. The response variable for this analysis would be the answer to the question of "Suppose that the 2016 presidential election was being held today, which of the following candidates would you vote for?" Due to the reason that there are five categories in the answer, we are going to use Multinomial Logistic Regression. It is a method that is similar to Binomial Logistic Regression but for categories in the response variable that have more than two options.

Before we apply this model, we will separate the original data into three sub-datasets based on people's parties. After separating them, we would apply the models. The variables that we are going to use are some demographic variables such as Age, Gender, Education Level, and Marital Status. The variable related to news outlets would be the frequency of news per day in a week and the political stances on the information they watch. The variables that will be used would be elaborated clearly in the Data and Model section below. By applying this model, we can determine how the news outlets affect people's choice in the presidential election. Several interesting conclusions can be drawn from our analysis. For example, what is the increase in odds of voting for the opposing candidate that you support? Or what party's supporters have a larger chance of changing their thoughts after including the effect of news outlet influences.

This report would talk about the analysis being done by introducing the data first. The Data section includes the source of the data, how the data is collected, and what data manipulation that we did for the model that we are fitting. After the data section, we will explain and present the model that we are using in this report. Following the Model section, we would demonstrate the model result and discuss what it means in the discussion section. Finally, we would also talk about what improvements can be made to improve our analysis.

Data

Data Introduction

This data is obtained from the result of the fourth wave Taiwan Communication Survey (TCS) in 2015: Political and Civic Communication. TCS's aim is to track developments and changes in media behavior in Taiwan. Specifically, this wave of survey focuses on politics and public affair information dissemination, people's knowledge on politics, and election options and acknowledgment. There are three common information dissemination methods: interpersonal communication, mass media communication, and new media communication. Here, new media communication is defined as mass communication using digital technologies such as the internet.

As we stated in the introduction, we are trying to find whether people obtain information in ways that fit the ideology of the political party that they support. The main reason why we are focusing on television news outlets is because Taiwan news outlets have obvious political parties they support, either DPP or KMT. In addition, from the observation of TCS survey, around 50% of the interviewers choose to get the latest information from the Internet, whereas 45% of the interviewers would choose to watch television for the latest news. However, under the question of which source of information they trust the most, around 60% of the interviewers choose the television, whereas only 27% of the interviewers choose the Internet. Because of this reason, news outlet television would provide the best explanation on the interviewers' political behaviors to the questions this paper is exploring.

Sampling Method

The methodology that TCS used was Stratified Three-Stage Probability Proportional to Size Sampling through Computer Assisted Personal Interviewing (CAPI) on Commsurvey application. For this survey, the population was people who were eighteen-year-old and above, and the sample frame was people who lived at their registered residences for at least four days per week on the main island of Taiwan. The final sample for this survey is the people who got sampled through TCS's sampling method. The first stage was conducting random sampling on District and Township. Within the District and Township samples, a random sampling method on Urban Villages and Rural Villages would be conducted. In the third stage, address numbers within the Urban Villages and Rural Villages would be randomly sampled. For those households being sampled,

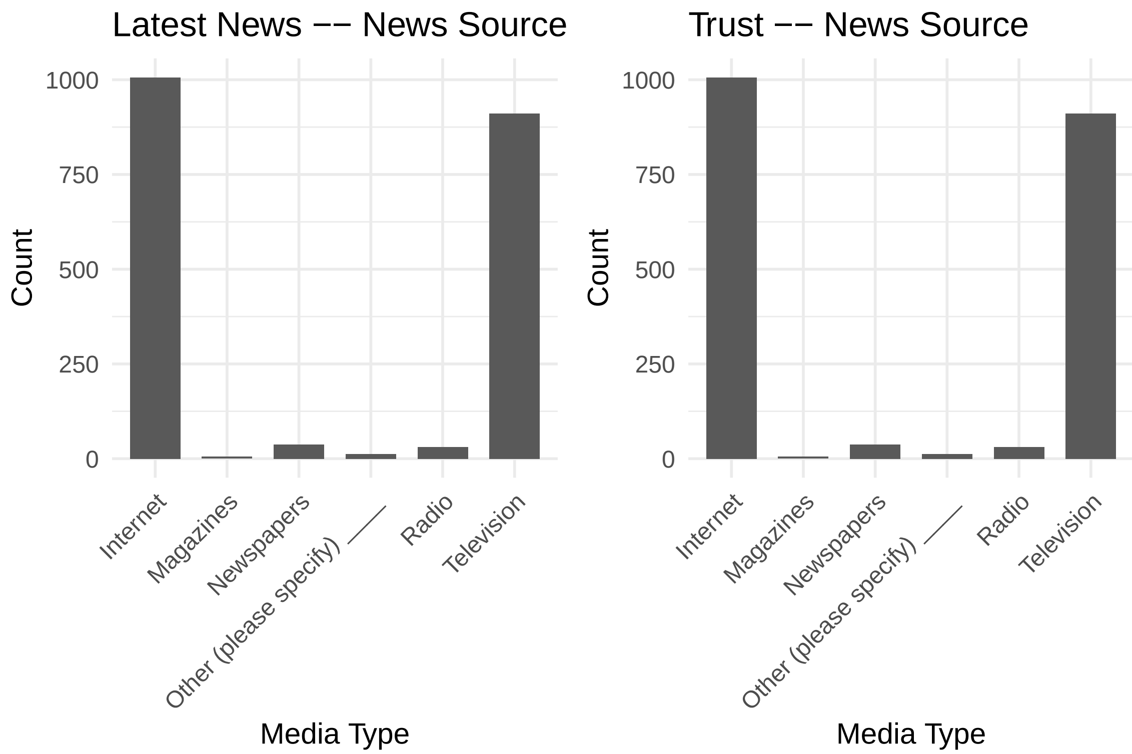

Figure 1: News source people watch vs News source people trust

interviewees would interview the household member within the pre-determined age range. The age range that was assigned was based on the age distribution of the entire population.

The nonresponse of the survey would occur in the survey. TCS member solved the nonresponse question by contacting the interviewees directly for their answers if the contact information was available. If the contact information was unavailable, the nonresponse questions would be filled in by the age-range's average. When there are "other" options in the questions, the reviewer after the survey was conducted would review the answers and categorize them into a pre-existent answer. If the pre-existent answers could not fit the the answers in the "other" options and the similar answers appeared several times in from the interviewees, the reviewer would add an option and categorize their answers in that option.

The noticeable feature from this survey was the well-structured sampling methods that could avoid the imbalanced situation among cities and townships. In addition, the flow that the questionnaire possessed made the questions easy to answer and the purpose of the questions were obvious. The intuitiveness not only benefit the goals that they tried to achieve, but also assisted people who would like to analyze their results. However, the way they conducted error correction could be very subjective. As we mentioned on how TCS managed to categorize "other" option into pre-existent options or newly-created option, this could really depend on how the reviewers thought at the moment during the review process. The answers would not be raw enough for the data analysis process.

Variables and Data Pre-Processing

Because we are trying to understand the relationship between interviewees' television news channel choices and political choices, this analysis would include several variables: people's political stances, the news channel they usually watch, the frequency of watching news, and who they would vote for on the 2016 Presidential Election. Most importantly, we would include interviewees' demographic variables because demographics could also contribute to who they will vote for during the 2016 Presidential Election. For example, according to an article from Taipei Times, KMT received no support from the age group of 20-24. This shows that the importance to include age groups in our analysis.

For converting the data into usable for our models, some variables required some adjustments. Before manipulating other variables, we needed to discard some observations because they did not receive news from television, which was the observations that answered zero on the question asking for the days they watch the news on television per day.

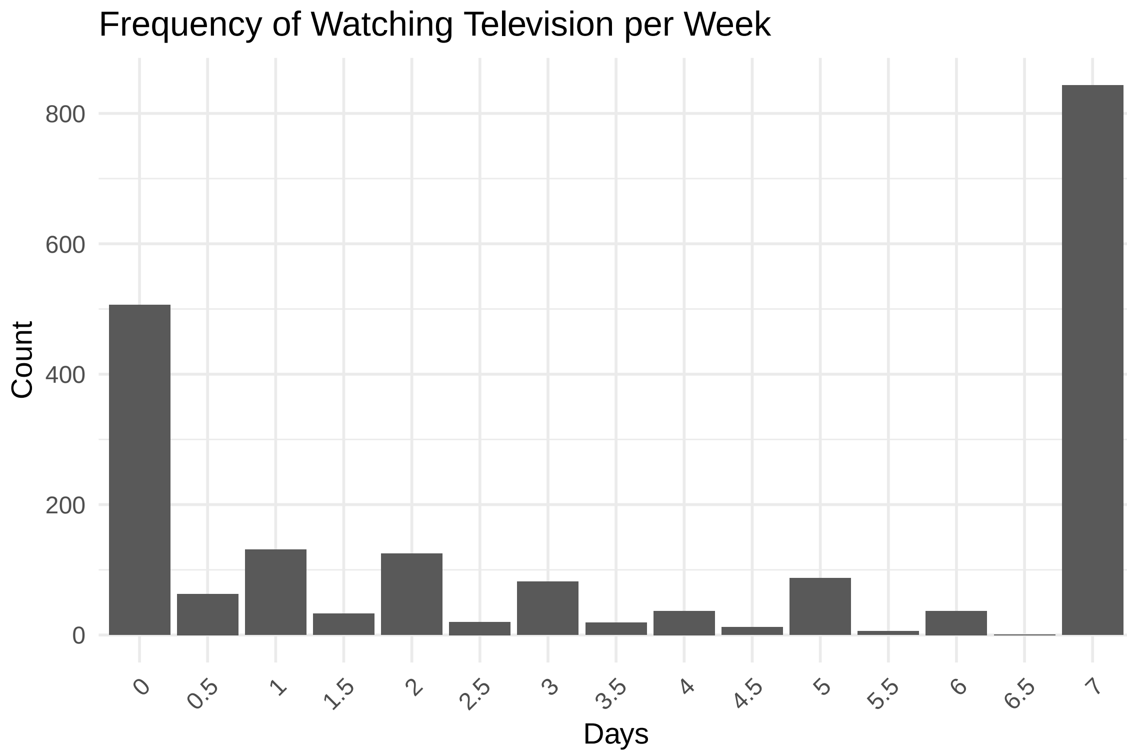

Figure 2: Frequency Distribution

For the answers to the question "During the 2016 presidential election, which television channel(s) do you watch most frequently for getting election news?" it allowed more than one choice. To determine whether the interviewees' options on news outlets were KMT (Blue Party), DPP (Green Party), or Neutral, we assigned each channel with either -1, 1, or 0 according to the news channel's political stances. After setting them, we would add all the interviewees' options and determine the political views on the media they watched. For the sums that were larger than 0, we assigned them as "Blue Party" viewers. For the sums that were lower than 0, we set them as "Green Party" viewers. For the sums that equaled 0, we put them into the "Neutral" category. Also, there was a variable that asked interviewees about their understanding of the channels' political affiliation. However, in this analysis, we would not be using it because it blurs the questions that we tried to answer in this paper.

Moreover, the survey asked about who the interviewees would support in the 2016 Presidential Election. If the interviewees did not know which candidate to support, the following question asked about which candidate they were leaning towards. We combined these two questions into one to better indicate whom the interviewees were going to support on the election. As we mentioned, this survey also asked about which political party the interviewees were leaning towards. Since we were only discussing KMT (Blue Party) and DPP (Green Party), the rest of the party would be put into the "Neutral" category to fit the political stances the news channels have (Blue, Green, and Neutral).

After categorizing them into three categories based on their political inclination: "Blue", "Green", and "Neutral," we separated them into three datasets according to their political inclination. The reason for us to

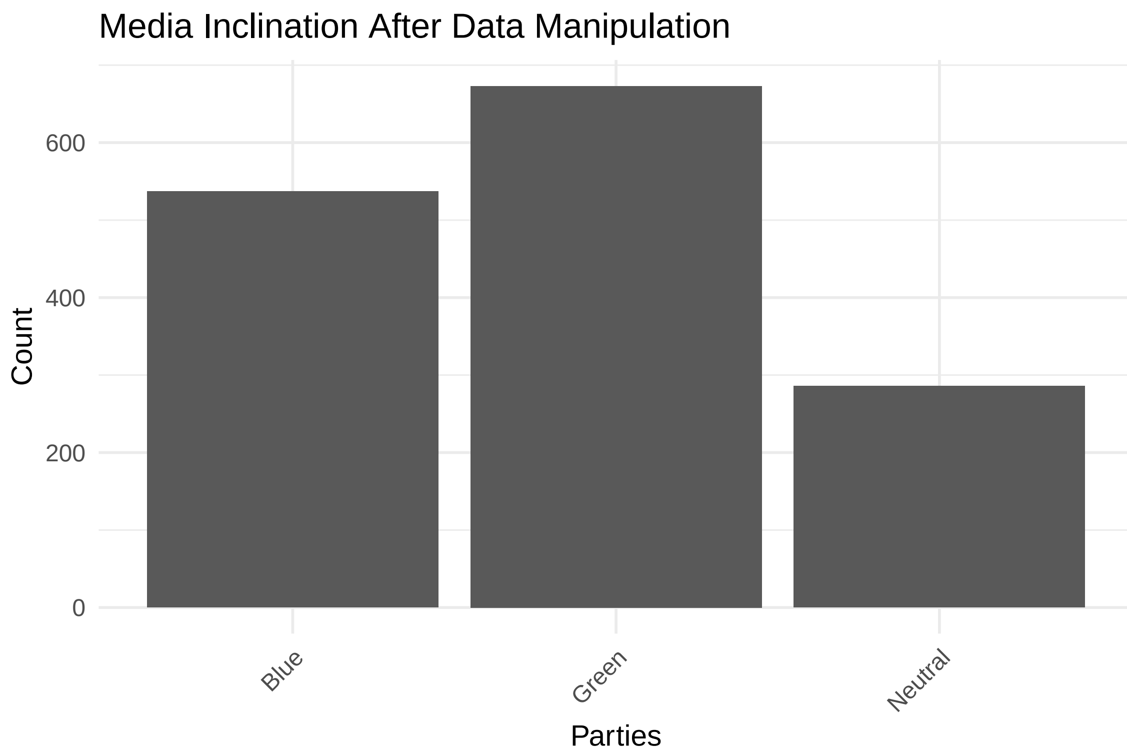

Figure 3: Media Inclination

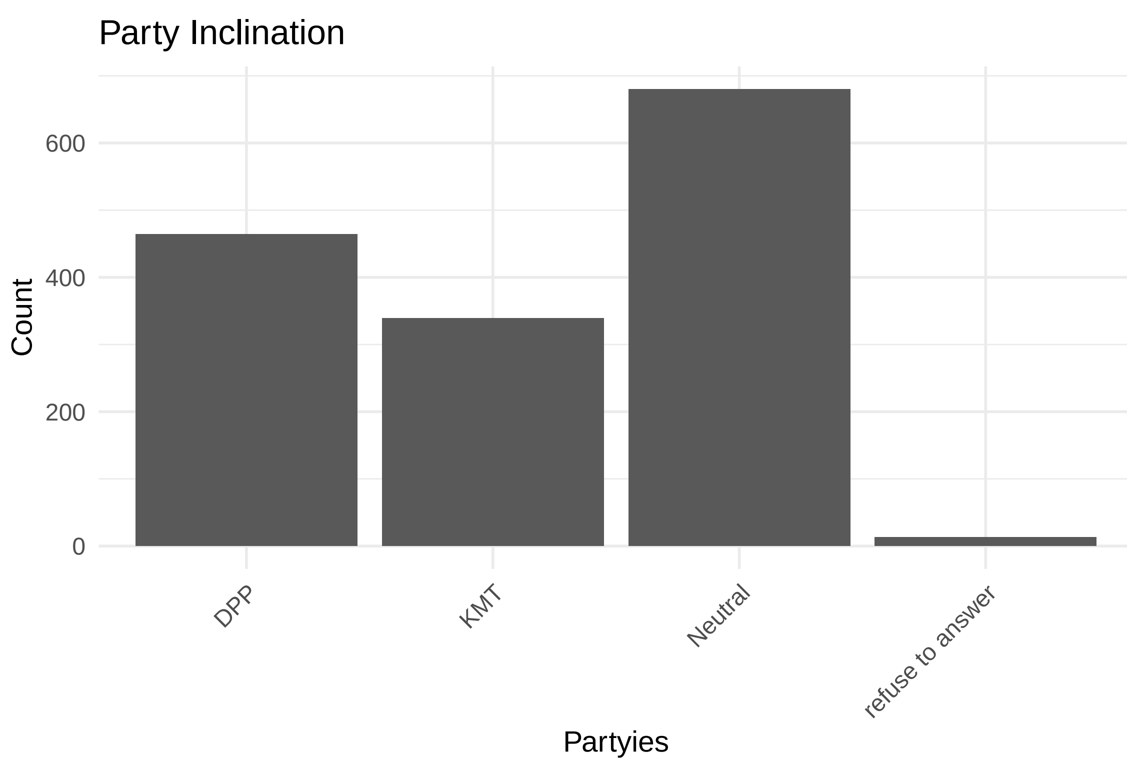

Figure 4: Parties that interviewees support

Table 1: First six interviewees for Green Sub-Dataset

Candidates	Gender	Age	Marital_Status	Education	Frequency	Media_Lean_Obj
Don't know	female	36	Married	university	7	Blue
James Soong from the PFP	male	51	Married	senior high school(general class)	7	Blue
Tsai Ing-wen from the DPP	male	63	Married	junior high school	7	Green
Tsai Ing-wen from the DPP	male	71	Married	senior high school(general class)	7	Neutral
Tsai Ing-wen from the DPP	male	56	Divorced or separated	junior high school	7	Green
Don't know	male	20	Single and never married	university	1	Blue

Table 2: First six interviewees for Green Sub-Dataset

Candidates	Gender	Age	Marital_Status	Education	Frequency	Media_Lean_Obj
Eric Chu from the KMT	female	47	Married	university	7	Green
Eric Chu from the KMT	male	57	Married	two-year junior college	7	Green
Eric Chu from the KMT	male	80	Married	junior high school	7	Blue
Eric Chu from the KMT	male	38	Married	Military/police college	7	Neutral
Eric Chu from the KMT	male	46	Married	master's degree	7	Neutral
Tsai Ing-wen from the DPP	male	57	Married	elementary school	7	Green

separate them into three datasets is mainly because we can find whom political inclined interviewees and political neutral interviewees would vote for in 2016 election under the influence of different political views' news outlets. The graph below shows the basic structure of our analysis and data and the table below shows the first few observation for each type of political inclination datasets.

Model

Model Introduction

This analysis tries to find the effects on interviewees' television news channel choices on the candidate they intended to vote for during the 2016 election. As we mentioned in the Data section, we separated the data with 2002 variables into three datasets according to interviewees' political inclination, which is "Blue Party," "Green Party," and "Neutral." The same model would be applied to these three datasets. Due to the nature of the Candidate variable, which has five categories: "Tsai Ing-wen from the DPP" (Green Party), "Eric Chu from the KMT" (Blue Party), "James Soong from the PFP" (Neutral Party), "Don't Know," and "You would not cast a vote," we are going to use Multinomial Logistic Regression. Similar to Binomial Logistic Regression, the response variable's log-odds are the independent variable's linear combination. The below explanations would briefly introduce logistic regression and multinomial regression.

$$log(\frac{P(A)}{1-P(A)}) = x_0 + \beta_1 x_1 + ... + \beta_n x_n$$

In the model above, the equation's left-hand side is the log-odds of the events, which is taking the logarithm of the probability of Event A happening over the probability of Event B $(1 - P(A))$ happening. On the right

Table 3: First six interviewees for Green Sub-Dataset

Candidates	Gender	Age	Marital_Status	Education	Frequency	Media_Lean_Obj
Eric Chu from the KMT	female	47	Married	university	7	Green
Eric Chu from the KMT	male	57	Married	two-year junior college	7	Green
Eric Chu from the KMT	male	80	Married	junior high school	7	Blue
Eric Chu from the KMT	male	38	Married	Military/police college	7	Neutral
Eric Chu from the KMT	male	46	Married	master's degree	7	Neutral
Tsai Ing-wen from the DPP	male	57	Married	elementary school	7	Green

hand side, $x_0 \dots x_n$ are the observations for each variable and $\beta_1 \dots \beta_n$ are the coefficients of the logistic regression.

The description above is the explanation of logistic regression. Multinomial regression works similarly, with more than two outcomes. To explain the regression, let us assume three categories in the response variables: Event A, Event B, and Event C. The equations are in the following model.

$$log(\frac{P(A)}{P(B)}) = x_{01} + \beta_{11}x_{11} + \dots + \beta_{n1}x_{n1}$$
$$log(\frac{P(C)}{P(B)}) = x_{02} + \beta_{12}x_{12} + \dots + \beta_{n2}x_{n2}$$

The difference is that rather than having $1 - P(A)$ on the denominator in the odds function, we would have to decide a base category that other events would regress against. In this case, $P(B)$ is the base category, and the results of Event A and Event C would be fitted relative to $P(B)$. Thus, if we have N categories in our response variable, we would have N-1 separate logistic regressions with distinct coefficients. In the example we are using, there would be two regressions being trained. On the right hand side of the regressions, $x_{01} \dots x_{n1}$ and $x_{02} \dots x_{n2}$ are the observations according to the regressions being ran. In addition, $\beta_{11} \dots \beta_{n1}$ and $\beta_{12} \dots \beta_{n2}$ are the coefficients of the regressions accordingly.

Model

Based on the introduction of the model that we are using in the above subsection, we will talk about the models that we are using in our analysis. Since we separate the original data into three datasets based on the interviewees' parties, we will build models upon them separately. In general, the response variable would be the Candidate variable. There are five options like we mentioned above: "Tsai Ing-wen from the DPP" (Green Party), "Eric Chu from the KMT" (Blue Party), "James Soong from the PFP" (Neutral Party), "Don't Know," and "You would not cast a vote." The categories satisfy the assumption of performing multinomial logistic regression because it is a categorical variable and has more than two types. The base category we will use in each dataset would be the dataset political party's candidate. For example, in Green Party's dataset, the base category would be Green Party's candidate Tsai Ing-wen. By doing this, we can see the performance of the odds based on the base category. From the number of the Candidates variable types, we can know that we will fit 5-1 = 4 logistic regression with different events on the numerator in the log-odds for each data.

DPP (Green Party) Dataset

For the Green Party dataset, the base category would be people who would vote for "Tsai Ing-wen from the DPP" in the 2016 Presidential Election.

$$log(\frac{P(KMT)}{P(DPP)}) = \beta_{01} + \beta_{11}Gender + \beta_{21}Age + \beta_{31}Marital + \beta_{31}Education +$$
$$\beta_{41}MediaBlue + \beta_{51}MediaNeutral$$
$$log(\frac{P(Neutral)}{P(DPP)}) = \beta_{02} + \beta_{12}Gender + \beta_{22}Age + \beta_{32}Marital + \beta_{32}Education +$$
$$\beta_{42}MediaBlue + \beta_{52}MediaNeutral$$
$$log(\frac{P(DontKnow)}{P(DPP)}) = \beta_{03} + \beta_{13}Gender + \beta_{23}Age + \beta_{33}Marital + \beta_{33}Education +$$
$$\beta_{43}MediaBlue + \beta_{53}MediaNeutral$$
$$log(\frac{P(NotVoting)}{P(DPP)}) = \beta_{04} + \beta_{14}Gender + \beta_{24}Age + \beta_{34}Marital + \beta_{34}Education +$$
$$\beta_{44}MediaBlue + \beta_{54}MediaNeutral$$

As we mentioned, the base category would be DPP, which is the Green Party's official abbreviation. The variables that we will use in the models are the following: Gender of the interviewees, Age of the interviewees, Marital Status of the interviewees, Education Level of the interviewees, and the Media Political Inclination status of the interviewees that we created during the data cleaning process. In our data, there is actually a variable about what the interviewee thought about the political stances that the news outlets have. However, to make the result that we have as objective as possible, we choose to use the variable that we created based purely on the channels they normally watched. Thus, we decided to exclude such variable.

For the coefficients, β_{0m}, where $m = 1, 2, 3, 4$, are the intercepts for each logistic regression that we are fitting. For β_{1m}, where $m = 1, 2, 3, 4$, are the coefficients for Gender variable. β_{2m} , where m = 1, 2, 3, 4, are the coefficients of the Age variable. β_{3m}, where m = 1, 2, 3, 4, are the Marital Status Variable coefficients. β_{3m} , where m = 1, 2, 3, 4, are the coefficients for the Education Level variable. Since we want to understand the effects of the Blue side of the news outlet and Neutral news outlet on people who support the Green Party's final choice, we make the Green news outle a reference category. Thus, the coefficients of β_{4m} and β_{5m} will be presented as odds against Green news outlet coefficients automatically equal to 1. We can assess the effects of different news outlets' effects on people voting for the candidates.

KMT (Blue Party) Dataset

For the Blue Party dataset, the base category would be people who would vote for "Eric Chu from the KMT" in the 2016 Presidential Election.

$$log(\frac{P(DPP)}{P(KMT)}) = \beta_{01} + \beta_{11}Gender + \beta_{21}Age + \beta_{31}Marital + \beta_{31}Education +$$
$$\beta_{41}MediaGreen + \beta_{51}MediaNeutral$$
$$log(\frac{P(DontKnow)}{P(KMT)}) = \beta_{03} + \beta_{13}Gender + \beta_{23}Age + \beta_{33}Marital + \beta_{33}Education +$$
$$\beta_{43}MediaGreen + \beta_{53}MediaNeutral$$
$$log(\frac{P(NotVoting)}{P(KMT)}) = \beta_{04} + \beta_{14}Gender + \beta_{24}Age + \beta_{34}Marital + \beta_{34}Education +$$
$$\beta_{44}MediaGreen + \beta_{54}MediaNeutral$$
$$log(\frac{P(Neutral)}{P(KMT)}) = \beta_{02} + \beta_{12}Gender + \beta_{22}Age + \beta_{32}Marital + \beta_{32}Education +$$
$$\beta_{42}MediaGreen + \beta_{52}MediaNeutral$$

The structure of each logistic regression function is similar to Green Party Multinomial Logistic Regression, including the coefficients. The only difference would be the MediaGreen variable. The reason for changing **MediaBlue** to **MediaGreen** is because by doing this, we can find out the effects that Green News Outlets have on the people who support the Blue Party. Thus, the coefficients, β_{4m} where m = 1, 2,3, 4, are the coefficients for the variable MediaGreen.

Due to the reason that we separate the data into three parts, Green, Blue, and Neutral, we also are going to find the effects of the news outlets have on people who think they are neutral on the political stances.

$$
\begin{aligned}
log(\frac{P(DPP)}{P(Neutral)}) =& \beta_{01} + \beta_{11}Gender + \beta_{21}Age + \beta_{31}Marital + \beta_{31}Education + \\
& \beta_{41}MediaGreen + \beta_{51}MediaBlue \\
log(\frac{P(KMT)}{P(Neutral)}) =& \beta_{02} + \beta_{12}Gender + \beta_{22}Age + \beta_{32}Marital + \beta_{32}Education + \\
& \beta_{42}MediaGreen + \beta_{52}MediaBlue \\
log(\frac{P(DontKnow)}{P(Neutral)}) =& \beta_{03} + \beta_{13}Gender + \beta_{23}Age + \beta_{33}Marital + \beta_{33}Education + \\
& \beta_{43}MediaGreen + \beta_{53}MediaBlue \\
log(\frac{P(NotVoting)}{P(Neutral)}) =& \beta_{04} + \beta_{14}Gender + \beta_{24}Age + \beta_{34}Marital + \beta_{34}Education + \\
& \beta_{44}MediaGreen + \beta_{54}MediaBlue
\end{aligned}
$$

Similarly, the regression models' structure is similar to the ones that we mentioned above, with the base model being people who have neutral political stances. The changes made in this model on the Media Political Inclination variable is also similar to Blue Party Dataset. We change the MediaNeutral to Media Blue because we can easily increase or decrease odds relative to MediaNeutral. We can see the effects news outlets, either Green news outlets or Blue news outlets, have on Neutral Voters.

The R programming language would run the above models. The primary function would be multinom from the nnet package. Due to the reason that multinomial regressions are fitted via neural networks, we can see the convergence situation during the process of fitting the models. The above three models converged after 60, 30, and 30 iterations. For the assumptions to fit this model, we need to make sure the number of the response variable categories is larger than 2. In this case, the number of categories in the response variable is 5, which passes the assumption test. The second assumption would be the observation has to be independent of each other. Due to the sampling methods' nature, the sampling methods from the survey collection stage ensure the independence between the observations. For multinomial logistic model, we normally have to test an assumption called Independence of Irrelevant Alternative (IIA). This is an assumption "that when people are asked to choose among a set of alternatives, their odds of choosing A over B should not depend on whether some other alternative C is present or absent." (CITEEEEE) However, we would not test it for the following reasons. Due to the nature of our Candidate variable, it would be impossible for people to not recognize the presence of other candidates during the election. Also, the test created by Hausman and McFadden does not perform well enough in both small samples and large samples. Next, we have to check whether the inclusion of the news outlets' inclinations affects the events' log-odds to see if modifications on our model are needed. After conducting the Linear Hypothesis Test with linearHypothesis function in the car package, some coefficients of MediaBlue, MediaGreen, and MediaNeutral reject the null hypothesis of these coefficients equal to zero. Thus, we can say that the inclusion of Media Political Inclination is valid, which means that we do not need alternative models for our analysis. Not checking other demographic variables is that they could cause minor adverse effects if the variables are not included. The following section will present about the results after fitting the model.

Result

In this section, we will present the result that we obtain from our model. There would be three results due to the sub-datasets that we separated. The hat that in the equations indicates that they are the estimated values from the regressions.

The estimated coefficients and the intercepts are originally in log-odds form. For better interpretation, we take the exponential of each cells of the tables presented above. As we take exponential for the elements, the left hand side of the model would become $\frac{\hat{P}(Category)}{\hat{P}(BaseCategory)}$, which is the odds of "Category" happening. When the exponentiated coefficients are larger than 1, it means that the "Category" would have higher probability of happening compared to the "Base Category." On the other hand, if the exponentiated coefficients are smaller than 1, it means that the "Base Category" would have higher chance of happening.

DPP Result

Firstly, this is the result from Green Party dataset and the below four equations are the estimated model from our model.

$$
\begin{aligned}
log(\frac{\hat{P}(KMT)}{\hat{P}(DPP)}) =& \hat{\beta}_{01} + \hat{\beta}_{11}Gender + \hat{\beta}_{21}Age + \hat{\beta}_{31}Marital + \hat{\beta}_{31}Education + \\
& \hat{\beta}_{41}MediaBlue + \hat{\beta}_{51}MediaNeutral \\
log(\frac{\hat{P}(Neutral)}{\hat{P}(DPP)}) =& \hat{\beta}_{02} + \hat{\beta}_{12}Gender + \hat{\beta}_{22}Age + \hat{\beta}_{32}Marital + \hat{\beta}_{32}Education + \\
& \hat{\beta}_{42}MediaBlue + \hat{\beta}_{52}MediaNeutral \\
log(\frac{\hat{P}(DontKnow)}{\hat{P}(DPP)}) =& \hat{\beta}_{03} + \hat{\beta}_{13}Gender + \hat{\beta}_{23}Age + \hat{\beta}_{33}Marital + \hat{\beta}_{33}Education + \\
& \hat{\beta}_{43}MediaBlue + \hat{\beta}_{53}MediaNeutral \\
log(\frac{\hat{P}(NotVoting)}{\hat{P}(DPP)}) =& \hat{\beta}_{04} + \hat{\beta}_{14}Gender + \hat{\beta}_{24}Age + \hat{\beta}_{34}Marital + \hat{\beta}_{34}Education + \\
& \hat{\beta}_{44}MediaBlue + \hat{\beta}_{54}MediaNeutral
\end{aligned}
$$

Table 4:

	Dependent variable:			
	Don't know	Eric Chu from the KMT	James Soong from the PFP	You would not cast a vote.
	(1)	(2)	(3)	(4)
Gendermale	−0.595	0.705	1.213	−0.199
	(0.525)	(1.260)	(0.817)	(0.617)
Age	−0.022	−0.052	−0.037	−0.010
	(0.018)	(0.043)	(0.026)	(0.022)
as.numeric(as.factor(Marital_Status))	0.366	0.897	−0.339	0.184
	(0.353)	(0.905)	(0.563)	(0.438)
as.numeric(as.factor(Education))	−0.037	0.005	−0.040	−0.018
	(0.052)	(0.123)	(0.071)	(0.064)
Frequency	0.203	5.622***	0.100	0.019
	(0.178)	(0.611)	(0.220)	(0.177)
Media_Lean_ObjBlue	1.360**	1.747	−0.980	0.377
	(0.561)	(1.254)	(1.074)	(0.723)
Media_Lean_ObjNeutral	0.590	−10.680***	−13.285***	0.292
	(0.738)	(0.00001)	(0.00000)	(0.846)
Constant	−4.694**	−45.858***	−1.782	−3.775*
	(1.993)	(0.087)	(2.829)	(2.286)
Akaike Inf. Crit.	407.192	407.192	407.192	407.192
Note:				*p<0.1; **p<0.05; ***p<0.01

From the above table, the coefficients are shown in log odds form. We can see that Neutral News outlets variable's effects on voting for the Blue Party Candidate and Neutral Party Candidate are statistically significant. In this case, statistically significant means that the p-value is smaller than 0.1, 0.05, or 0.01. P-values in simple words means that the null hypothesis is true. Here the null hypothesis would be the coefficient equals to zero. The table below shows the exponentiated form of the coefficients. The results can be interpreted the way that we described above.

Table 5: Exponentiated Coefficients (Odds Form)

	Don't know	Eric Chu from the KMT	James Soong from the PFP	You would not cast a vote.
Intercept	0.00915	0.00000	0.16833	0.02295
GenderMale	0.55149	2.02390	3.36228	0.81936
Age	0.97835	0.94900	0.96335	0.98987
Marital Status	1.44219	2.45179	0.71258	1.20161
Education	0.96398	1.00451	0.96127	0.98171
Frequency	1.22544	276.48914	1.10485	1.01903
Media__ObjGreen	3.89708	5.73595	0.37515	1.45743
Media__ObjNeutral	1.80465	0.00002	0.00000	1.33966

KMT Result

The below models and the table are the results from the Blue Party dataset.

$$
\begin{aligned}
log(\frac{\hat{P}(DPP)}{\hat{P}(KMT)}) =& \hat{\beta}_{01} + \hat{\beta}_{11}Gender + \hat{\beta}_{21}Age + \hat{\beta}_{31}Marital + \hat{\beta}_{31}Education + \\
& \hat{\beta}_{41}MediaGreen + \hat{\beta}_{51}MediaNeutral \\
log(\frac{\hat{P}(DontKnow)}{\hat{P}(KMT)}) =& \hat{\beta}_{03} + \hat{\beta}_{13}Gender + \hat{\beta}_{23}Age + \hat{\beta}_{33}Marital + \\
& \hat{\beta}_{33}Education + \hat{\beta}_{43}MediaGreen + \hat{\beta}_{53}MediaNeutral \\
log(\frac{\hat{P}(NotVoting)}{\hat{P}(KMT)}) =& \hat{\beta}_{04} + \hat{\beta}_{14}Gender + \hat{\beta}_{24}Age + \hat{\beta}_{34}Marital + \\
& \hat{\beta}_{34}Education + \hat{\beta}_{44}MediaGreen + \hat{\beta}_{54}MediaNeutral \\
log(\frac{\hat{P}(Neutral)}{\hat{P}(KMT)}) =& \hat{\beta}_{02} + \hat{\beta}_{12}Gender + \hat{\beta}_{22}Age + \hat{\beta}_{32}Marital + \\
& \hat{\beta}_{32}Education + \hat{\beta}_{42}MediaGreen + \hat{\beta}_{52}MediaNeutral
\end{aligned}
$$

The estimated coefficients above are also shown in log-odds form. We can see that coefficients of the news outlets who support Green party have an effect on people who support Blue party's choice on voting for the presidential candidate. Also neutral news outlets' coefficient is also statistically significant in neutral candidate's model. The table below also shows the exponentiated form of the coefficients.

Table 6:

	Dependent variable:			
	Don't know	James Soong from the PFP	Tsai Ing-wen from the DPP	You would not cast a vote.
	(1)	(2)	(3)	(4)
Gendermale	-0.198	-0.339	0.399	0.114
	(0.374)	(0.427)	(0.462)	(0.460)
Age	-0.014	-0.041^{***}	-0.035^{*}	-0.040^{**}
	(0.013)	(0.016)	(0.019)	(0.017)
as.numeric(as.factor(Marital_Status))	0.558^{**}	0.693^{**}	0.226	0.519
	(0.236)	(0.283)	(0.366)	(0.320)
as.numeric(as.factor(Education))	-0.011	0.003	-0.015	-0.009
	(0.029)	(0.035)	(0.038)	(0.037)
Frequency	-0.034	0.073	-0.032	-0.133
	(0.124)	(0.138)	(0.142)	(0.131)
Media_Lean_ObjGreen	-0.130	1.146^{**}	1.569^{***}	-0.754
	(0.539)	(0.504)	(0.546)	(0.794)
Media_Lean_ObjNeutral	0.671^{*}	0.815^{*}	0.962	0.360
	(0.400)	(0.489)	(0.591)	(0.505)
Constant	-2.533^{**}	-3.058^{**}	-1.745	-1.061
	(1.266)	(1.512)	(1.772)	(1.555)
Akaike Inf. Crit.	767.913	767.913	767.913	767.913
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 7: Exponentiated Coefficients (Odds Form)

	Don't know	James Soong from the PFP	Tsai Ing-wen from the DPP	You would not cast a vote.
Intercept	0.07945	0.04697	0.17469	0.34599
GenderMale	0.82035	0.71235	1.49099	1.12047
Age	0.98569	0.95970	0.96559	0.96094
Marital Status	1.74732	1.99896	1.25420	1.67972
Education	0.98889	1.00261	0.98546	0.99079
Frequency	0.96628	1.07561	0.96894	0.87570
Media_ObjGreen	0.87831	3.14426	4.80184	0.47044
Media_ObjNeutral	1.95549	2.25979	2.61759	1.43352

Neutral Data

Other than assess the results from the Green Party and Blue Party dataset, we are also going to see the result from the Neutral dataset.

$$
\begin{aligned}
log(\frac{\hat{P}(DPP)}{\hat{P}(Neutral)}) =& \hat{\beta}_{01} + \hat{\beta}_{11} Gender + \hat{\beta}_{21} Age + \hat{\beta}_{31} Marital + \hat{\beta}_{31} Education + \\
& \hat{\beta}_{41} MediaGreen + \hat{\beta}_{51} MediaBlue \\
log(\frac{\hat{P}(KMT)}{\hat{P}(Neutral)}) =& \hat{\beta}_{02} + \hat{\beta}_{12} Gender + \hat{\beta}_{22} Age + \hat{\beta}_{32} Marital + \hat{\beta}_{32} Education + \\
& \hat{\beta}_{42} MediaGreen + \hat{\beta}_{52} MediaBlue \\
log(\frac{\hat{P}(DontKnow)}{\hat{P}(Neutral)}) =& \hat{\beta}_{03} + \hat{\beta}_{13} Gender + \hat{\beta}_{23} Age + \hat{\beta}_{33} Marital + \hat{\beta}_{33} Education + \\
& \hat{\beta}_{43} MediaGreen + \hat{\beta}_{53} MediaBlue \\
log(\frac{\hat{P}(NotVoting)}{\hat{P}(Neutral)}) =& \hat{\beta}_{04} + \hat{\beta}_{14} Gender + \hat{\beta}_{24} Age + \hat{\beta}_{34} Marital + \hat{\beta}_{34} Education + \\
& \hat{\beta}_{44} MediaGreen + \hat{\beta}_{54} MediaBlue
\end{aligned}
$$

Table 8:

	Dependent variable:			
	Don't know	Eric Chu from the KMT	Tsai Ing-wen from the DPP	You would not cast a vote.
	(1)	(2)	(3)	(4)
Gendermale	-0.652^{**}	-0.963^{**}	-0.571^{**}	-0.697^{**}
	(0.280)	(0.392)	(0.286)	(0.340)
Age	0.009	0.014	−0.005	−0.013
	(0.010)	(0.014)	(0.010)	(0.012)
as.numeric(as.factor(Marital_Status))	−0.108	0.089	−0.068	0.021
	(0.224)	(0.297)	(0.229)	(0.276)
as.numeric(as.factor(Education))	0.023	0.036	0.015	0.038
	(0.023)	(0.032)	(0.023)	(0.028)
Frequency	−0.029	0.075	−0.040	0.127
	(0.083)	(0.124)	(0.084)	(0.107)
Media_Lean_ObjBlue	0.144	0.755	−0.029	0.310
	(0.367)	(0.526)	(0.388)	(0.446)
Media_Lean_ObjGreen	0.012	0.041	0.581	−0.025
	(0.355)	(0.540)	(0.364)	(0.442)
Constant	1.496	−2.022	1.725	−0.284
	(1.118)	(1.553)	(1.145)	(1.410)
Akaike Inf. Crit.	1,919.270	1,919.270	1,919.270	1,919.270
Note:				*p<0.1; **p<0.05; ***p<0.01

The table above shows the log-odds results from the model and the table below shows the exponentiated form of the coefficients.

In the Discussion section below, we will discuss the result we just presented above.

Table 9: Exponentiated Coefficients (Odds Form)

	Don't know	Eric Chu from the KMT	Tsai Ing-wen from the DPP	You would not cast a vote.
Intercept	4.46240	0.13237	5.60975	0.75269
GenderMale	0.52113	0.38169	0.56522	0.49823
Age	1.00924	1.01364	0.99522	0.98677
Marital Status	0.89770	1.09358	0.93470	1.02164
Education	1.02324	1.03705	1.01554	1.03897
Frequency	0.97113	1.07823	0.96053	1.13588
Media__ObjGreen	1.15534	2.12825	0.97155	1.36324
Media__ObjNeutral	1.01254	1.04231	1.78766	0.97539

Discussion

This section will talk and discuss the result that we obtained that we presented in the above section. However, the outcome we are talking about here only speaks to the relationship between the media's stances and the interviewee's ideologies. It does not imply causation. We will discuss this section by explaining each sub-dataset model and what it means as a whole. After that, we propose what changes can be made to improve this analysis and apply the result for future use.

DPP dataset (Table 5)

Like what we mentioned in the Model section, the model category is "Tsai Ing-wen from the DPP." Thus, the results are shown in the form of log-odds. In order words, the numbers represent how other political stances' media would affect the interviewees to vote for those parties' candidates. Also, this sub-dataset contains all the interviewees who support the Green Party. In Table #####, we can see that people who watch news channels that favor Blue Party are five times more likely to vote for the candidate of the Blue Party — Eric Chu after switching from watching news outlets that support the Green Party. Here, you would not see the odds coefficients for the Media_Lean_ObjGreen because it is the baseline category. For Neutral Candidate, James Soong from the PFP, the probability for voting for him is significantly lower than voting for Tsai Ing-wen from the DPP, for both people who watch news that favor the Blue Party or the Neutral Party. This also means that the interviewees would still support the Green Party's candidate. However, we can see that switching to watching the news from opposite parties or neutral stances would increase the odds of not voting or do not who to vote for. In other words, by switching to the news outlets to Blue Parties or Neutral, it would be less likely for the interviewees to decide who to vote for in the 2016 presidential election.

Let's also examine what other demographics would affect the voting odds for the Blue Party Candidate Eric Chu. The noticeable would be the frequency. The considerable number indicates that the increase in the frequency of watching the news per week would vastly increase the odds for voting for the Blue Party's Candidate. From the Gender variable, we can see that the Male interviewees would be more likely to vote for the candidates from the Blue Party and the Neutral Party, and Female interviewees would be more likely to choose not to vote or do not know who to vote for.

KMT dataset (Table 7)

The interpretation would be the same as the DPP dataset. From Table #### above, people who support the Blue Party and watching Green Party sided news are four times more likely to vote for Green Party's Candidate — Tsai Ing-wen. Eric Chu from the KMT is not shown because it is the base category of this model, which means it is the odds function's denominator. Even for Blue Party's supporters who watch the neutral news would have a higher odds (also higher probability) of voting for Tsai Ing-wen. When we look at the odds of voting for James Soong from the PFP against Eric Chu from the KMT, for the blue party supporters, it is three times likely for them to change their minds and vote for Jame Soong if they watch the news outlets in favor of the Green Party. It is also two times more likely for people to support James Soong if they watch neutral news. However, watching the news that favors the Green Party would not increase the probability of making people choose not to know who to vote for and not cast a vote. On the other hand, if people who support Blue Party watch neutral news, it is 95% more likely to answer and do not know how to vote for and 43% more likely not to vote. In general, we can say that Blue party supporters are more likely to base on the news and understand the news's facts to choose the people they are going to vote for.

Similar to the DPP dataset, let us take a look at how demographics affect the voting odds. For the Age coefficients, we can see that the odds are almost equal to 1, which means that the one unit of increase in age only has a small effect on the odds of voting for the respective candidates, not going to answer do not know. Also, the increased frequency of watching the news per week would slightly affect the interviewees' choice.

Neutral Dataset (Table 9)

In this dataset, people selected that they are neutral in their political position. Thus, the model result would show how much and how likely the Taiwanese media would affect its citizens' choice of candidates. In Table XXX, for people who watch news outlets that favor the Blue Party, the odds of voting for Eric Chu from the KMT would double and only slightly affect other candidates. Also, for people who watch news outlets that favor the Green Party, it would increase the odds of voting for Tsai Ing-wen by 70%, whereas voting for other candidates would only affect voting. From this result, we can objectively say that there is a relationship between how media would affect people's decisions even though they have their political position.

For the demographics, we can see that female interviewees have higher odds of voting for the neutral candidate — James Soong. We can see this from checking out the result table. From Table XXX, we can see that the odds for voting for males decrease around 50% for voting for other candidates or choosing "don't know" or not casting a vote. We can also see that the increase in frequency does not affect the odds of voting for any candidates.

Conclusion

After discussing the three sub-datasets we separated from the primary dataset based on the parties the interviewees support, there are several conclusions we can imply. However, this analysis only focuses on the relationship between the media and th interviewees' voting decisions, rather than causation. Firstly, people watching the news from opposite news sources would affect their candidate choice. We have shown it through the result of our models above. If we only consider the Blue Party and the Green Party, the Blue Parties' supporters are more likely to change their minds based on the news outlets they watch. If we interpret the result in this way, it means that for Green supporters, more people strongly believe in the Green Party's ideology and would be less affected by the information from the media. To conclude the question that we mentioned in the Introduction section: whether people watch the news or the parties, they support confirmation bias. Based on this analysis, the result would be people do watch the news for the information and support the candidates that fit their ideology rather than support the candidates only for the party.

There are some improvements and space for future work available for this work. Most importantly, the sample size would be a concern. Since we separated this data into three datasets based on the parties they support, the dataset with 2002 observations was divided into three small datasets. The better way to do this would be to have a survey dedicated to this research and sample by cluster sampling based on the party people support. Another concern of this analysis would be the way we decide the media inclination of the interviewees. Even though there is a question in the survey asking which party do the news channels that the interviewees watch support, the reason for not using it is because of the concern of the result being subject and would affect our interpretation's accuracy. Thus, a potential solution for this concern would be changing the levels that we assigned to the interviewees into different inclination levels rather than only Green, Blue, and Neutral. By doing this, we believe that we can obtain more insight from this analysis.

References

Library References

- Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra
- Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A (2019). "Visualization in Bayesian workflow." *J. R. Stat. Soc. A*, *182*, 389-402. doi: 10.1111/rssa.12378 (URL: https://doi.org/10.1111/rssa.12378).
- Hadley Wickham, Jim Hester and Romain Francois (2018). readr: Read Rectangular Text Data. R package version 1.3.1. https://CRAN.R-project.org/package=readr
- Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.
- Hadley Wickham (2016). ggplot2: Elegant Graphics for Data Analysi. Springer-Verlag New York
- JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2020). rmarkdown: Dynamic Documents for R. R package version 2.3. URL https://rmarkdown.rstudio.com.
- Kay M (2020). *tidybayes: Tidy Data and Geoms for Bayesian Models*. doi: 10.5281/zenodo.1308151 (URL: https://doi.org/10.5281/zenodo.1308151), R package version 2.1.1, <URL: http://mjskay.github.io/tidybayes/>.
- Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01
- Paul-Christian Bürkner (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal, 10(1), 395-411. doi:10.32614/RJ-2018-017
- Paolo Di Lorenzo (2020). usmap: US Maps Including Alaska and Hawaii. R package version 0.5.1. https://CRAN.R-project.org/package=usmap
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- Sing T, Sander O, Beerenwinkel N, Lengauer T (2005). "ROCR: visualizing classifier performance in R." *Bioinformatics*, *21*(20), 7881. <URL: http://rocr.bioinf.mpi-sb.mpg.de>.
- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.29.