Can we predict United States' 2020 Presidential Election with Twitter Data

By:
Ahmad Maseeh Faizan
Student Number : 1009767469
Email : Ahmad.Faizan@mail.utoronto.ca

ECO225: Big-Data Tools for Economists
University of Toronto
Department of Economics
5 April 2023

Abstract

The last few elections in the US have demonstrated how social media has been used to shift the election balance to a certain candidate. I study how Twitter's analytics data, like the number of tweets mentioning Donald Trump and Joe Biden respectively, and the number of likes and retweets of those respective tweets to predict the likelihood of each candidate winning. The prediction is analyzed at state-level, but we also go deeper at county level for our regressions. I find that the number of Tweets mentioning each candidate is not well suited to predict the actual election for each state. The number of interaction that each of those states get, through likes and retweets that the tweet mentioning each candidate, does seem to have some correlation with the election outcome for certain states. It appears that the correlation does not hold for all the US States. These bias and might be explained through demographic or cultural differences in the underlying voters and who might or might not be on the platform creating these results. We also observe that Trump does have higher Tweet count throughout the country, while Biden has a higher interaction rate through retweet count. At the county-level, our regressions do not seem to indicate any useful information.

Introduction

The Internet from its genesis was a communication tool. It helped people all around the world to connect easily and efficiently. Information now is spreading at a pace and scale never seen before but it also has negative impacts like the spread of fake news regarding elections and public health matters (*Hunt Allcott, al. 2017*[1]). Ordinary people can share their thoughts on social media platforms such as Twitter. We have also seen the rise of technology of the social media platforms being increasingly used in recent elections (*John H Aldrich 2015*[2]). In this

[1] DOI: 10.1257/jep.31.2.211

[2] https://doi.org/10.1177/1354068815605304

study, I will focus our attention on the United States 2020 presidential election. Most studies do sentiment analysis, analyzing the content of the tweet like the (*Chaudhry 2021*[3]). This analysis however covers the amount of tweets, likes and retweets (shares) of tweet that mention either candidate, Joe Biden, or Donald Trump. Instead of looking into the sentiment analysis of the tweets, the goal of the study is to analyze data and to simply see whether there is a correlation between these analytics and their interaction with the tweets mentioning the candidates and the actual election outcome. I found out that the number of likes and retweets are highly correlated and will group them together as the number of interactions.

I am focusing on popular voters' outcome rather than Electoral Colleges even though they do not always have the same outcome, they are highly correlated. Sometimes, popular votes don't win the election in all the states, but considering popular vote is a good proxy for the actual presidential winner. In the study, we find that the number of interactions through likes and retweets does correlate with better election outcome for both candidates. These results are however, not applicable to all the states. Counting the number of tweets mentioning each candidate has little, no correlation with voter outcome in all the states. At the county level, our data is very unreliable and is heavily biased towards zero. In other words, most of the counties do not tweet or retweet even though they are active in voter outcome, while other counties have higher number of Twitter activity with varying degree of voting power. We see, however, that apart from the number of tweets, there is a positive bias towards Biden. In other words, people are talking more about trump through tweets, while Biden is getting most of the traction through the number of likes and retweets. I also found that income per capita of each state correlates very well with the election outcome and is fairly associated with the number of retweets. In what follows, we will discuss how the twitter analytics data correlates with election outcome and attempt to explain why.

[3] **https://doi.org/10.3390/electronics10172082**

Body

The Context and Data:

The data[4] at hand was collected towards the end of year 2020, from Twitter and contains tweets with #Trump and #Biden for both candidates which will allow us to draw comparison between both candidates and draw insights on how they perform relative to actual election outcome. The initial data contains tweets from all around the world, but since we are only interested in the United States voters, we will constrain our data with only the population of interest which is the US residents. Before we could analyze further how well our Twitter analytics data is able to predict the election outcome, we will have to know the actual outcome that we will then be able to compare. I am therefor importing actual election outcome Data[5] collected from Federal Election Commission, which contains both popular and electoral college vote to each state in the United States. The original twitter data contains a good deal of useful information such as latitude and longitude, username, and state location. With this data, I was able to describe the data by visualizing maps and other plots. The maps were build using geographic data collected from the US Census Data[6]. For the quantitative analysis, I am computing and adding to the data, total number of tweets, like and retweets for each candidate and further computing those variables in relative terms for each state. Similarly, I am computing relative popular vote outcome for each state. Comparing each variable in relative terms helps the analysis because it helps to have all the values in the same unit. It means that the data will not be biased due to population size. To enrich this data, I further scrapped income data from *Wikipedia*[7] for each state and combining this new data with our initial analysis, was extremely valuable. Trying to have a richer data, by adding another social media data such as Facebook would have also been

[4] www.kaggle.com/datasets/manchunhui/us-election-2020-tweets

[5] www.fec.gov

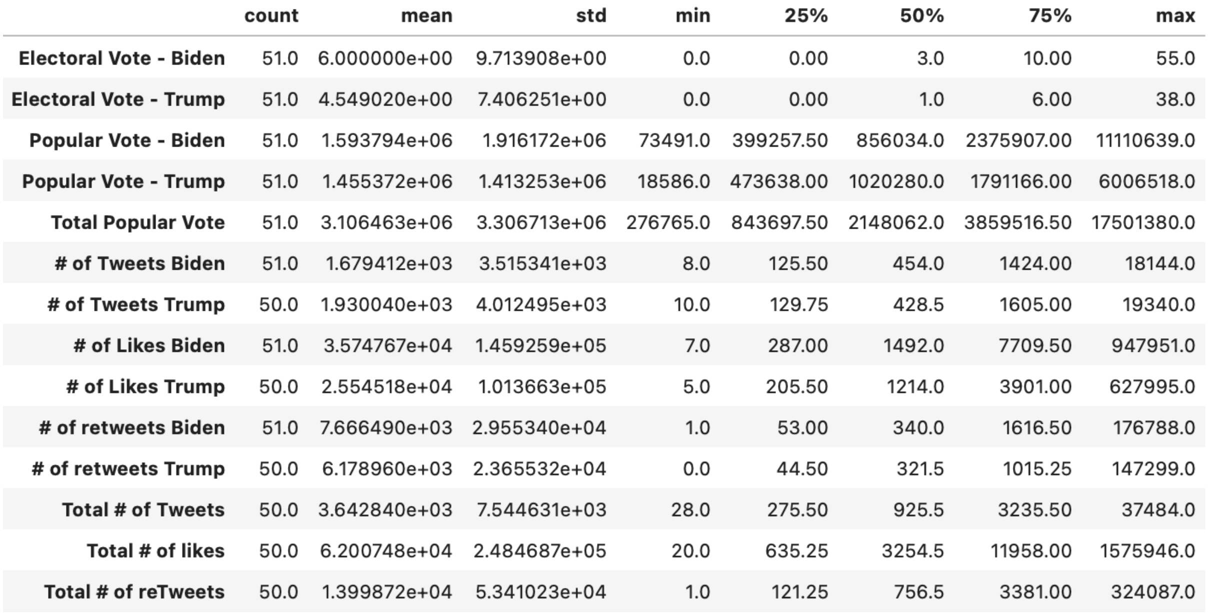
[6] www.census.gov/

[7] https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_income

valuable, however, Facebook policy[8] doesn’t allow its data to be collected freely online like Twitter and Reddit data was very resource intensive for the purpose of this study, which meant that I was not able to have an overall analysis of different social media platforms. I also collected and included COVID-19 data[9] collected from the Center of Disease Control and prevention (CDC) website, since this was an election held during a world-wide pandemic and had a huge impact on the outcome.

Summary statistics & Findings:

Figure 1



	count	mean	std	min	25%	50%	75%	max
Electoral Vote - Biden	51.0	6.000000e+00	9.713908e+00	0.0	0.00	3.0	10.00	55.0
Electoral Vote - Trump	51.0	4.549020e+00	7.406251e+00	0.0	0.00	1.0	6.00	38.0
Popular Vote - Biden	51.0	1.593794e+06	1.916172e+06	73491.0	399257.50	856034.0	2375907.00	11110639.0
Popular Vote - Trump	51.0	1.455372e+06	1.413253e+06	18586.0	473638.00	1020280.0	1791166.00	6006518.0
Total Popular Vote	51.0	3.106463e+06	3.306713e+06	276765.0	843697.50	2148062.0	3859516.50	17501380.0
# of Tweets Biden	51.0	1.679412e+03	3.515341e+03	8.0	125.50	454.0	1424.00	18144.0
# of Tweets Trump	50.0	1.930040e+03	4.012495e+03	10.0	129.75	428.5	1605.00	19340.0
# of Likes Biden	51.0	3.574767e+04	1.459259e+05	7.0	287.00	1492.0	7709.50	947951.0
# of Likes Trump	50.0	2.554518e+04	1.013663e+05	5.0	205.50	1214.0	3901.00	627995.0
# of retweets Biden	51.0	7.666490e+03	2.955340e+04	1.0	53.00	340.0	1616.50	176788.0
# of retweets Trump	50.0	6.178960e+03	2.365532e+04	0.0	44.50	321.5	1015.25	147299.0
Total # of Tweets	50.0	3.642840e+03	7.544631e+03	28.0	275.50	925.5	3235.50	37484.0
Total # of likes	50.0	6.200748e+04	2.484687e+05	20.0	635.25	3254.5	11958.00	1575946.0
Total # of reTweets	50.0	1.399872e+04	5.341023e+04	1.0	121.25	756.5	3381.00	324087.0

Now that our dataset set up and ready, we want a first impression of our dataset. First looking at the mean of Popular vote in Figure 1, Joe Biden did win both popular and Electoral Vote, which, based on this first impression, leads us to believe in saying that usually higher vote count did lead to election win, during this election. Joe Biden did get on average 1’594'794 at

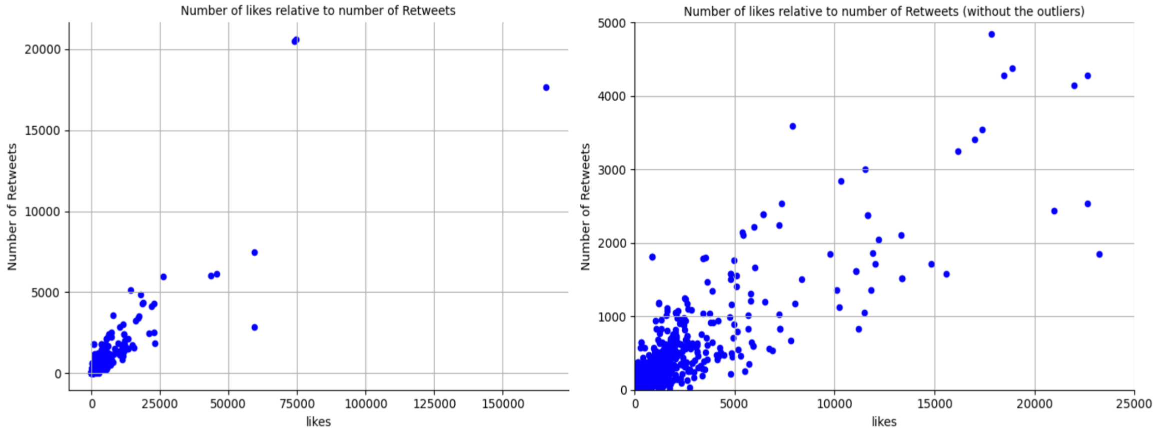
[8] https://developers.facebook.com/docs/development/terms-and-policies/automated-data-collection/

[9] https://covid.cdc.gov/covid-data-tracker/#cases_casesper100k

state level compared to Trump who got 1’455'372 votes. In our study moving forward we will be looking into the Popular vote for both of our candidates and comparing them to number of tweets, retweets and likes. Looking at the averages for all of our data points, we can see there is a visible correlation between the popular vote for each candidate and the number of retweets and likes that the tweets of each candidate got. The only Twitter analytic metric where Trump win is the number of Tweets where on average Trump got 300 extra tweets at state level, from around 1'680 for Joe Biden to 1'900 for Trump. To explain this phenomenon, one of the hypothesis is that, Trump is a more controversial figure where he gets most of the attention. People will be talking about him more than Biden, but that does not necessarily mean that people agree with him. We can see that by looking at the amount of traction that his tweets gets on average. We have a very good measure of traction for all the tweets, which is the number of likes that specific tweet gets. In this department, Biden on average has higher likes even with lower initial tweet count. This implies that people agree or relate more with tweets regarding Biden, even if they are fewer than the overwhelming amount of tweets about Trump, where they relatively get fewer likes.

Before, we try to predict the election outcome through twitter analytics, we want to confirm our claim that the number of like and retweets are strongly correlated, so we could run estimations on the number of retweets on election outcome and know that the number of likes runs on a similar pattern. To do so, we will run a scatter plot, where on the horizontal axis we have the number of likes and on the vertical axis, the number of retweets.

Figure 2

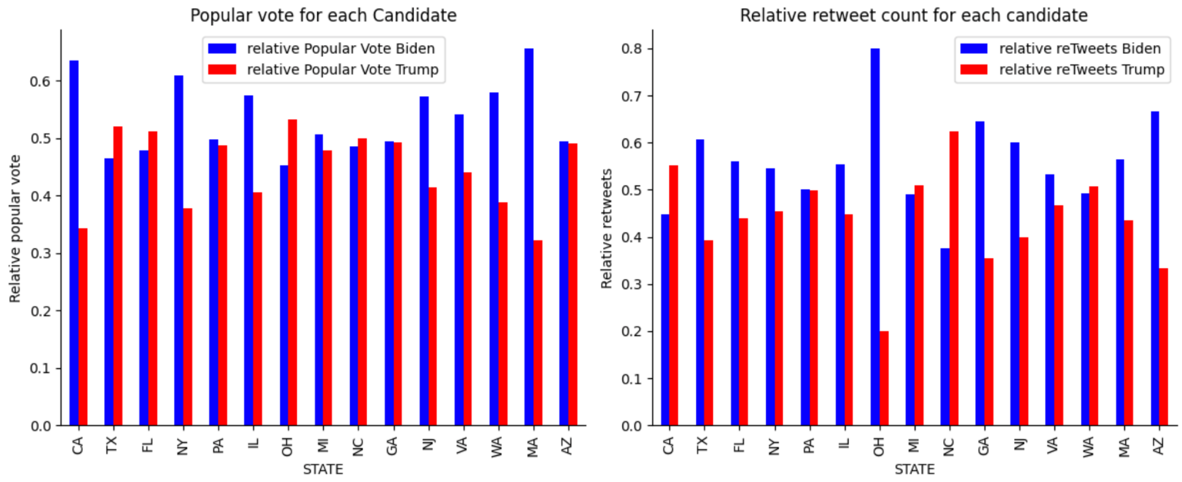


The figure, on the left, shows how the number of likes and retweets correlate with one another. However, we have few outliers that we need to control for. After controlling for these outliers, we get the scatter plot showed on the right. Most of our observations are around 0 and 5000 likes and between 0 and 1000 retweets. It is not unusual to see a higher share of like than retweets, but what is important is the proportional growth. The figure shows that there is a strong correlation between both variables, which means we can consider one of them further in our analysis and have an insight on the other.

Voter outcome and Twitter metrics for most populous states

Now, let’s consider the top 15 states with the highest voter outcome. These are big states, with high voting power like California, New York, Florida, or Texas. We want to visually see how the relative popular vote compares to the relative retweet count for each candidate in these states.

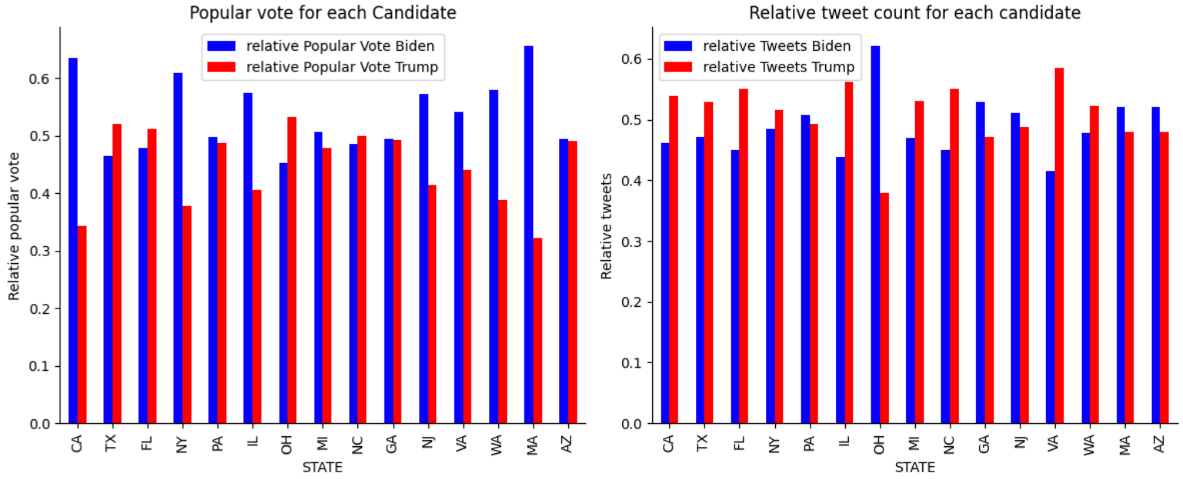
Figure 3 :



The blue bar is the relative vote for candidate, Joe Biden, and the red bar for Donald Trump. The plot on the left shows the relative popular vote for each candidate. We are measuring these variables in relative terms, because this will enable us to have all the data in similar units. The plot on the right shows the relative retweets for the same states. Looking at this data, surprisingly, we find an inverse correlation that emerges. For example, in California (CA) and Florida (FL) voters voted for Biden however, the relative vote count shows that tweets about trump had higher relative retweets. On the other hand, we can see the inverse for Texas (TX) and many other states. For other states, like New York (NY) or Illinois (IL) the relative retweet count does correlate with voter outcome. This makes us question the relative retweet count and how well is it able to predict the voter outcome. We also observe that the relative number of retweets is quite higher for Joe Biden than for Donald Trump across many states.

We, therefore, want to see how the number of tweets compare with relative voter outcome. We find that even for the number of tweets, there isn't an obvious pattern that emerges.

Figure 4:



A similar pattern emerges, for instance in California (CA) and New York (NY) the relative tweet count is higher for Trump relative to Biden, even though Biden wins the election. However, in Texas (TX) and Florida (FL) the number of tweets mentioning each candidate, does relate to the actual voter outcome measures. Furthermore, there is an overall higher number of tweets about Trump than Biden.

One of our hypotheses to explain these values is that there are a higher number of tweets about Trump than Biden, which I believe comes from the fact that Trump is a controversial figure and attract more attention than Biden does. The correlation between the number of interaction (i.e., retweet level) and voter outcome is not obvious at state level. Most of the states has a favorable outcome for Joe Biden in our Twitter data then Trump, even when Trump wins in terms of popular votes. This, I believe, is a sign of a bias towards Biden. When only analyzing Twitter data we only get the data for certain demographic which mainly consists of Urban and young demographic, or most celebrities and politician. Twitter is a social media platform which is generally considered being polarizing[10] and doesn't actually have any political alliance[11] as often mentioned in the news. Nevertheless, the site is popular among young and urban demographic who are more likely to vote for Biden and are most importantly more likely to agree and interact with his tweet rather than those of Trump.

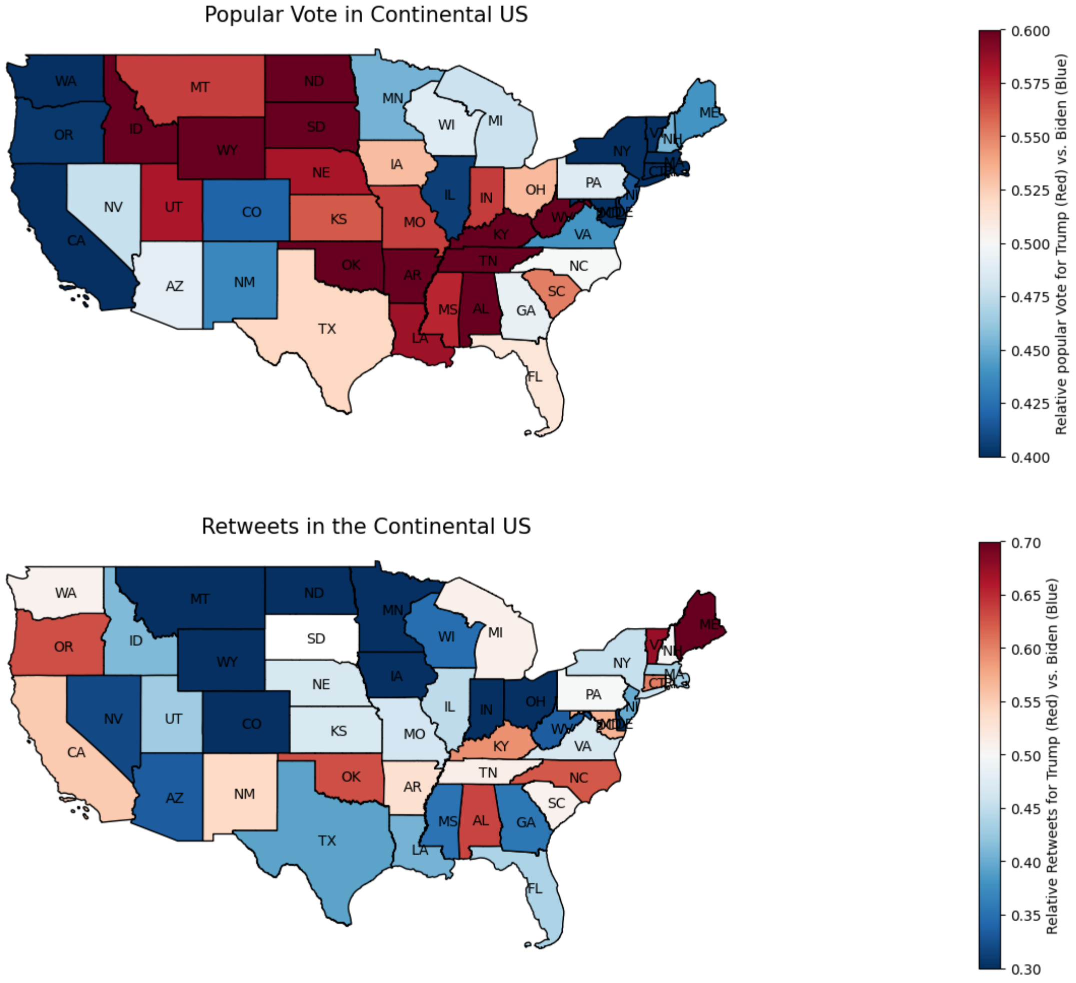
[10] doi: 10.1007/s10796-021-10222-9

[11] https://doi.org/10.1038/s41467-021-25738-6

Voter outcome and Twitter metrics for all states

To further see whether our hypothesis holds through all the states, not just for the most populated ones, we would need to investigate the continental US and have a visual representation of how each state compares. First, let's compare the popular vote outcome for each candidate and the relative retweet count through a Map. The map on the top shows the relative popular vote outcome for the continental US. Looking into the heat plot, the red states indicate that they voted for Trump relative to Biden and the Blue states indicate they were more likely to vote for Biden relative to Trump. The map on the bottom indicates the relative retweet share for both of our candidates in each state with the same color structure.

Figure 5 :



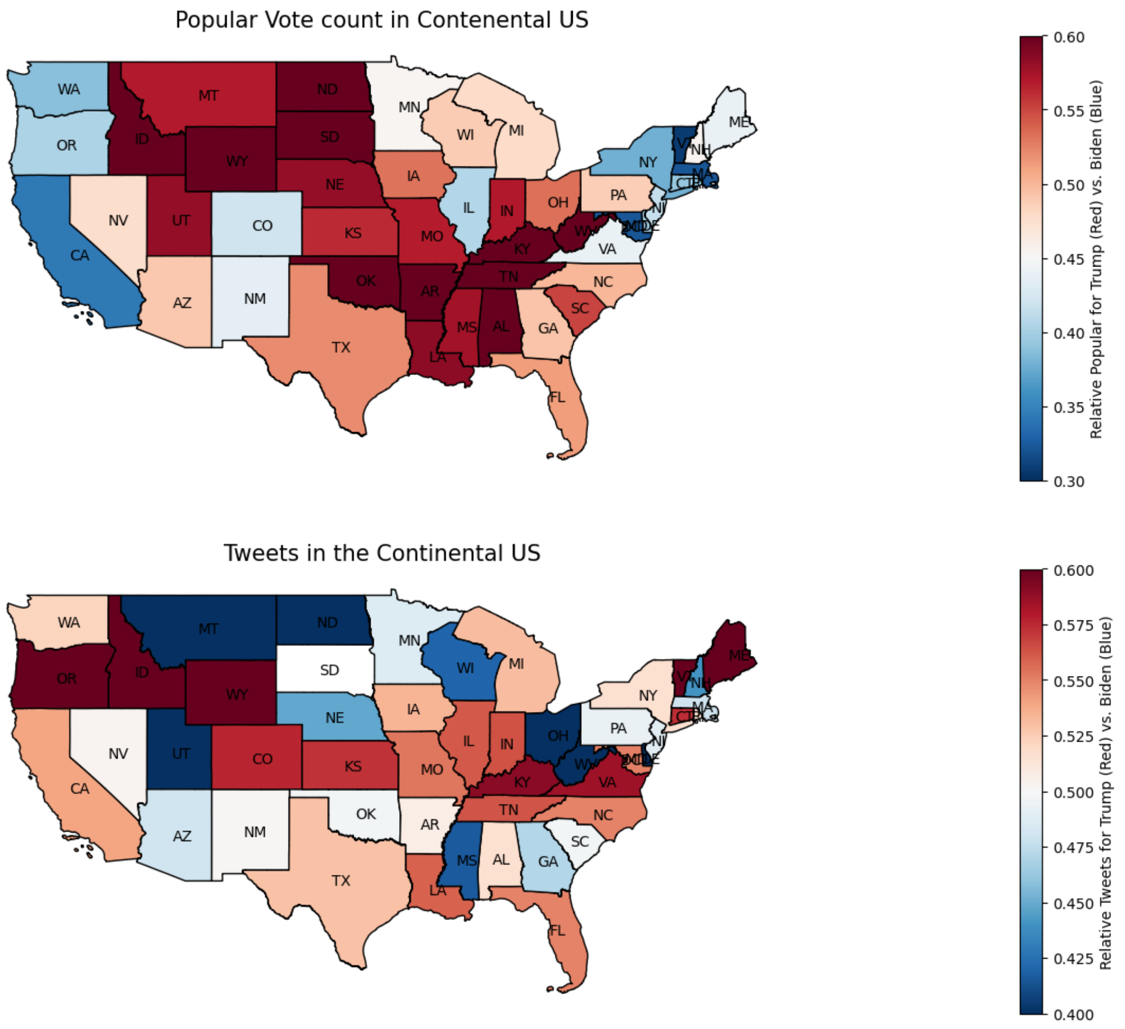
Popular Vote in Continental US

Retweets in the Continental US

Having a visual representation of the continental US with the heat plots, surprisingly, shows an opposite correlation between the relative number of retweets and relative popular voter outcome. States where people generally voted for Trump were more likely to retweet or share the tweets about Biden and vise versa, for instance California and Texas. These two states have large population and large electoral college vote. We can see that in California, people mostly voted for Biden, but the relative retweet count shows that they, were also the ones retweeting most about Trump. The inverse is True for Texas or a swing state like Florida. I believe that further analysis might be needed in terms of the content of those tweets. We can also observe that throughout the whole continental US, the bias towards Joe Biden is present. At a macro level, most of the states are more blue leaning, which indicates they are more likely to vote for Joe Biden relative to Trump.

Now, that we have analyzed the pattern of traction that each retweet about the tweet mentioning each candidate throughout each state, we will investigate the actual tweet count, with a similar map.

Figure 6 :


Popular Vote count in Contenental US

Tweets in the Continental US

Our first observation about the relative tweet count is that compared to the relative retweet count, it is more red leaning, which means most of the tweets are mentioning Trump rather than Biden across states.

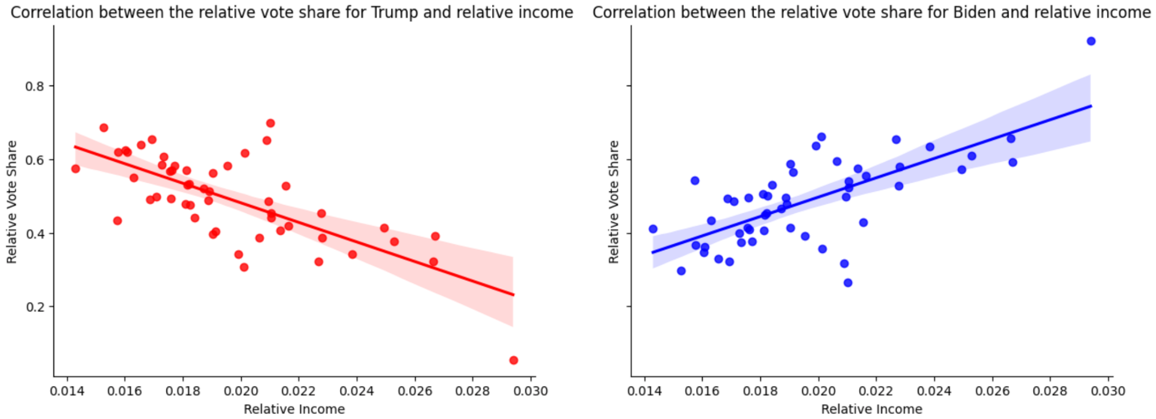
Analyzing deeper the map we see that most of the time there is an opposite pattern emerges where in states where Trump wins has a higher number of Tweets mentioning Biden like Utah (UT), while in states where Biden wins has higher number of Tweets mentioning Trump, like in California (CA), Washington (WA) or even New York (NY). This pattern, however, isn't consistent enough for us to draw any conclusions. We ought to investigate the content of these tweets and whether they were positive or negative news about the candidate in question. It is likely that people tweet and share more negative news about the opposing candidate, which makes the retweet count negatively correlated with the actual voter outcome. This implies that without looking into the content of the Tweet to see whether it is positive or negative, the number of Tweets mentioning each candidate doesn't predict well the actual outcome.

The Income effect on Voter outcome

There is some research done by *Kim Parker and Al*[12], that rural America is more likely to vote for Trump and therefor republican rather than Democrat. We have also mentioned that there might be a correlation between rural and urban demographic and twitter traffic. In other words, people who are more urban and high income are also more likely to use social media platforms like Twitter.

First, let's see the states with higher income per capita are more likely to vote for Biden rather than Trump and vice-versa. Once again, we have put all the values in relative terms, so we could analyze the data with the same units.

Figure 7:

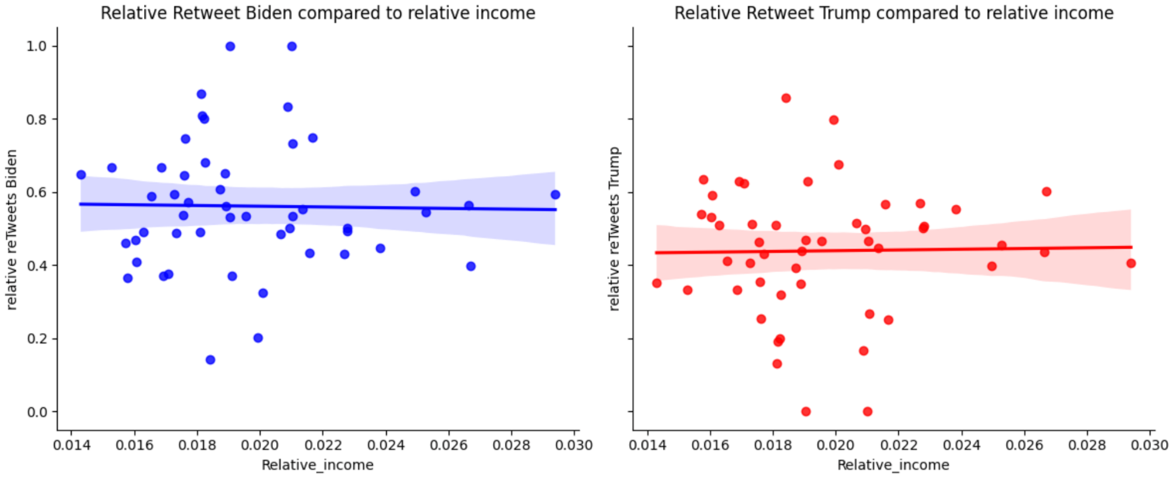


[12] https://www.pewresearch.org/social-trends/2018/05/22/urban-suburban-and-rural-residents-views-on-key-social-and-political-issues/

The scatter plot on the left shows the relative vote share of Trump compared to relative income for each state. We can see that the correlation is negative, which means states with higher income per capita are less likely to vote for Trump and therefore vote for Biden. Since all the terms are in relative terms and there are only two candidates the relative voter share compared to relative income per capita is the opposite of Trump, where high income states are more likely to vote for Biden rather than Trump.

Finally, we want to know whether income influences the number of retweets on the tweets mentioning each candidate. The horizontal axis shows the relative income for all the continental US states, while the vertical axis shows the relative retweet. The figure on the left are observations for Biden color coded in blue, while the right is color coded in red and are observation for Trump.

Figure 8 :

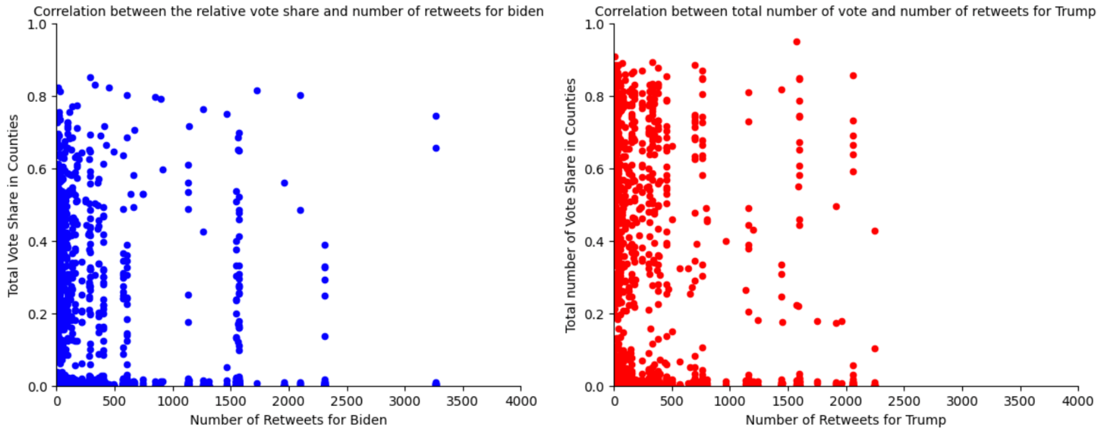


We can see that the number of retweets and relative income has little to no correlation. We are observing a straight line which indicates that both high income and low-income states are evenly likely to retweet tweets mentioning either candidate. This as another insight that shows that just counting the number of retweets mentioning either candidate is not enough to be able to predict voter outcome, and need to further analyze the content of the tweets.

Regression analysis:

As seen in the previous analysis, the retweet count, or the number of interaction that people had with the tweets, was able to better predict the outcome at state level. We will now run a regression at county level and see how will it be able to predict the election outcome.

Figure 9:



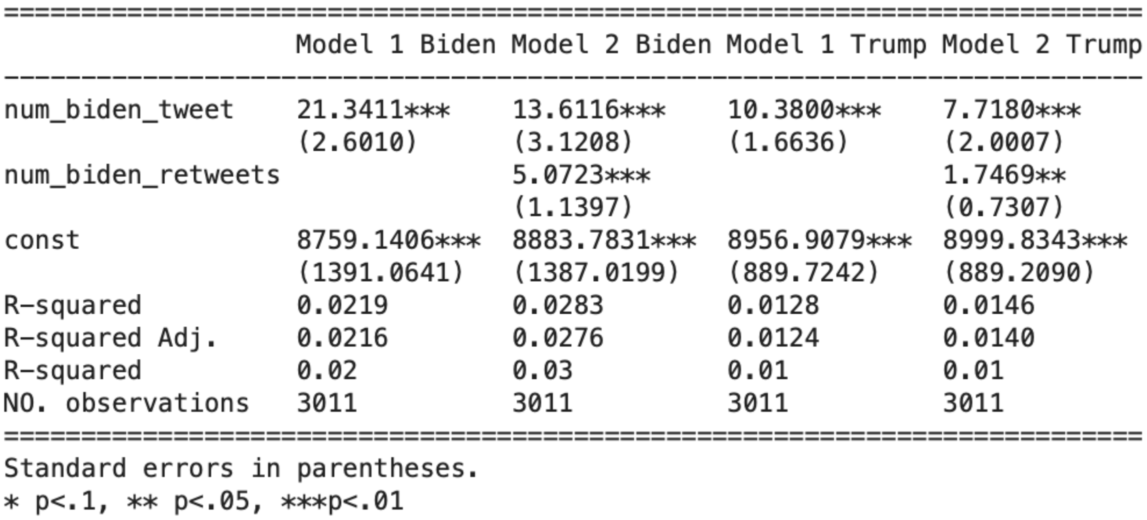
Before we run any regressions, we would first want to have an initial idea of the data at the county level. In this case, we will look into a scatter plot and see whether the data is linear or non-linear. Our initial impression doesn't tell us much. Looking at the data, we are unable to say whether the data is linear or non-linear. The data is scatter very close to zero. Most of the data is close to zero retweet count, even though the relative outcome spreads widely across the spectrum. This might be because at county level, we have very few tweets and retweets data. Even if some tweets get a lot of traction in terms or retweets, the majority lay close to zero.

Regression outcome at county level:

We are now running the following regression for both of our candidate at county level:

$$outcomei=\beta 0+\beta 1Numtweeti+\beta 2Numretweeti+ui$$

Figure 10 :



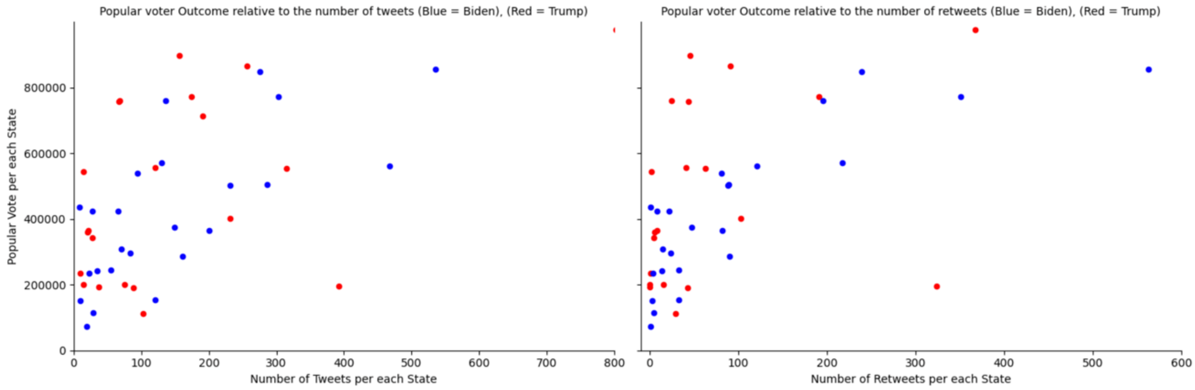
```
=========================================================================
                    Model 1 Biden Model 2 Biden Model 1 Trump Model 2 Trump
-------------------------------------------------------------------------
num_biden_tweet     21.3411***    13.6116***    10.3800***    7.7180***
                    (2.6010)      (3.1208)      (1.6636)      (2.0007)
num_biden_retweets                5.0723***                   1.7469**
                                  (1.1397)                    (0.7307)
const               8759.1406***  8883.7831***  8956.9079***  8999.8343***
                    (1391.0641)   (1387.0199)   (889.7242)    (889.2090)
R-squared           0.0219        0.0283        0.0128        0.0146
R-squared Adj.      0.0216        0.0276        0.0124        0.0140
R-squared           0.02          0.03          0.01          0.01
NO. observations    3011          3011          3011          3011
=========================================================================
Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01
```

As seen in the regression data, We can see that the data is quite similar for both our candidate. We have our intercept around 8’800, which is the average number of votes each candidate gets per county. We can then see that the number of tweet only adds around 21 extra vote count per county for Biden and only 10 extra votes for Trump. In other words, having higher number of tweets will give a candidate around 10 to 20 extra votes. This data is statistically significant at 1% significance level, but we can see that it is not economically significant. Since the number of votes only explain 10–20 votes out of 8000, we cannot be confident in the number of tweets to predict election outcome for any candidate in each county. Moreover, the number of retweets holding the number of tweets constant has even lower score of predicting the number of voter outcome in each county. It is only able to predict 5 extra votes per retweets for Biden and around 2 extra votes per retweet for Trump, which is quite low.

Regression for voter outcome at state level :

Above, we can see two scatter plots. The red dots are data points for the republican candidate Donald Trump and the blue dots are data points for the Democrat party candidate Joe Biden. We have set a limit on the x and y-axis because there are outliers which will deform the figure and will make it hard to analyze the data points. We were then able to focus our attention on the major trends with the bulk of the data points.

Figure 11 :



The figure on the left is the number of popular voter outcome relative to the number of tweets per candidate in each state. We can see that there is a positive correlation between the two variable, but the correlation seems to be non-linear. With the higher number of tweets we will also get a higher voter outcome, but this trend gets much flatter the closer we get to the number of tweets equal to around 500

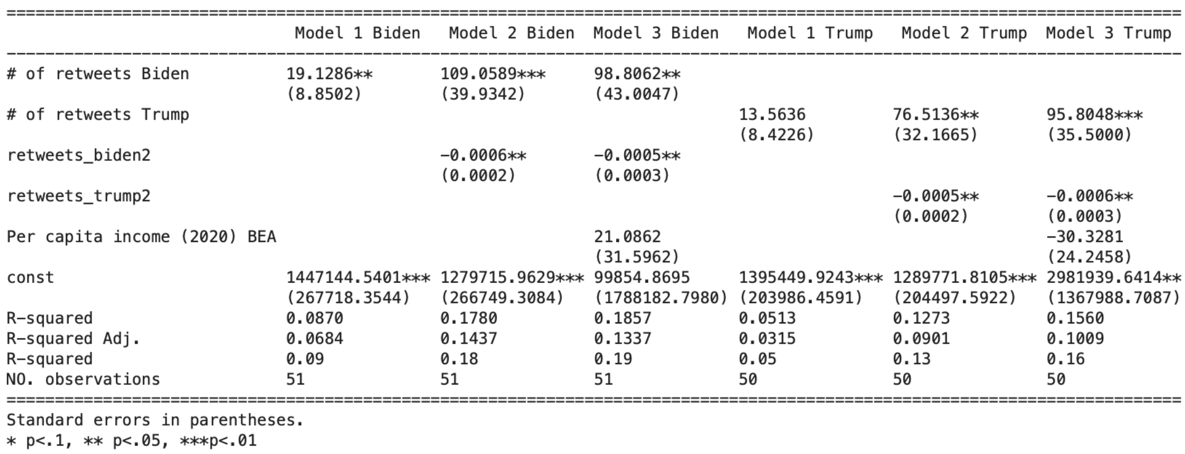
In the figure on the right, we can that the y-axis is the same, but we want to see how the voter outcome compares with the number of retweets or the number of interactions with the tweet. Here we can see that there is a similar story as the figure on the right, but the trends are much more pronounced. We can see that the correlation is positive and even more non-linear.

These scatter plots show that there is a correlation between the number of tweets and integration to those tweets and voter outcome, but this correlation is non-linear. After a certain number of tweets or retweets, there is no positive association between our Twitter data and the voter outcome

We are running the regression one by one, each time controlling for one extra variable. Furthermore, we are running the regression separately for each candidate.

$$voteroutcome\hat{}i=\beta 0+\beta 1numretweets+\beta 2numberretweet2+\beta 3income+ui$$

Figure 2 :



```
==========================================================================================================================
                           Model 1 Biden   Model 2 Biden  Model 3 Biden   Model 1 Trump   Model 2 Trump  Model 3 Trump
--------------------------------------------------------------------------------------------------------------------------
# of retweets Biden        19.1286**       109.0589***     98.8062**
                           (8.8502)        (39.9342)       (43.0047)
# of retweets Trump                                                        13.5636         76.5136**       95.8048***
                                                                           (8.4226)        (32.1665)       (35.5000)
retweets_biden2                            -0.0006**       -0.0005**
                                           (0.0002)        (0.0003)
retweets_trump2                                                                            -0.0005**       -0.0006**
                                                                                           (0.0002)        (0.0003)
Per capita income (2020) BEA                               21.0862                                         -30.3281
                                                           (31.5962)                                       (24.2458)
const                      1447144.5401*** 1279715.9629*** 99854.8695      1395449.9243*** 1289771.8105*** 2981939.6414**
                           (267718.3544)   (266749.3084)   (1788182.7980) (203986.4591)   (204497.5922)   (1367988.7087)
R-squared                  0.0870          0.1780          0.1857          0.0513          0.1273          0.1560
R-squared Adj.             0.0684          0.1437          0.1337          0.0315          0.0901          0.1009
R-squared                  0.09            0.18            0.19            0.05            0.13            0.16
NO. observations           51              51              51              50              50              50
==========================================================================================================================
Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01
```

Before analyzing and interpreting the data, one should note that in this data, we only have 50 to 51 Observation, which represents each states. Since most of our previous analysis was at state level, where we were able to gather all the relative data, in terms of income per capita in each state, number of tweets and retweets to predict each state's election outcome. This also means that the following table has regression with low observations, which might lead to some biases in some cases.

In model 1 for Biden, with no retweet count (i.e, the constant in our regression) shows that the average vote for Biden is at 1’447’144 when the number of retweet is at zero. with our Number of retweets coefficient, we can see that a unit increase of retweet has the impact of increasing voter outcome by 19.12 votes. Both our coefficients are statistically significant at 5% confidence interval. However, the coefficient in the number of retweets has relatively low economic significance. We would need a fairly high retweet count for the coefficient to be significant. This is not the case for most rural states. In certain states like New York or Washington DC, however, where the number of retweet is high, the coefficient can have practical significance. Looking in Model 1 Trump, we can see a similar phenomenon happening, where the coefficient for the number of retweets for Trump is even lower and has no statistical significance since the standard errors is fairly high. Looking at the R-Squared, we can see that there is a fairly low correlation between number of retweets and voter outcome.

In Model 2, we can see that our R-Squared coefficient is higher, which means that our hypothesis is true that the regression is non-linear. In other words, there is a non-linear relationship between voter outcome and the number of retweets that the tweet of each candidate, gets. The number of retweets coefficient is now much higher. For example, in Model 2 Biden we can see that an extra retweet is associated with 109 extra votes, which with a large enough number of retweets can be considered as economically significant. Once, again, all of our coefficient in model 2 regressions are also statistically significant, even with low number of observation.

Finally, in model 3 we are also controlling for the income effect. As seen in previous analysis, income is highly correlated with relative vote share, with higher income states voting for Biden and low income states voting for Trump. This is shown in model 3 with our per capita income. For Biden the coefficient is positive and is around 21, which means for a marginal increase in income per capita is associated with 21 extra votes and for Trump the coefficient is negative 30. The coefficient however is not statistically significant. This is because the income across states are fairly similar. The income does relate with voter outcome, however, there is a lot of variation among states. For example, Texas, is a high income state and voted for Trump which creates variability among states even if on average the relationship between income per capita and voter outcome is there.

Conclusion

In summary, we can see that, when looking into the United States data, even though Donald Trump has higher tweet count people interact more with Tweets related to Joe Biden. Looking into it is quite hard to draw a correlation between election outcome and the number of interactions on Tweets for instance likes and retweets. Looking deeper into our data, we notice that these correlations hold at the country level. When looking at our heat plot, we see that at the state level, there is little to no correlation between voter outcome of a candidate and the number of retweets that candidate gets. Our data suggests that throughout the country, with few exceptions, Joe Biden has a higher retweet count while Trump has a higher tweet count. Meanwhile, we cannot rule out the fact that the tweets about Biden can be negative news which is shared more intensively throughout the countries. In other words, Biden might be getting the majority of the traction on Twitter, but it might be negative traction. Moreover, other facts like income also shed a light onto our analysis. We can see that states with higher income are more likely to vote for Biden, and have relatively high interaction with tweets about Biden rather than Trump, and the inverse is also true. We are also assuming that high income states are more urban than rural, which means they are also more likely to use new technology and social media platforms like Twitter.

As for shortcomings that need to be improved, we need to enlarge our data set and include other social media where supporters of other candidates are also present. Looking solely at Twitter data, we notice a bias towards Joe Biden, because he appeals to the younger more urban and politically left leaning demographic, which happens to be the demographic of Twitter. By enriching the data set with other data from other social media, we might get a better, more unbiased image. When analyzing the Twitter analytics data, which includes, the number of likes, retweets and tweets we can only have one version of the facts. To have a complete study we would then need to combine this analysis with sentimental analysis, which will give us a deeper understanding of the content of the tweets, and we can then compare these analytics more carefully, by weighing them, for how positive or negative they are. We can further use satellite data to better understand tweet and voter outcome patterns between rural and urban America.

Reference:

1. DOI: 10.1257/jep.31.2.211

2. doi.org/10.1177/1354068815605304

3. https://doi.org/10.3390/electronics10172082

4. www.kaggle.com/datasets/manchunhui/us-election-2020-tweets

5. www.fec.gov

6. www.census.gov/

7. https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_income

8. https://developers.facebook.com/docs/development/terms-and-policies/automated-data-collection/

9. https://covid.cdc.gov/covid-data-tracker/#cases_casesper100k

10. doi: 10.1007/s10796-021-10222-9

11. https://doi.org/10.1038/s41467-021-25738-6

12. https://www.pewresearch.org/social-trends/2018/05/22/urban-suburban-and-rural-residents-views-on-key-social-and-political-issues/