Why Elon Musk Buys Twitter and Why Donald Trump Lanchs Truth Social?*

Social media becomes more and more important in affecting the outcome of voting in US election

Yinuo Zhang

27 April 2022

Abstract

study aims to investigate whether social media like twitter, facebook is indeed affect the US election based on the Democracy Fund + UCLA Nationscape survey data by logistic regression model. Logistic regression model is a simple but widely used model which is appropriate for modelling binary outcome and interpreting factors affect the binary outcome. This study finds that not usage group of socila media twitter and facebook is about 45% times higher in odds of voting for Trump than who uses socila media twitter and facebook and we have 95% confidence this would fall into the range from 13% to 86%.

Contents

*Code and data are available at: https://github.com/Amy527/finalproj304.

Introduction

Recently, in addition to the Russian Ukrainian war occupying the hot spot of world news, another hot spot is that Elon Musk has become the largest shareholder of Twitter and wants to wholly acquire twitter. During the general election more than a year ago, Trump was blocked by twitter due to inciting people to make trouble on Capitol Hill. As we all know, musk has been close to the Republican Party for many years, After trump was elected as president in 2016, he immediately invited musk to join the presidential Advisory Committee. Therefore, some Republican lawmakers hope musk can let twitter lift the ban on Trump's account. Since Trump lost the last election, he has repeatedly proposed to prepare for the next election in public, Trump even has recently launched the social media platform truth social.

It is known that Musk's relationship with Democratic President Biden is not good for many reasons. For example, musk believes that trade unions will affect productivity which is obviously different from the basic line of the Democratic Party. We all know that no social media can beat twitter in terms of rhythm in the election process. Therefore, whether Trump launches truth social or Musk buys twitter, it indicates that the outlet of the next US election must be social media, and this is likely to indicate that Musk will strongly support the Republican Party in the next US election, and may even run for president himself. Musk's reason for running for president is also very simple. Musk is a very potential leader and genius, many very powerful projects such as neuralink, SpaceX and Starlink star chain were founded by Musk. However, Musk also has an ultimate dream of colonizing Mars, but no matter how rich and appealing he is, colonizing Mars is not a task that an enterprise can complete, so musk may run for president to use the power of the whole country to realize the dream of colonizing Mars. As Twitter is known as the mouthpiece of political public opinion in the United States, buys twitter means that Musk will be easier to lead the voting tendency of American voters in the presidential election.

Based on the above backgroud, this study aims to investigate whether social media like twitter, facebook is indeed affect the US election based on the Democracy Fund + UCLA Nationscape survey data (Tausanovitch and Vavreck (2020)) by logistic regression model. Logistic regression model is a simple but widely used model which is appropriate for modelling binary outcome and interpreting factors affect the binary outcome. This study uses whether voters voted Trump in 2020 US election as response which is binary outcome and this study also interested in how the related factors affect the outcome of the 2020 US election especially for the usage of social media twitter, facebook. This study applied the Logistic regression model to investigate the question of interest and obtained the estimates of effects of social media twitter, facebook on voting outcome for Trump based on the Democracy Fund + UCLA ationscape survey data (Tausanovitch and Vavreck (2020)).

The study is important because we would get estimates of effects of social media twitter, facebook on voting outcome for Trump to investigate whether social media like twitter is important in affecting US election, we could use the results to explain why Trump launched the social media platform truth social and why Elon Musk wants to wholly acquire twitter. Also, this study might be a good start point for people who are interested in this topic area for their work.

The study was organized as following: In data section, we introduce the Democracy Fund + UCLA ationscape survey data and give an overview of reflections from the survey data. In model section, we introduce the logistic regression model. In results section, we interpret the model results as well as model diagnostics. At last, in the discussion section, we discuss the findings, results and weaknesses of the study as well as future work. The study is carried out using R (R Core Team (2020)), Rmarkdown (Allaire et al. (2021)), tidyverse (Wickham et al. (2019)), ggplot2(Wickham (2016)), ggthemes(Arnold (2021)), ggrepel(Slowikowski (2021)), dplyr (Wickham et al. (2021)), scales (Wickham and Seidel (2020)), broom (Robinson, Hayes, and Couch (2021)), knitr (Xie (2021)).

Data

The source of the data used in this study comes from the U.S.presidential election 2020 survey data by Democracy Fund + UCLA Nationscape released on 12 Jan, 2021 (Tausanovitch and Vavreck (2020)). This study uses the survey data to build a logistic regression model to interpret the effects of social media like twitter on the outcome of voting for Trump after controlled several important factors like age, gender, income level, race, voted outcome in 2016 and so on.

Age differences played an important role in the history of US election, this study also included it and divided into 4 groups: 18-35, 36-50, 51-65 and 65+ which is already widely used by lots of former studies and supposed to best describe the differences between age groups. Gender difference is also important, males tend to be more conservative. Race is important too, it is known whites support Trump more than others. This study merged difference races and divided races into 3 groups: white, black and other group. It is similar for hispanic or not. And for income, richer people like Trump more than others, Trump supported by richer people not only due to himself is rich but also he has lots of policy beneficial for rich voters. This study uses the median income level which is around 39999 dollars in the survey to divide the voters, voters who are above this level is marked as “above median” which stands for rich voters and voters who are below this level is marked as “median or below” which stands for poor voters. For education, voters with different education levels also have different attitudes on voting for Trump too. This study divided voters into 3 groups: High school or below, BA or below and Above BA. Voted outcome in 2016 is also included, it might be useful in describing the changes or consistency of attitudes on voting for Trump between 2016 and 2020. Also, this study considers a new covariate which is whether voters worried ban of twitter, there are 2 groups: worried and not worried, this factor can be used to describe the degree of attention on social media like twitter. The usage of social media like twitter and facebook is used as the main independent variable of interest. At last, the response is created from the outcome of voting in 2020, voted Trump is encoded as 1, otherwise encoded as 0.

The survey data comes from comes from the U.S.presidential election 2020 survey data by Democracy Fund + UCLA Nationscape released on 12 Jan, 2021 (Tausanovitch and Vavreck (2020)). The sampling method used in the survey is the stratified sampling, the voter study group in Nationscape divided the potential voters into subgroups by age, race, gender, education and so on, and then the voter study group applied simple randomly sampling in each of the stratified group. The population of the survey is the United States voters. The sampling frame is first dividing voters into subgroups by age, race, gender, education and so on using stratified sampling and then pplied simple randomly sampling in each of the stratified group. The sample unit of the survey is US adult who is supposed to be a voter and who conducted the survey online.

The sample size of the sample by the UCLA Democracy Fund Voter Study Group is 4138 originally, after data cleanning, this study uses a cleaned version with 3023 voters. There are originally over 260 variables, this study considers 12 of them. The survey offered two language options - Spanish or English. The survey was conducted online anywhere the respondent has access to networked device.

The survey was designed by the UCLA Democracy Fund Voter Study Group and was aimed to cover lots of possible factors which could affect the voting and reflect the voting behavior of voters. There are many vriables included to make the survey to be a high quality one and reliable one. It also includes different plotical opinions like medical health care, ban of gun and so on as well as covid-19 related information.

The survey was reliable and could be a good sample of US voters because it was carefully designed by the UCLA Democracy Fund Voter Study Group, although it was conducted online, it was desigend weekly by the Voter Study Group so that the Voter Study Group could compare the results among weeks to select reliable answers of the survey. People who have too fast answers which completed in several seconds or too short time would be dropped, people who have similar answers across the questionnaires would be dropped too. The Voter Study Group also removed people who did not complete the survey. Also the Voter Study Group designed many types of question to make the survey as accurate as possible.

However, the survey also has some weaknesses. First, it was an online survey, it still could have selection bias either due to division of subgroups and selection process. Second, people received the online survey might not answer it or people might forgot to answer it which lead to non-response bias. Third, there is some imbance of subgroups in the survey sample that voters from some states is more than others. At last, the survey can only ensure people complete the survey not too fast or not with too many similar answers, as the survey has too many questions and it was conduceted weekly online, it is hard to make sure people answer all questions seriously.

Figure 1 shows Donald Trump supporters' profiles in the survey, clealy, older white males are the main supporters for Trump while the non-supporters are maninly blacks.

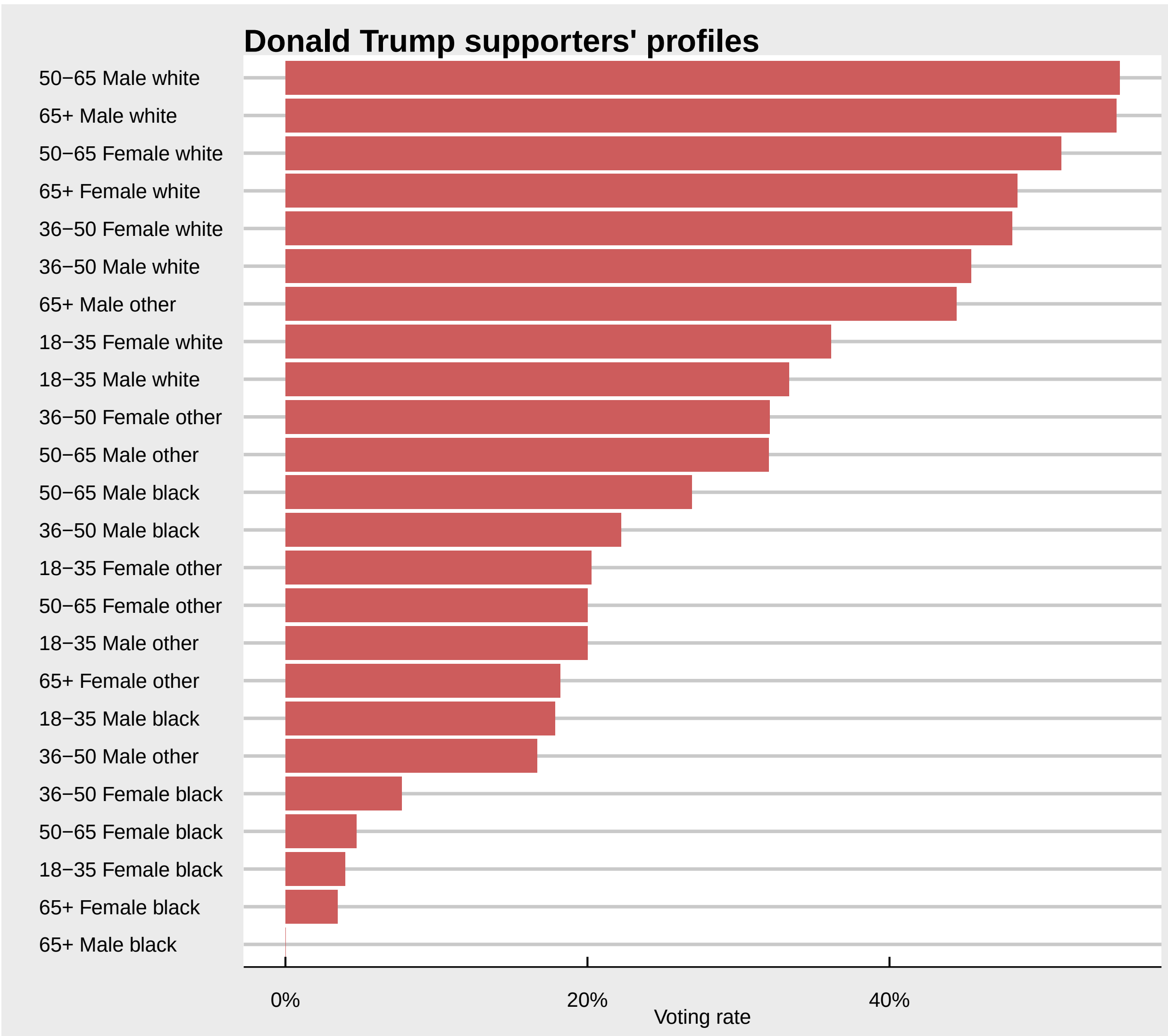

Figure 1: Donald Trump supporters' profiles in the survey

Figure 2 shows Donald Trump supporters' profiles by states in the survey, clearly, the differences in votes of Trump among states are obvious, it can be found that some voters in states like wyoming are most likely to vote for Trump.

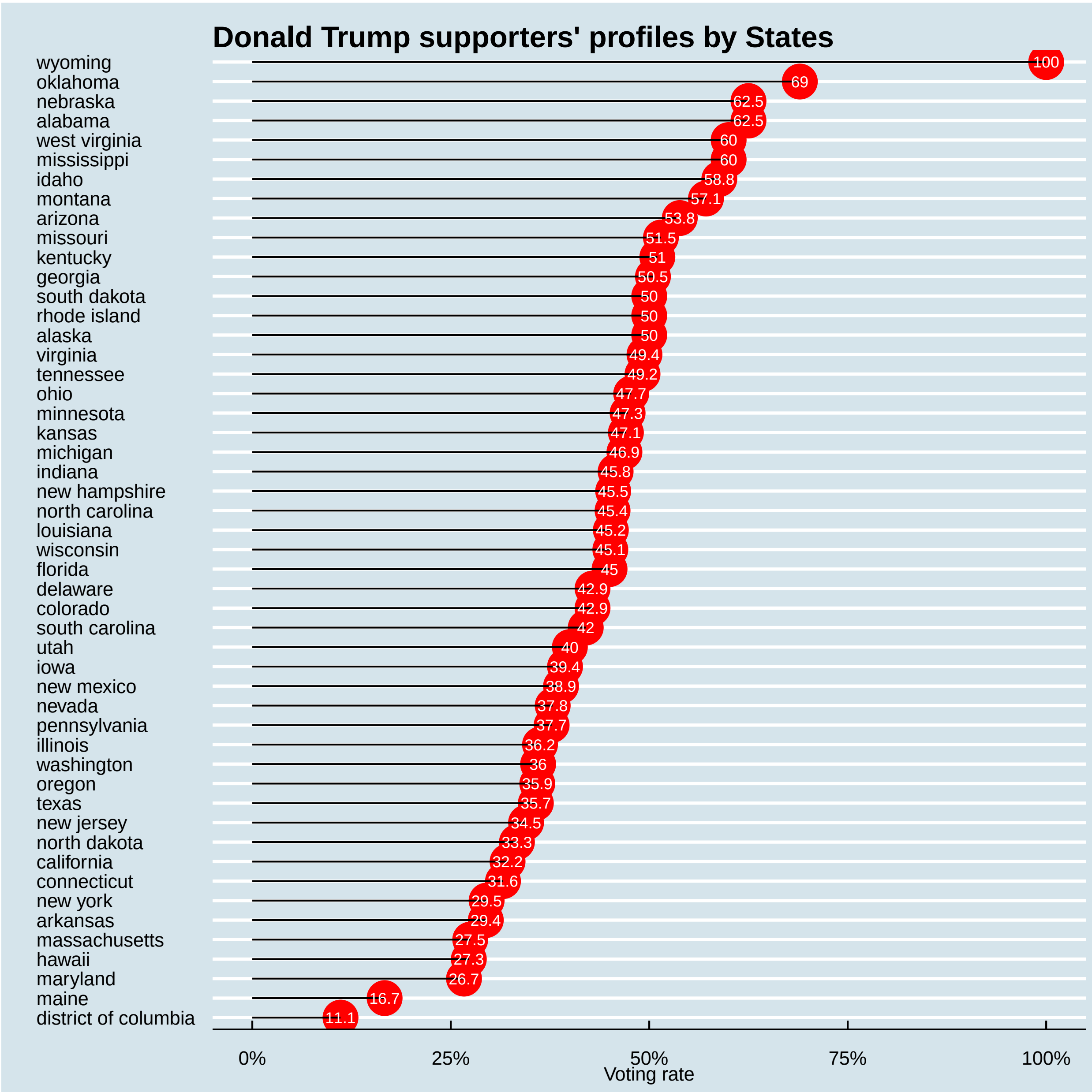

Figure 2: Donald Trump supporters' profiles by states in the survey

Figure 3 shows the Donald Trump supporting rate by States in 2016 and 2020, clearly, there is a strong positive linear relationship between the vote rates in 2016 and 2020 for Trump. It indicates the supports across years are consistent.

Figure 4 shows the Donald Trump supporting rate in 2020 vs. Percentage of usage social media like twitter,

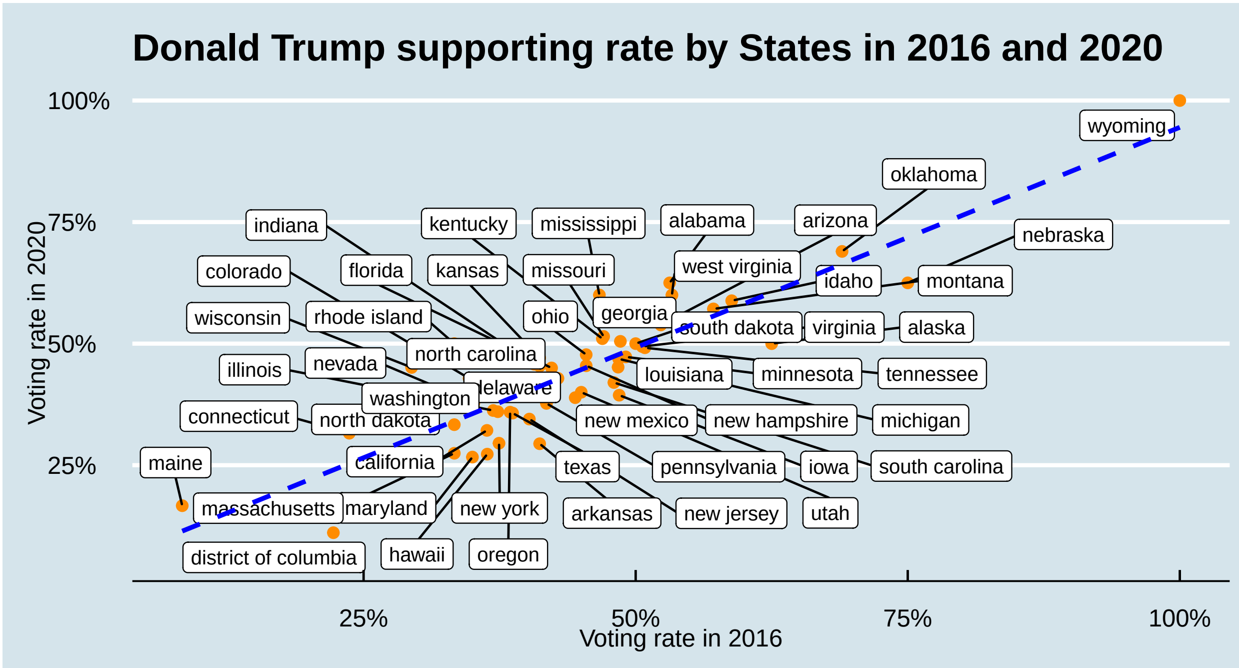

Figure 3: Donald Trump supporting rate by States in 2016 and 2020

facebook. Clearly, there is also a high correlation, there is a strong positive linear relationship between supporting rate in 2020 vs. Percentage of usage social media like twitter, facebook. It means that more usage of twitter, facebook higher likely in supporting Trump.

Figure 5 shows the Donald Trump supporting rate in 2020 vs. Percentage of worried ban of twitter, also, people who are more care about ban of twitter is obviously more likely to support Trump. Thus, twitter is important in US election.

Figure 6 shows the Donald Trump supporting rate grouped by income, race and social media twitter&facebook usage, clearly, for both high and low income levels, for all race groups, using social media twitter&facebook group has a lower rate for Trump, it indicates Trump needs to occupy twitter to change twitter so he can own better supporing rate.

Figure 7 shows the Donald Trump supporters' age distribution grouped by usage of social media twitter, facebook, clearly, younger voters below about 40 years old are more likely not to support Trump who are using social media twitter&facebook than those who are not using social media twitter&facebook, it indicates Trump needs to change twitter to own supports from younger voters.

Figure 8 shows Donald Trump supporters' age distribution grouped by worried of twitter ban, clearly, voters who are more care about ban of twitter is more likely not to support Trump, it indicates Trump needs to change twitter to own supports from these groups of voters too.

Figure 9 shows Donald Trump supporting rate grouped by vote group in 2016 and social media twit-

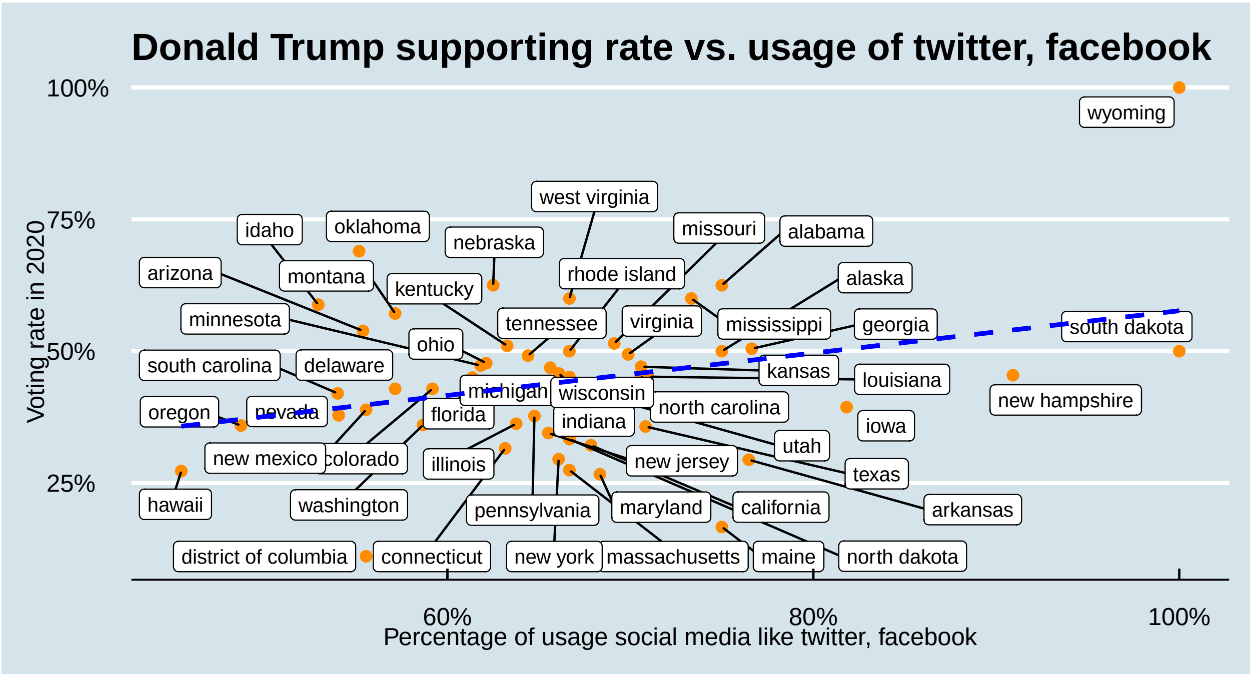

Figure 4: Donald Trump supporting rate in 2020 vs. Percentage of usage social media like twitter, facebook

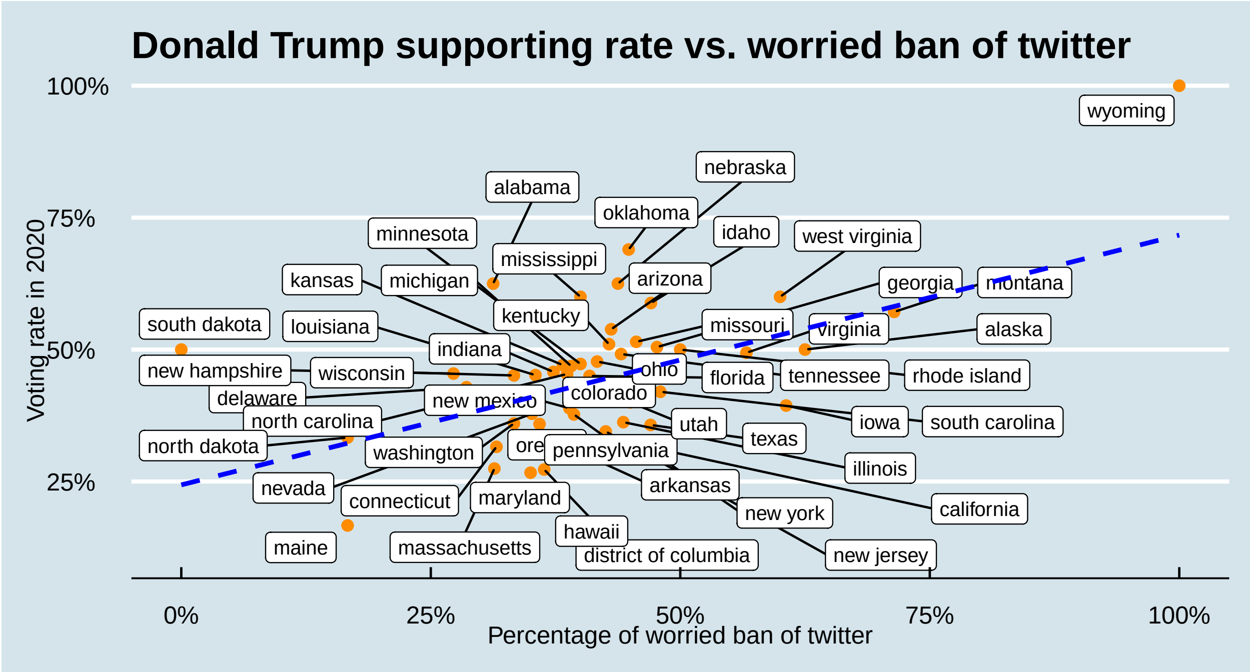

Figure 5: Donald Trump supporting rate in 2020 vs. Percentage of worried ban of twitter

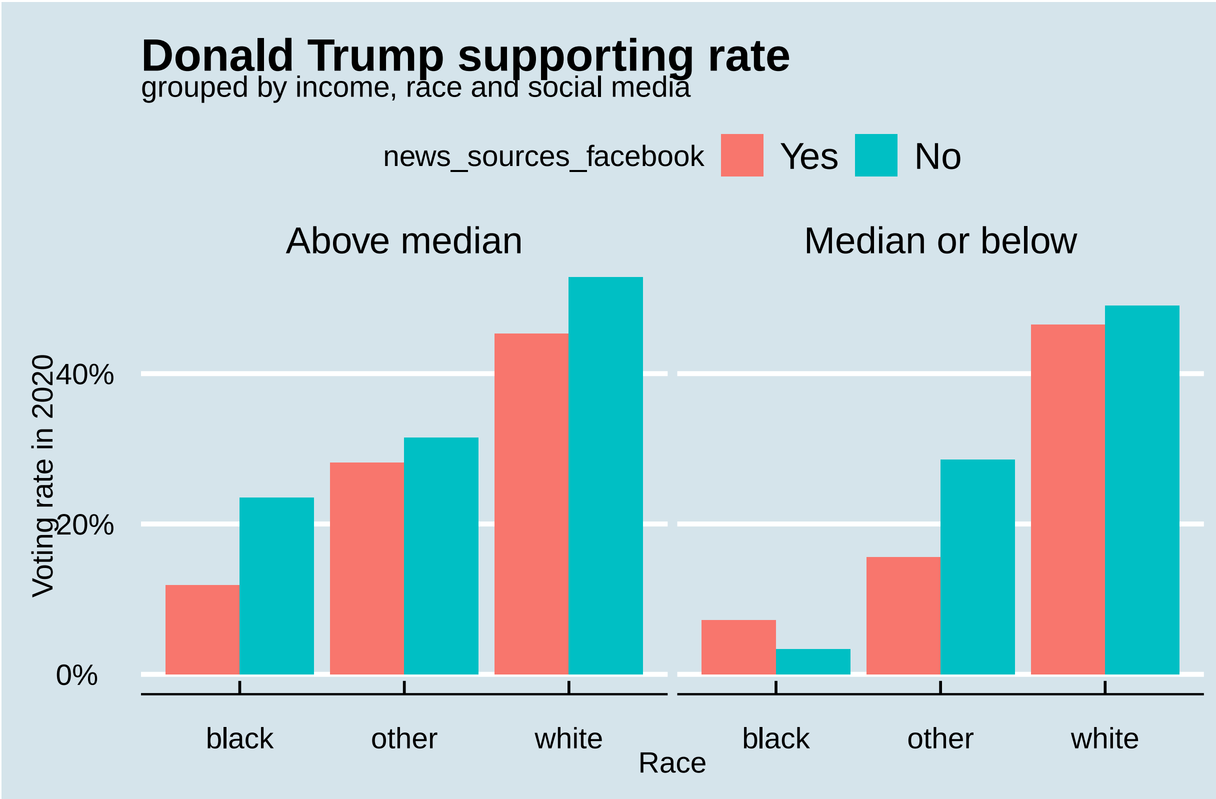

Figure 6: Donald Trump supporting rate grouped by income, race and social media twitter&facebook usage

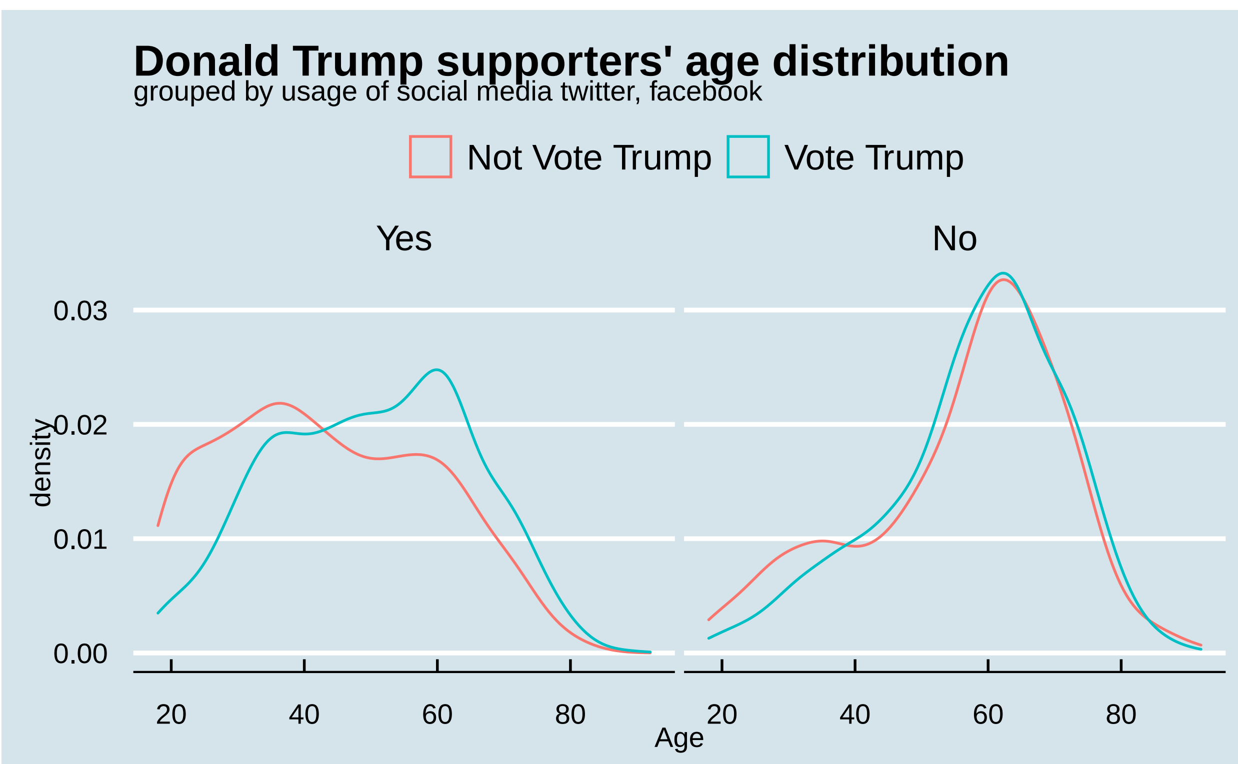

Figure 7: Donald Trump supporters' age distribution grouped by usage of social media twitter, facebook

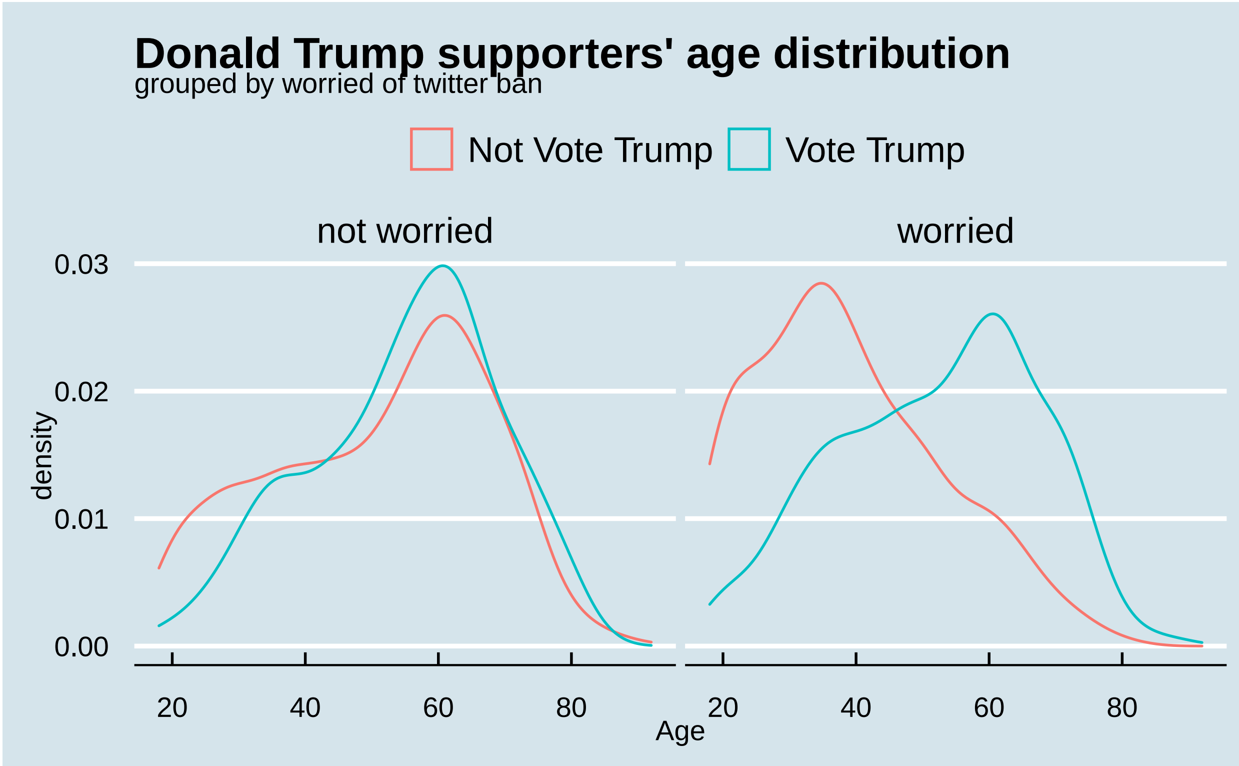

Figure 8: Donald Trump supporters' age distribution grouped by worried of twitter ban

ter,facebook, clearly, poeple who ever voted for Trump in 2016 has changed not to vote him in 2020 is more likely happend in group of using social media twitter,facebook.Thus, Trump needs to change twitter to re-own their supports.

Overall, we could find that social media twitter,facebook and selected covariates are indeed related with voting outcome for Trump in 2020, so we can model the outcome on these factors.

Model

This study plans to use the multiple logistic regression based on the Democracy Fund + UCLA Nationscape survey data (Tausanovitch and Vavreck (2020)). Logistic regression model is a simple but widely used model which is appropriate for modelling binary outcome and interpreting factors affect the binary outcome. This study uses whether voters voted Trump in 2020 US election as response which is binary outcome, this study uses the usage of social media twitter, facebook as the main independent variable of interest, this study also uses several important covariates including age, gender, income level, education leve, race groups, voted outcome in the year 2016 and so on. This study would use the Logistic regression model to investigate the effects of social media twitter, facebook on voting outcome for Trump.

The equation (1) shows the math form of the logistic model. The logistic model is appropriate as it is

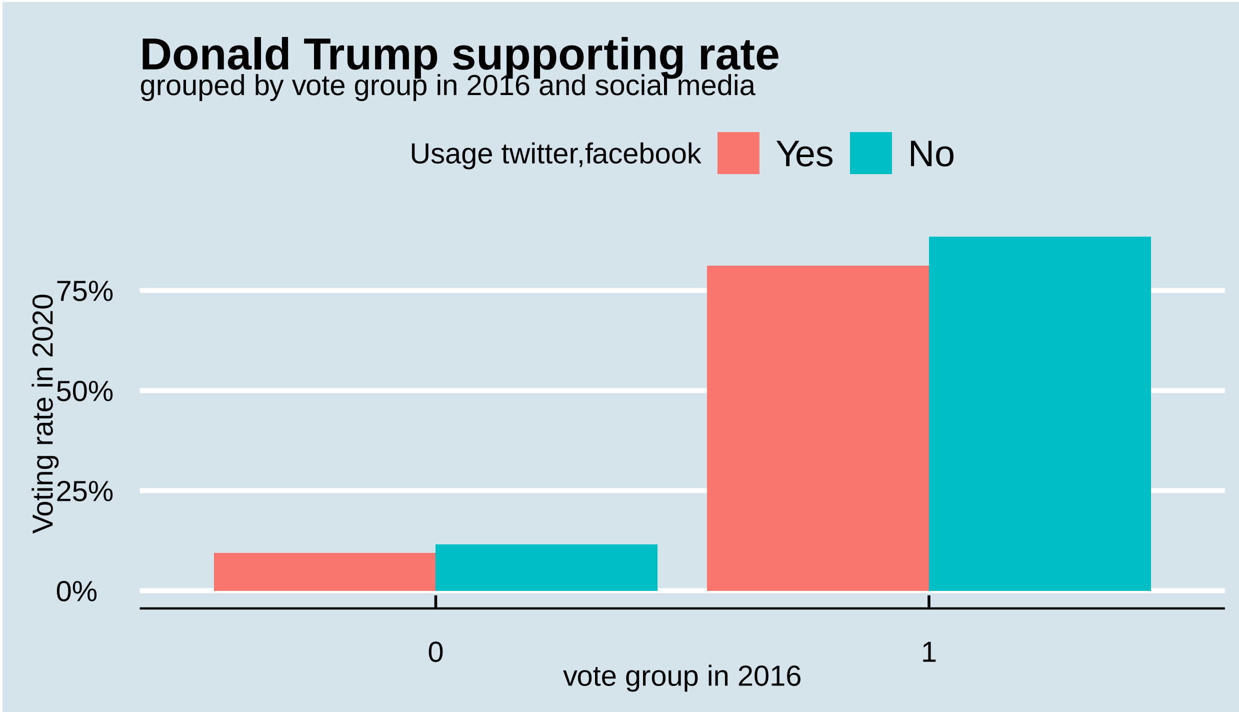

Figure 9: Donald Trump supporting rate grouped by vote group in 2016 and social media twitter,facebook

appropriate for modelling binary outcome and interpreting factors affect the binary outcome, in this study, we encoded outcome voting for Trump as 1 and otherwise as 0 which is a binary dummy variable, also we want to know the effect of social media twitter, facebook on voting outcome for Trump and it can be explained by the estimates from the logistic model.

$$logit(p) = \beta_0 + \beta_1 age + \beta_2 gender + \beta_3 race + \beta_4 race + \beta_5 income + \beta_5 education + ... + \beta_i twitterusage + \epsilon \quad (1)$$

Another form is:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 age + \beta_2 gender + \beta_3 race + \beta_4 race + \beta_5 income + \beta_5 education + ... + \beta_i twitterusage \epsilon \quad (2)$$

Note, the p is probability of voting for Trump in 2020, β_i are coefficients to be estimated, ϵ is the error term. The estimates are log-odds of coefficients, we also can have odds ratio by using exp() function in R (R Core Team (2020)).

The estimated probabilty in voting for Trump by logistic model is in equation (3):

$$p = \frac{exp(\beta_0 + \beta_1 age + \beta_2 gender + \beta_3 race + \beta_4 race + \beta_5 income + \beta_5 education + ... + \beta_i twitterusage)}{1 + exp(\beta_0 + \beta_1 age + \beta_2 gender + \beta_3 race + \beta_4 race + \beta_5 income + \beta_5 education + ... + \beta_i twitterusage)} \quad (3)$$

We fit this model on the Democracy Fund + UCLA ationscape survey data (Tausanovitch and Vavreck (2020)) use the `glm` function in R (R Core Team (2020)) with the logit link and then interpre the results to investigate the question of interest in this study.

The main advantages of using logistic model are as below:

1). Fast enough for large sample

2). Easy to understand of the model estimates

3). Easy to include variables as many as possible

4). Easy for using both continuous variables like age or categorical variables like gender, income level, education level

5). Good model's interpretability of effects of factors

6). Nature for vote election problem which has response to be binary outcome

The main dis-advantages of using logistic model are as below:

1). Not suitable for non-linear effects

2). Multi-collinearity issues

3). Sensitive to imbalanced responses

4). Worse performance in predicting probability

5). Can not handle random effects

However, in this study we do not consider random effects of individual voters which would make the model too complicated as there are too many voters. The assumption of logistic model is not normality of errors but independence of errors and large sample size. We have large sample size which is over 3000, also, we have independence of errors as the data for voters which are supposed to be independent.

Also, note in this study that, the model has some weakness. First, not all covariates are included, so the results might have omitted important variables biasness. Second, only Biden and Trump are considered, no other candidates, the voting group encoded as 0 might be some complicated as it stands for voting for Biden not voting for other candidates who is not Trump. At last, the results of the model dependend on the survey data seriously, different data lead to different estimates.

Results

The reference group in the model is the young voters 18-34 age group female hispanic and blacks with above median income level and above BA education level who are not voted for Trump and who used social media twitter and facebook but not worried ban of twitter.

The results of estimates are lots for dummy variables, this because the logistic model would created dummies for the categorical variables, the baseline group and the difference with the baseline group are the estimates of log-odds for these dummies. For example, the variable "genderMale" is compared with females which means the difference between females and males in voting for Trump when we fixed effects of all other groups.

Table 1 and figure 10 show the Logistic model estimates in log-odds and the table 2 shows the Logistic model estimates in odds ratio. From the table 1 and figure 10, we can easily find which variables have negative effects and which ones have positive effects on the response as well as the 95% confidence intervals with the effects. However, it is more easier to explain the effects of variables in odds ratio than log odds by using table 2. Figure 11 shows the model diagnostics, clearly, the residuals plot show no big issues, the cook's distance shows few strong influence points.

Table 1: Logistic model estimates in log-odds

term	estimate	std.error	statistic	conf.low	conf.high
(Intercept)	-4.840	0.350	-13.840	-5.538	-4.166
agegroup36-50	0.370	0.171	2.166	0.035	0.705
agegroup50-65	0.700	0.169	4.140	0.370	1.033
agegroup65+	0.634	0.197	3.223	0.249	1.020
genderMale	-0.495	0.121	-4.106	-0.734	-0.261
raceother	0.863	0.303	2.848	0.274	1.462
racewhite	1.307	0.231	5.645	0.865	1.774
incomegroupMedian or below	0.071	0.125	0.569	-0.174	0.318
educationBA or lower	0.711	0.175	4.058	0.369	1.055
educationHigh school or Lower	1.054	0.201	5.251	0.662	1.449
hispanicnot hispanic	0.025	0.199	0.123	-0.364	0.417
Trump2016	3.582	0.121	29.684	3.349	3.822
news_sources_facebookNo	0.372	0.128	2.914	0.122	0.623
twitter_banworried	1.122	0.123	9.088	0.881	1.365

Table 2 shows that Logistic model estimates in odds ratio as well the 95% Confidence intervals for the corresponding odds ratio. The odds ratios of age groups above 18-34 are all over 1s, for age group 36-50, the

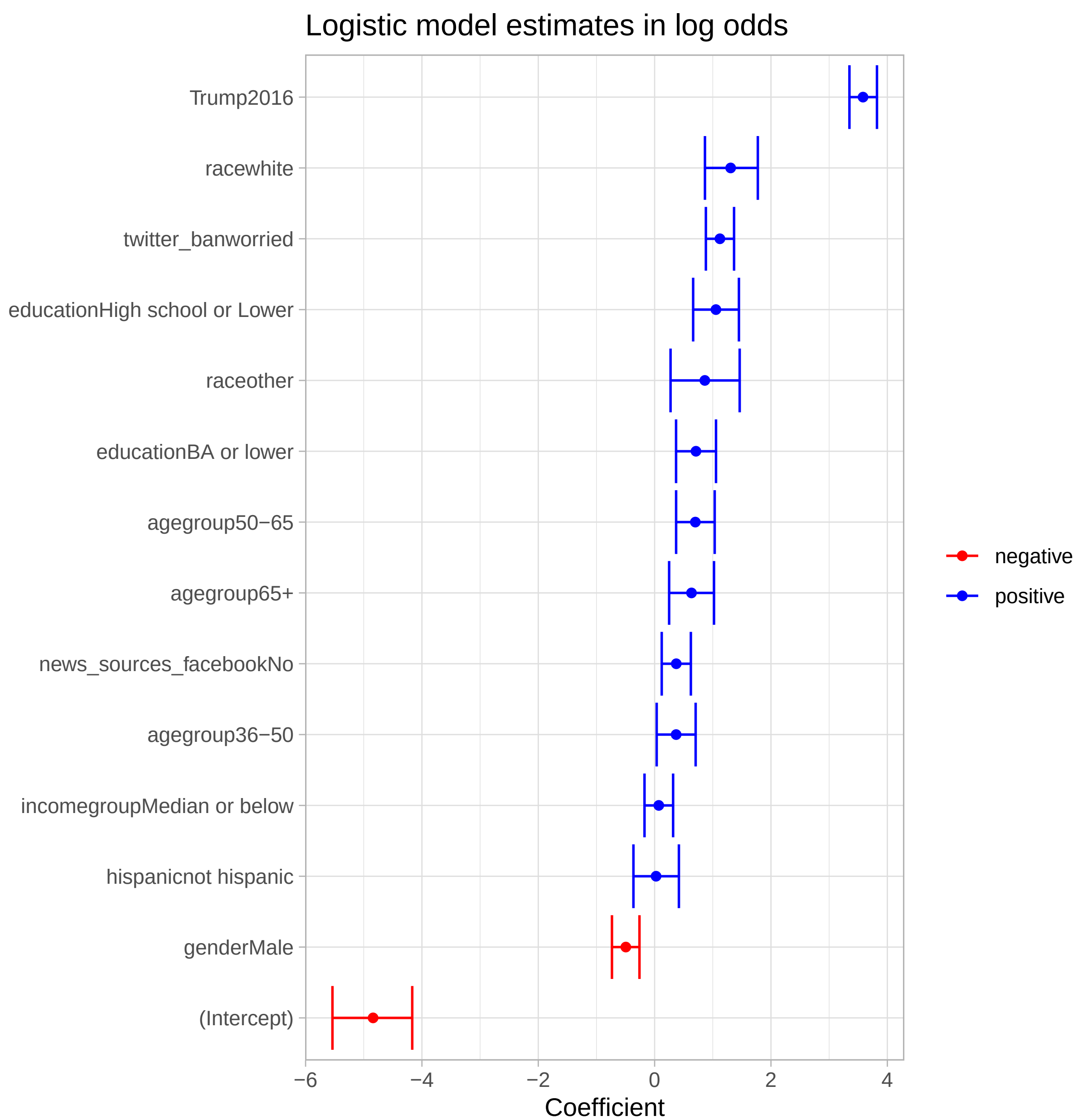

Figure 10: Logistic model estimates in log-odds

OR is 1.48, for age group 50-65, the OR is 2.01, for age group 65+, the OR is 1.88, all these groups have 95% CIs' lower bounds larger than 1, so it means the older age groups are more likely to support Trump significantly. The OR for males is 0.61 which means males is 39% times lower in odds of voting for Trump compared with females fixed others. The races whites and other groups all show OR much larger than 1 which are above 2s, so it means whites and other groups are much more likely to support Trump significantly than blacks fixed others. For income level, high income group is more likely to support Trump while low educated groups are also more likely to support Trump, non-hispanic is also more likely to support Trump.

Table 2 also shows voters who ever voted Trump in 2016 are about 35 times higher in odds to vote Trump in 2020 than who not voted Trump. This is an important variable to control. Also, the worried ban of twitter group has about 2 times higher in odds of voting Trump than who are not worried ban of twitter.

Table 2: Logistic model estimates in odds ratio

Variable	OR	conf.low	conf.high
(Intercept)	0.008	0.004	0.016
agegroup36-50	1.447	1.036	2.024
agegroup50-65	2.014	1.447	2.808
agegroup65+	1.885	1.283	2.774
genderMale	0.609	0.480	0.771
raceother	2.370	1.315	4.315
racewhite	3.695	2.376	5.894
incomegroupMedian or below	1.074	0.840	1.374
educationBA or lower	2.035	1.446	2.872
educationHigh school or Lower	2.868	1.939	4.259
hispanicnot hispanic	1.025	0.695	1.518
Trump2016	35.942	28.469	45.697
news_sources_facebookNo	1.451	1.130	1.864
twitter_banworried	3.072	2.413	3.917

And controlled all other effects, it was found from table 2 that the estimate of usage of socila media twitter and facebook has OR to be 1.45, so it means that controlling all other important variables in this model, the not usage group of socila media twitter and facebook is about 45% times higher in odds of voting for Trump than who uses socila media twitter and facebook.

Because this study aims to investigate the question that why Musk buys twitter and why Trump lanchs truth social, the hypothesis is that social media like twitter could not affect the voting willingness of voters in US election. And from the model estimates, this study rejects this null hypothesis, it was found that twitter

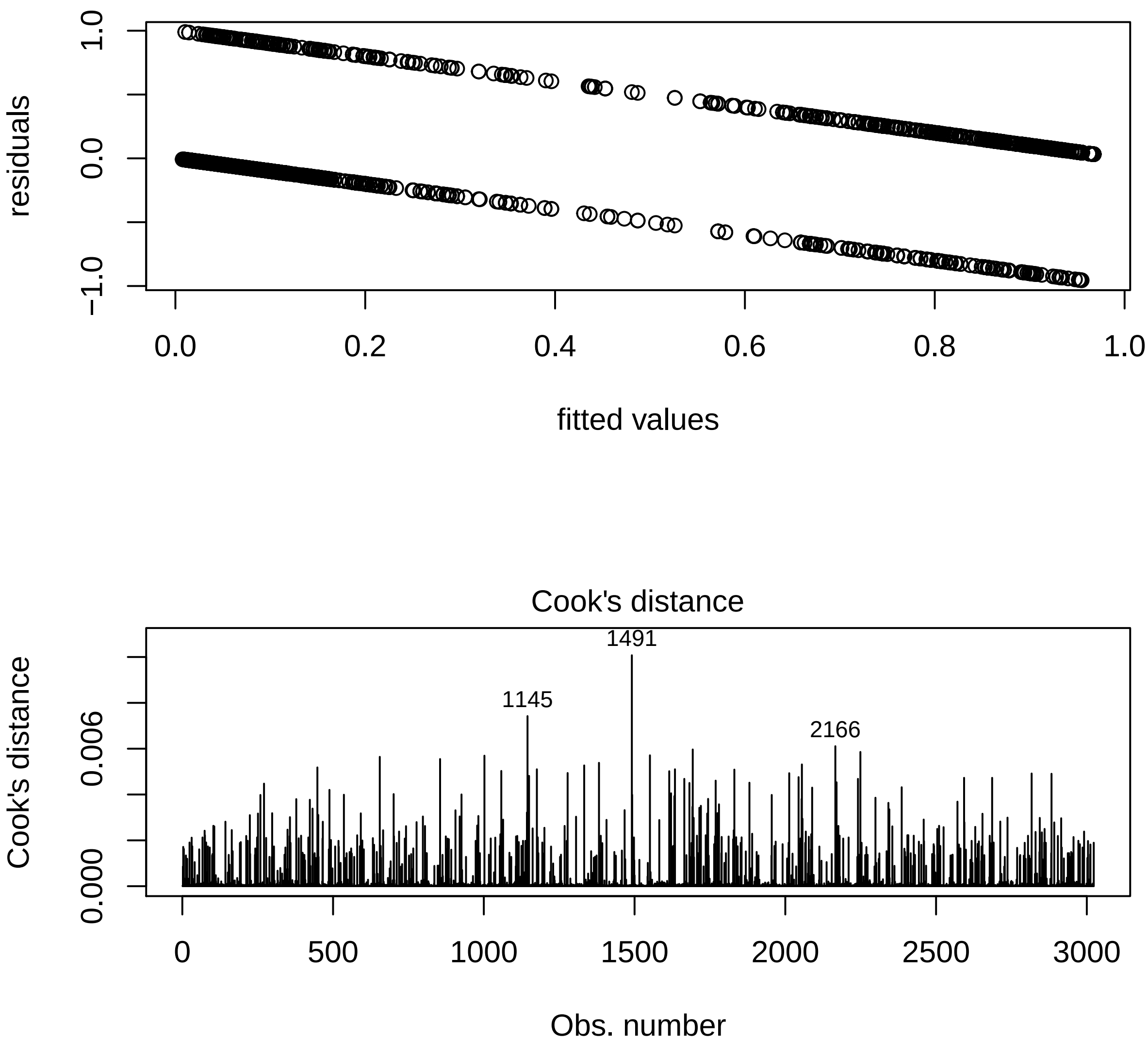

Figure 11: Model diagnostics plots

could affect the voting willingness of voters in US election significantly after controlling other important factors.

Discussion

Findings

After all of the above work in the study, the main findings from exploring analysis is that age group,gender, income level, education, race diferences are all related to voting outcome for Trump in the year 2020, also, for states, the differences are more clearly, some states are much more likely to support Trump than other states. More importantly, it was found that social media twitter,facebook and selected covariates are indeed related with voting outcome for Trump in 2020, so we can model the outcome on these factors.

And based on the model estimates, after controlling all other effects, it was found the estimate of usage of socila media twitter and facebook has OR to be 1.45 with 95% confidence interval to be (1.13, 1.86), which means that controlling all other important variables in this model, not usage group of socila media twitter and facebook is about 45% times higher in odds of voting for Trump than who uses socila media twitter and facebook and we have 95% confidence this higher difference would fall into the range from 13% to 86%.

Because this study aims to investigate the question of interest why Musk buys twitter and why Trump lanchs truth social, from the model estimates, this study rejects the null hypothesis that social media like twitter could not affect the voting willingness of voters in US election. It was found that usage of social media like twitter could affect the voting willingness of voters in US election significantly after controlling other important factors.

Importance of social media twitter and facebook

Besides this study found that usage of social media like twitter could affect the voting willingness of voters in US election significantly after controlling other important factors. With the development of networking technology, social networking sites such as Twitter are playing an increasingly important role in the US presidential election. The Twitter social networking site has become a basic campaign tool. Public candidates, political more and more use tweets to disseminate and capture ampaign information to attract voters who are very interested in politics to spread information widely and create news so in orde to earn supports from them. This might be the reasons why Musk buys twitter and why Trump lanchs truth social as Twitter provides voice and power to millions of voters and allows candidates to monitor the effects of their information and adjust it anytime on time to earn more supports.

Controlling covariates

This study used lots of Controlling covariates to adjust the effects of the usage of social media like twitter and facebook, it was found that most of these Controlling covariates are indeed affect the voting outcome of Trump in 2020. From the model estimates, for age group 36-50, the OR is 1.48, for age group 50-65, the OR is 2.01, for age group 65+, the OR is 1.88, all these groups are more likely to support Trump significantly so older groups are mode likely to vote for Trump, this because they might be affected by the slogan of Trump that makes USA great again, due to these older voters have saw the better USA in the past decades when countries like China has not been grown up. Also, there are significant differences in supports of males and females, income level, education level. Males is about 39% times lower in odds of voting for Trump compared with females fixed others, also, low educated groups are also more likely to support Trump. This might be due to the reference group used in this study and imbalance data in the study, generally, males is more likely to vote Trump. The races whites and other groups all show OR much larger than 1 indicating whites and other groups are much more likely to support Trump significantly than blacks fixed others. This is consistent with findings from most of studies, also, rich people is likely to support Trump. Due to these covariates have significant effects in affecting estimates, it is reasonable to include them in the study as controlling covariates.

Weaknesses and next steps

There are also some weakness for this study. First, the survey data might be selection bias and non-response bias, as the survey was conducted online, voters could not respone to it lead to on-response bias and as the survey uses stratified sampling, several subgroups in this survey data might be has selection biasness. Second, imblanced data issues existed in this survey data, for examples, the proportions of whites are much higher than others.

Also, the model results might have omitted important variables bias, although this study considered lots of related factors and show they are indeed needed to be inclueded, however, there are still other important variables should be considered, these variables might be opinions to policies like ban of gun, medical health, covid-19 issues, tax and so on. There is also one weakness that the model depend on cleaned data, we divided several groups for variables in this study, different divisions could affect model estimtes clearly.

In future work, we might consider more other important variables like opinions to policies like ban of gun, medical health, covid-19 issues, tax and so on. We might also consider mixed effects model using states as groups, we could expect the results to be improved in the next steps in future studies.

Appendix

Additional details

Table 3: Voting rates, usage twitter,facebook, worried ban of twitter percentages by states

stateicp	rate	rate2016	news_sources_facebook	twitter_ban
alabama	0.625	0.531	0.750	0.312
alaska	0.500	0.625	0.750	0.625
arizona	0.538	0.523	0.554	0.431
arkansas	0.294	0.412	0.765	0.412
california	0.322	0.363	0.678	0.453
colorado	0.429	0.347	0.592	0.388
connecticut	0.316	0.237	0.632	0.316
delaware	0.429	0.429	0.571	0.286
district of columbia	0.111	0.222	0.556	0.556
florida	0.450	0.423	0.614	0.409
georgia	0.505	0.486	0.766	0.477
hawaii	0.273	0.364	0.455	0.364
idaho	0.588	0.588	0.529	0.471
illinois	0.362	0.369	0.638	0.443
indiana	0.458	0.407	0.661	0.373
iowa	0.394	0.485	0.818	0.606
kansas	0.471	0.412	0.706	0.382
kentucky	0.510	0.469	0.633	0.429
louisiana	0.452	0.484	0.710	0.355
maine	0.167	0.083	0.750	0.167
maryland	0.267	0.350	0.683	0.350
massachusetts	0.275	0.333	0.667	0.314
michigan	0.469	0.484	0.656	0.391
minnesota	0.473	0.491	0.618	0.400
mississippi	0.600	0.467	0.733	0.400
missouri	0.515	0.471	0.691	0.456
montana	0.571	0.571	0.571	0.714

stateicp	rate	rate2016	news_sources_facebook	twitter_ban
nebraska	0.625	0.750	0.625	0.438
nevada	0.378	0.351	0.541	0.351
new hampshire	0.455	0.455	0.909	0.273
new jersey	0.345	0.402	0.655	0.425
new mexico	0.389	0.444	0.556	0.389
new york	0.295	0.374	0.661	0.480
north carolina	0.454	0.412	0.706	0.387
north dakota	0.333	0.333	0.667	0.167
ohio	0.477	0.455	0.621	0.417
oklahoma	0.690	0.690	0.552	0.448
oregon	0.359	0.385	0.487	0.359
pennsylvania	0.377	0.418	0.648	0.393
rhode island	0.500	0.333	0.667	0.500
south carolina	0.420	0.480	0.540	0.480
south dakota	0.500	0.500	1.000	0.000
tennessee	0.492	0.508	0.644	0.441
texas	0.357	0.387	0.708	0.470
utah	0.400	0.450	0.700	0.450
virginia	0.494	0.506	0.699	0.566
washington	0.360	0.373	0.587	0.333
west virginia	0.600	0.533	0.667	0.600
wisconsin	0.451	0.294	0.667	0.333
wyoming	1.000	1.000	1.000	1.000

References

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2021. *Rmarkdown: Dynamic Documents for r*. https://github.com/rstudio/rmarkdown.

Arnold, Jeffrey B. 2021. *Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2'*. https://CRAN.R-project.org/package=ggthemes.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2021. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Slowikowski, Kamil. 2021. *Ggrepel: Automatically Position Non-Overlapping Text Labels with 'Ggplot2'*. https://CRAN.R-project.org/package=ggrepel.

Tausanovitch, Chris, and Lynn Vavreck. 2020. *Democracy Fund + UCLA Nationscape*. October 10-17, 2019 (version 20200814). https://www.voterstudygroup.org/publication/nationscape-data-set.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.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.

Wickham, Hadley, Romain Fran?ois, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Dana Seidel. 2020. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.