An Observational Study to COVID-19 and 2020 US Presidential Election: Counties with High Deaths per Case Rate Showed Diminished Support For Trump*

Yiliu Cao

December 14, 2023

This study investigate the causal inference between COVID-19 and Donald Trump's loss during the 2020 US Federal Election using the data from MIT Election Data Science Lab and Johns Hopkins University CSSE. The main methodology used in this paper is Propensity Score Matching with with an exploratory analyzing the optimal treatment initially. The key finding suggest that the counties with a death per case rate exceeding the 0.4 quantile threshold shows a reduced voting preference for Trump. Additionally, this paper conducts a counterfactual analysis based on the treatment effect, indicating that Trump might have be secured to re-elect if the disparities in death per case rates were addressed. Future research should aim to refine the propensity score calculations by incorporating additional variables, such as the winning party in each county during the 2016 election.

Table of contents

*Code and data from this analysis are available at: https://github.com/yiliuc/covid_and_trump_loss.git

1 Introduction

At the beginning of 2020, US reported the first COVID-19 cases. This number grows to ten thousand within two months on March 19 and increase dramatically to 100K on March 27 [citation]. At the same year, the U.S. Federal Presidential Election started off and Trump lost to Biden. As of Election Day, there was above 9 million cases and about 0.2 million deaths in US. When people talk about Trump's lost, people always connect it to the pandemic, saying that he "mishandled his greatest test" [citation] and 2016 Election was a historical accident [citation]. However, there is another voice that COVID is not the only or the decisive factor attributing to his lost, his limits as a politican and wasted his advantage on the economy are also important explaining why he lost. This paper will investigate whether the is a causal effect between COVID and Trump's lost. If there is, whether it is strong enough to beath Trump at 2020.

There are existing researches on this. Leaonardo et al. [citation] suggested that Trump could win if there was no COVID but it will be too naive to simply connect his lost to the COVID-19 infection rate. Their research highlighted that the COVID-19 had the most significant negative impacts on those urban areas where there was no stay-at-home orders, especially for those "swing" states. They also argue that the race diversity and education attainment are also important when analysing voting patterns for Trump. In addition to that, Marcus et al. [citation] suggested that the deaths per case is a more crucial indicator than infection rate when analysing the voting for Trump. However, they suggest the voting itself may not capture the impact of COVID as we do not know when did the voters make their decision, especially for early vote. Compared to these researches, Harold [citation] conduct read survey before and after the election to illustrate the impact of COVID pattern, they found that even though COVID has impacts on voting, it is not the dominant factor. Without COVID,

Trump was also less likely to re-elect due to US highly polarized political landscape. Similarly, Shang et al. [citation] also conduct a survey and all the participants show natural or negative altitude. They suggest some social activity like "Black lives matter" may have impacts on voting. Furthermore, the

While all the above researches had investigated the correlation between COVID-19 and votings for Trump. Except conducting real surveys, the methods they used were to conduct a model and find the relevant variables in predicting the (change) vote for Trump. Then conduct counterfactual analysis to predict what will be the voting for Trump if there was no COVID or a certain amount reduce in COVID deaths. Instead using a similar way like those studies, this paper will introduce the method of Propensity Score Matching to find the treatment effect, to see whether those more impacted counties will have different voting patterns for Trump. However, since all counties experienced COVID-19, this paper will firstly find the optimal treatment groups. After that, I will also conduct a counterfactual analysis to see if Trump could re-elect.

There will be five parts in this paper. I will introduce the data used in this paper and present data suammries and visualization in Data part. After that, I will introduce the methods in this paper and the corresponding results in Methods and Results part, respectively. Then the results will be interpreted and discussed in the Discussion part and conclude with limitations and drawbacks in [Conclusion] part..

2 Data

2.1 Data sources

The data used in this paper comprised five data sets from three different sources corresponding to different topics. The primary source is MIT Election Data Science Club which build open online data collections of the US Federal or Senate Election results, spanning from nation to county levels. The data extracted is called "County Presidential Election Returns 2000-2020" with about 70,000 rows containing the voting patterns for each candidate and party by county since 2000. Besides that, the data also indicate the types of voting, such as "EARLY VOTE" and "ELECTION DAY," for the same party in the a county. To analyze the voting patterns for Donald J. Trump, I only select the data for 2016 and 2020 US Federal Election for all counties and parties. Thus the filterd data set contains the number of votes through various ways for each county and party.

In addition to that, the study also uses the data regarding COVID-19 from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. CSSE collects and reports the local, national, and global multidimensional data, including medicine, health care, disaster response, etc. During the pandemic, they collected the U.S. and international COVID cases and deaths and reported them on their GitHub, summarized by daily reports from April 12, 2020, to March 9, 2023. To access the impact of COVID on 2020 Election, this paper used

the daily report on November 3, 2020 which is the election day of 2020 US Federal Election. The resulting data include the aggregates number of cases, deaths and recovers, as well as the incidence rate and case fatility ratio for each county by the election day.

Lastly, this paper also employ the socio-economic data from American Community Survey (ACS). ACS is a online open source database conducted by the U.S. Census Bureau, containing the various soci-economic factors in either geographic levels. The data sets extracted from ACS is 2020 five-year estimates of DP02, DP03 and DP05 which covers the social, economic and demographic characteristics in county level. Since there are thousands of variables, I will only use the variables which are found from my previous paper that are significant in predicting the COVID-19 mortality rate such as the mean household income, high-education attainment etc. The descriptions of all variables can be found on Table 1.

After having five data sets from three sources, they are merged into one big data by counties. Using the merged data, there are some new variables calculated by the exiting variables. In terms of election, I calculate the percentage of vote for each party (candidate) in both 2016 and 2020 Election, and calculate the corresponding change of percentage votes. I also create new variables indicating the winning party in both elections for each county. Additionally, to accurately compare the COVID cases and deaths across all counties, I transform the number of cases and deaths to infection and mortality rate by 10,000 citizens in each county. The final data now consists of 3107 rows with 36 columns. All the important variables are described in Table 1.

Table 1: Descriptions of all important variables in the analyze data

Variables	Coded name	Descriptions
Income Percentile	income_pctile	The income percentile of each county
High-Education Attainment	prop_high_education	The proportion of residents in a county having a at least bachelor degree
Private Insurance	private_insurance	The proportion of local residences having private insurance
No Insurance	no_insurance	The proportion of local residences without any health insurance
White Population	white_pct	The proportion of White population
Black Population	black_pct	The proportion of Black population
Males Population	males	The proportion of Males population
Infection Rate	infrate	The COVID infection rate, calculated by number of case per 10,000 residences
Mortality Rate	mortrate	The COVID mortality rate, calculated by number of death per 10,000 residences
Death per Case	dpc	The COVID death per 10,000 confirmed cases
Votes for Democrat (2016)	pct_vote_demo16	The percentage of votes for the Democrat in 2016

Variables	Coded name	Descriptions
Votes for Democrat (2020)	pct_vote_demo	The percentage of votes for the Democrat in 2020
Votes for Republican (2016)	pct_vote_rep16	The percentage of votes for the Republican in 2016
Votes for Republican (2020)	pct_vote_rep	The percentage of votes for the Republican in 2020
Change of Vote for Republican	change_vote_rep	The change of vote for the Republican from 2016 to 2020

2.2 Data summaries and visualizations

Table 2: The summary of COVID cases and deaths as of Election Day.

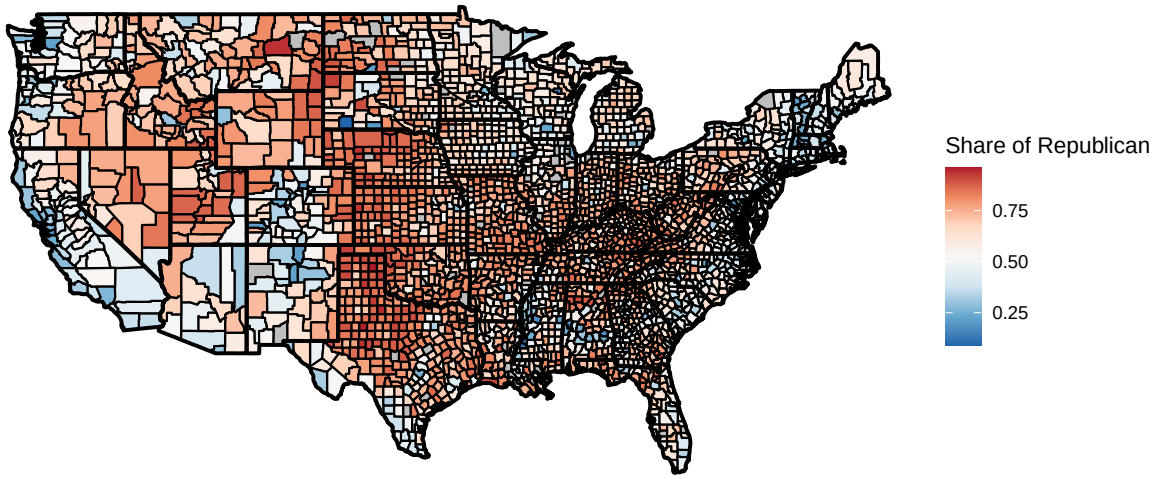
Winning Party	Count	Cases	Infection Rate	Deaths	Mortality Rate	DPC	Income
Democrat	539	10437.514	2989.381	292.82931	77.394	2528.575	83540.12
Republican	2576	1427.246	2970.135	28.21584	55.643	1887.507	69800.52

Table 2 summarize the COVID-19 impacts and income levels between the counties voting for Democrat and Republican. Even though the Republican won about five sixths counties, there is a stark contrast where counties supporting Biden had nearly ten times the average number of COVID-19 cases and deaths compared to those supporting Trump. Despite a similar infection and mortality rates between the two groups, there is still a notable higher death-to-case rate in counties voting for the Republican. Furthermore, these counties also has a lower mean household income than the "Democrat-win" counties. These patterns suggest a potential correlation between the extent of COVID-19 impact and voting behavior in the 2020, where areas more severely affected by COVID-19 were less likely to vote for the Trump, vice-versa. This aligns with mainstream media and intuitive expectations.

We can validate our guess from Figure 1 which compares the relative ratio of votes for the two parties and the death to cases rate in maps. From the maps, it is clearly more counties had preference for the Republican but the counties with higher death per case rate have a smaller ratio of vote for the Republican and these counties are usually richer ones. The most significant example can be California and New York. In contrast, the states with less impacted by COVID, such as Utah, vote more for Republican.

Given the insights from Table 2 and Figure 1, Figure 2 further examine the impact of COVID and income levels on the voting behaviors for the two parties. Figure 2 shows the correlation between the votes of the two parties to the death per case rate and income levels. The two graphs on the left indicates the share of votes whereas the right two graphs indicate the number

The relative share of votes between Republican and Democrat party



The distribution of death per case rate by 100K

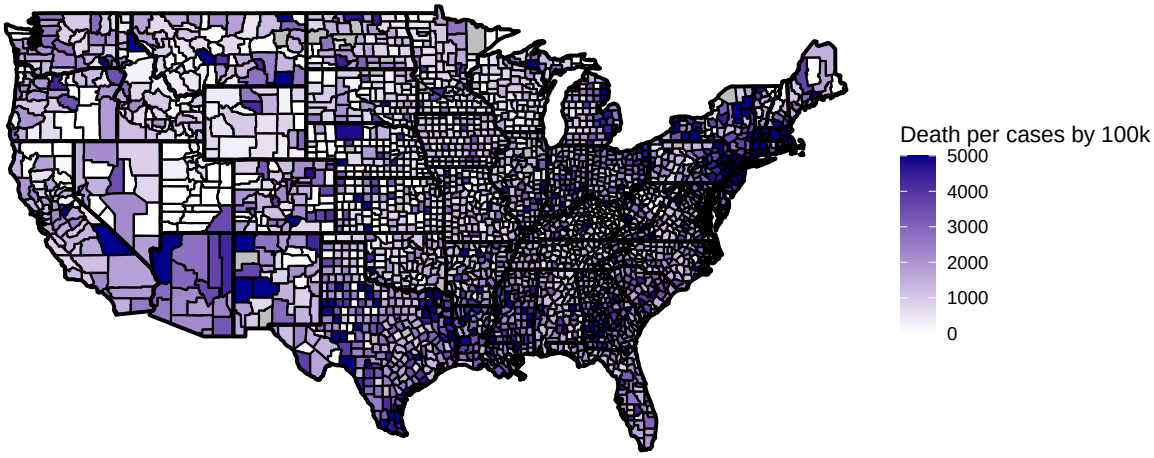


Figure 1: The ratio of votes for the Republican and the infection rate per 100k in each county

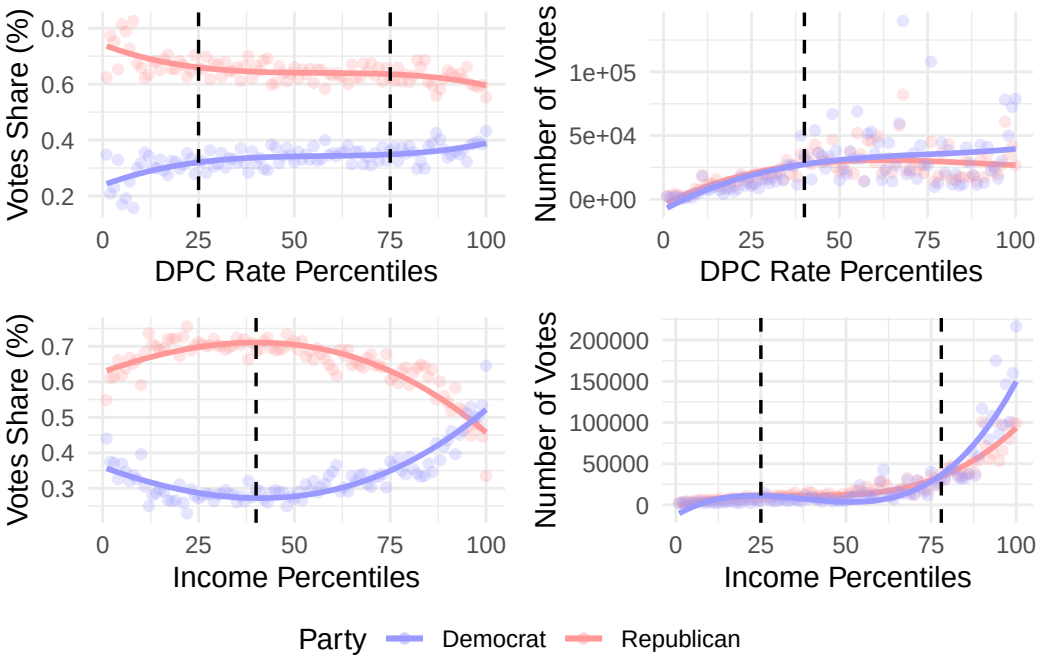
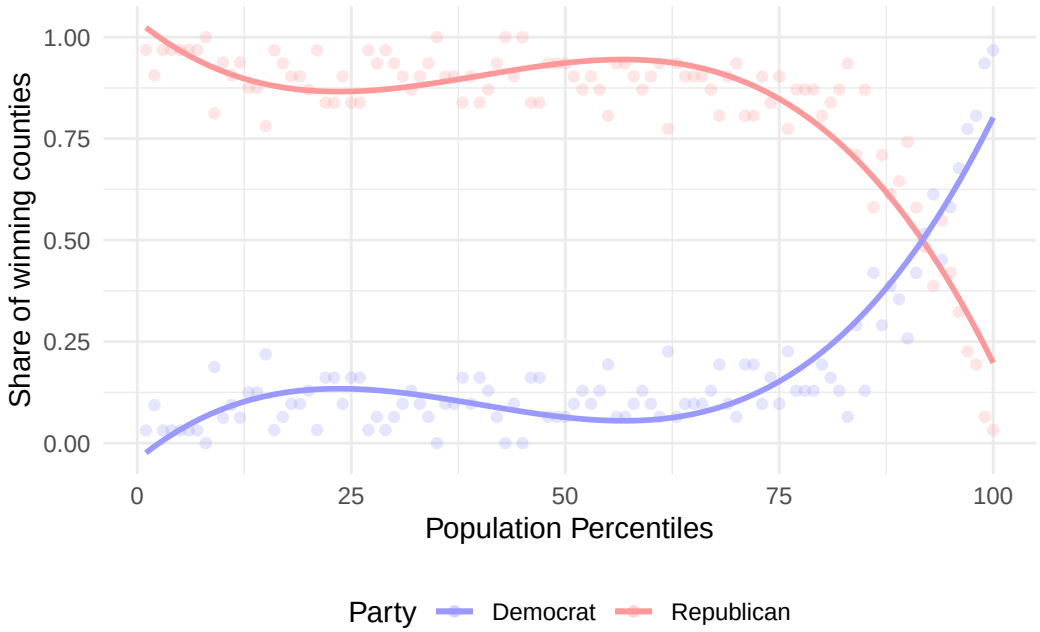


Figure 2: The correlation between voting behaviors to income and DPC rate for the Republican and Democrat

of votes. The two parties are represented by different colors. In terms of share of votes, in general, there is a positive relationship between the DPC and votes for the Democrat but a negative one for the Republican. However, the counties with medium value of DPC is less sensitive to the voting for the two parties, compared to the "tail" counties which their support change significantly as DPC increases or decreases. Meanwhile, there is a quadratic patterns between the vote for the two parties and income, it seems that the Republican is in favor of low income counties, but as income increase, more counties vote for Bide not Trump.

In addition, even though the average share of votes for Republican is always higher than 0.5, it does not necessarily mean that the Democrat lost in all counties. The reason behind this is, as mentioned in Table 2, the counties which the Republican won is about five times as higher as the Democrat. Since Republican has more counties, it definitely has a higher average share of votes than Democrat. This pattern also indicate that Democrat usually won the densely populated, urban areas but Republican won in more sparsely populated, rural areas [citation]. The high population density areas has more votes and explains why Biden beat Trump. We can verify this from the number of votes for the two parties, where the number of votes for Biden exceeds Trump as DPC and income level increases. This pattern is especially significant for the top 50% richest counties and the counties with DPC larger than 75 quantile.



From four graphs in Figure 2, we can conclude that, in general, the higher the DPC rate, the lower the support for Trump but higher for Biden. Besides, the Democrat is in favor of richer voters, where as Republican is in favor of low income counties. It is direct to ask whether these two variables have interaction effect on the voting patterns? However, it will be difficult to visualize the interaction effect between two continuous factors. Instead, I will manually set a new categorical variable `high_income` to denote whether the county has a high income with certain cutoff. Then the difference of correlations of the two groups as death per case rate changes. I will set different cutoffs to fully test the interaction effect.

Figure 3 compares the distribution of county income levels for the two parties. In general, counties voting for the Democrats have higher income levels than the Republicans. In addition, we can observe that the counties are intensively concentrated around income levels \$50k to \$75k, compared to the Democrats, where the counties are approximately uniformly distributed at each income level. Furthermore, barely any county has at least 150k voting for the Republican party. Both Table 2 and Figure 3 indicates that the Democrat is in favour of wealthy counties but poorer counties for Republican. Consequently, the more affluent counties (states) usually have more electoral votes than the poorer ones; this may provide insights into why Trump lost.

3 Methods

The objective in this paper is to conduct the causal inference between COVID-19 and Trump's loss during 2020, the main methods implemented in this study will be the Propensity Score

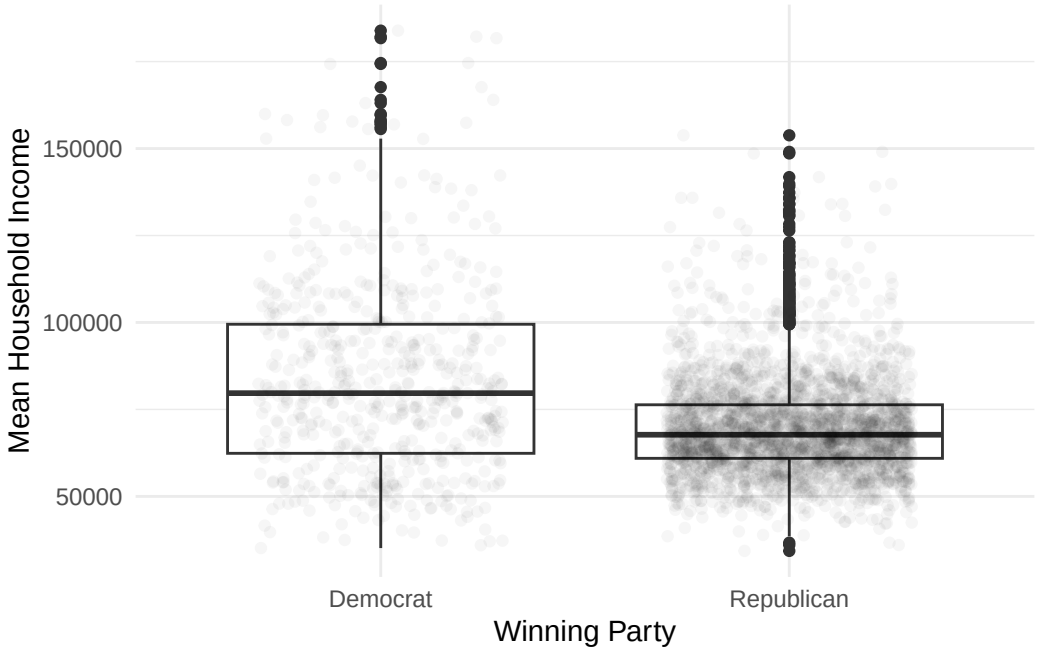


Figure 3: Summary of income levels for counties voting for Democratic and Republican

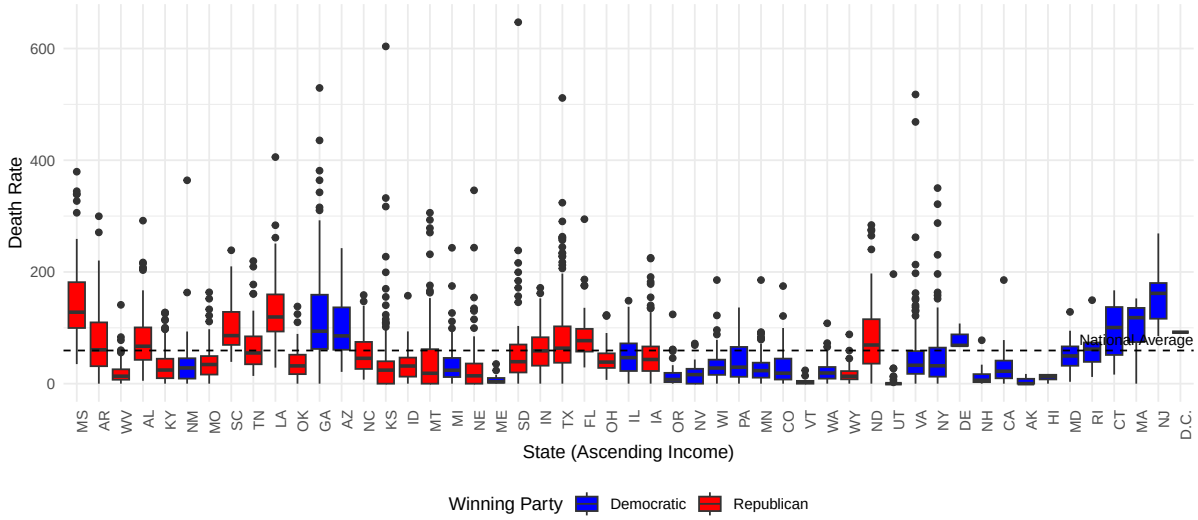


Figure 4: Variation in COVID-19 Mortality Rates Across States, Categorized by Income Levels and Dominant Political Preferences

Matching (PSM). After find the causal inference, this paper will also conduct a counterfactual analysis to see whether Trump can re-elect if there was no COVID.

3.1 Treatment

The treatment refers to the intervention or exposure that is being studied to understand its potential impact on the outcome. In other words, we implement certain interventions to a group of people but not for the rest. The group of people receiving the treatment is called the treatment group and control group for the rest. For instance, Jochen [citation] examined whether a new medical treatment called clampless off-pump coronary artery bypass (clampless OPCAB) will reduce the in-hospital mortality rate, compared to the traditional way. In their studies, the treatment is the new medical surgery and patients who took this surgery is the treatment group. They compared the mortality rates between the two groups and concluded that the new treatment can significantly lower the stroke and mortality rate.

Normally, the treatment should be established before conduct the analysis, this aims to reduce the risk of p-hacking. However, the objective of this paper is to find whether Trump lost the last Election due to COVID-19, it may not be possible to state the treatment in advance as all the counties in US had experienced COVI-19. Therefore this paper will firstly perform exploratory analysis to firstly find the treatment group which has the most significant difference in voting for Trump. This paper will set a cutoff defining whether a county has a high or low death per case rate. In addition, from Figure 2, low and high income counties also seem to have different voting behaviors. Therefore the treatment in this paper will incorporate whether a county has a high DPC and a high income levels with certain cutoffs.

However, we need to be very careful about the cutoffs for defining "high" in DPC and income is crucial as the risk of p-hacking or data dreging may raise. P-hacking means the manipulation of data analysis until we can find the statistically significant results [citation]. If we simply choose one cutoff and find the significant results that match the common sense, it may raise the p-hacking risk and the analyses will be less convincing. Therefore, to make the analysis more comprehensive, instead of choosing a single cutoff, I will set a grid of cutoffs for each variable to see how the separate or combine treatment effects varies. For example, one treatment groups can be counties with death per case and income levels higher than 0.3 quantiles. By this way, we should be able to find the optimal treatment group and avoid the risk of p-hacking.

3.2 Propensity Score Matching (PSM)

To estimate the treatment effects, randomized controlled trials (RCT) is the most ideal and the golden way to do this [citation]. The reason is that RCTs is the only way that guarantees the equal distribution of known and unknown parameters. Therefore the outcomes between the treatment and control groups will not be confounded. However, conducting a RCT is nearly impractical and impossible. Back to the coronary artery bypass example [citation], you

can not force a patient to take new or traditional treatment, the researchers can not control it and hence RCT is impossible. Instead, we can take observational studies to find the treatment effects.

One method that can be applied to observational studies to find the treatment effect is Propensity Score Matching (PSM). Propensity score is the probability of an observation that receive the treatment based on the existing covariates using logistic regression. Using propensity scores have one main advantages which is dimension reduction [citation]. Dimension reduction means we can describe an observation by only its propensity score instead of a list of covariates. Therefore, we can match two observations having similar propensity scores but from treatment and control groups separately. With this algorithm, we can have many matched pairs and the matching process is called propensity score matching. Here, we have another advantage of PSM which is design separation [citation], meaning that it can separate the covariates balancing and effects estimation. I will show the balance of covariates later. Finally, we can directly estimate their outcomes and solve the drawback that we can not have causal inference from observational studies. That is why it can also separate designs and this process is called propensity score matching.

In this paper, the propensity scores are used to estimate how likely a county having a high DPC or high income levels (or both) using multiple linear logistic regression. The model will be look like:

$$\log\left(\frac{P(T=1|X)}{1-P(T=1|X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

where:
* $P(T=1|X)$ is the probability of a county receiving treatment
* X_j $(1 \leq j \leq p)$ are the covariates
* β_0 is the intercept * β_j $(1 \leq j \leq p)$ is the corresponding coefficients of each covariate

Using the model, I will calculate the propensity scores for each county and match the data with similar propensity scores. Notice that the values of propensity scores changes as the treatment changes. That said, I will calculate the propensity scores and match the data for number of treatment times.

3.3 Counterfactual Analysis

Using the PSM described previously, we can find the treatment effects under different settings of treatment groups. However, to conclude whether Trump was lost due to COVID, we need to perform a counterfactual analysis with re-calculating the votes for Trump of counties in the treatment group.

The couterfactual analysis is particularly necessary for those "swing" states in 2020. In 2016 Federal Election, one key factory why Trump defeated Hillary is that he won seven out of eleven swing states. However, he only won three of them in 2020 [citaion (wikipedia)]. Besides, if Donald Trump were able to hold onto three swing states which are Georgia, Arizona and Wisconsin where the average margin of Democrat is only about 0.3%, the result would have been a 269-269 electoral tie decided in the House of Representatives [citation (wikipedia)]. Then the presidential election is left up to members of the House of Representatives and it is possible that Trump can win.

To find the voting patterns for Trump, especially for the "swing" states, I will follow the official election procedure and use the winners-takes-all rule. That said, I will re-calculate the votes for Trump in county level and summarize by county level. Then the party with higher total votes will take all the electoral votes in this state. Eventually, I will calculate the total electoral votes for each party to see whether Trump could re-elect if the treatment effects diminish. Besides, I will also predict the votes if the death number can reduce 5%, 10%..., to see if a tiny decrease in deaths will make Trump to re-elect.

4 Results

4.1 Choise of treatment groups

Figure 5 shows the correlation between votes for Trump and income levels, segmented into "High-DPC" and "Low-DPC" county groups at varying DPC cutoffs. For instance, a cutoff of 0.3 classifies counties with a DPC rate below this threshold as "Low DPC," and those above as "High DPC.". By comparing these groups across different cutoffs, we aim to identify the cutoff where the difference in voting for Trump across difference income levels is most pronounced. With that said, we can find that except cutoff 0.4, the two line converge as income increases (right tail). Only for the 0.4 cutoff, the lines of two groups converge at the middle-income but do not overlap and diverge at the tails. Additionally, regardless of the cutoff applied, there is a common pattern that the support for Trump increases in poorer counties up to around the eighth bin, which is approximately the 0.4 income quantile, and then decreases in wealthier counties. This indicates a significant voting disparity between counties below and above the 0.4 income quantile.

Furthering on our analysis, Figure 6 shows the correlation between Trump's vote share and the DPC rate but this time dividing counties based on income levels into "High Income" and "Low Income" groups. Each subplot in this analysis corresponds to a different income cutoff. Compared to Figure 5, all curves, regardless of the cutoff, display a linear or near-linear relationship with voting patterns. As the income cutoff increases, the difference in voting behavior between the high and low-income counties becomes more distinct, especially at the 0.9 cutoff. This suggests that in counties with higher incomes, voting patterns vary significantly from those in lower-income counties as the DPC rate changes. Therefore, the

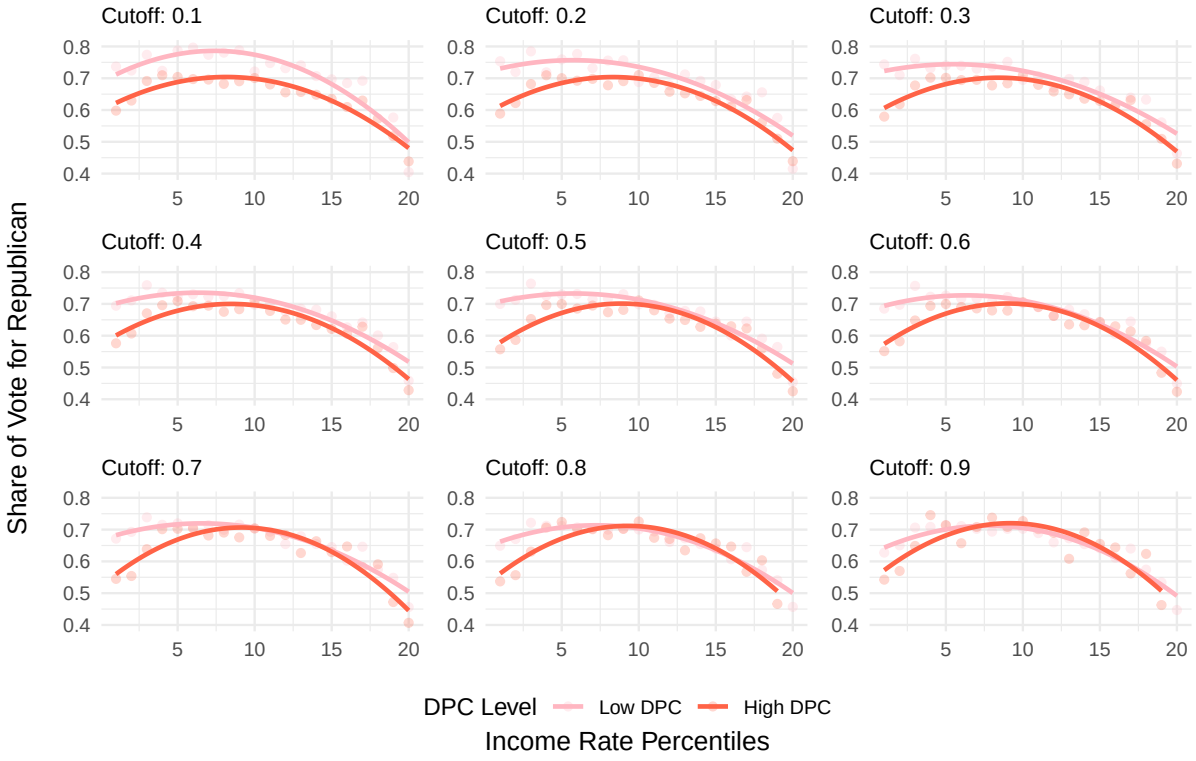


Figure 5: The Correlation Between Voting for Trump and Income Levels, Categorized by DPC Rate at Varied Quantile Cutoffs (e.g., Top 40% as High-DPC and bottom 60% as Low-DPC for "Cutoff: 0.6")

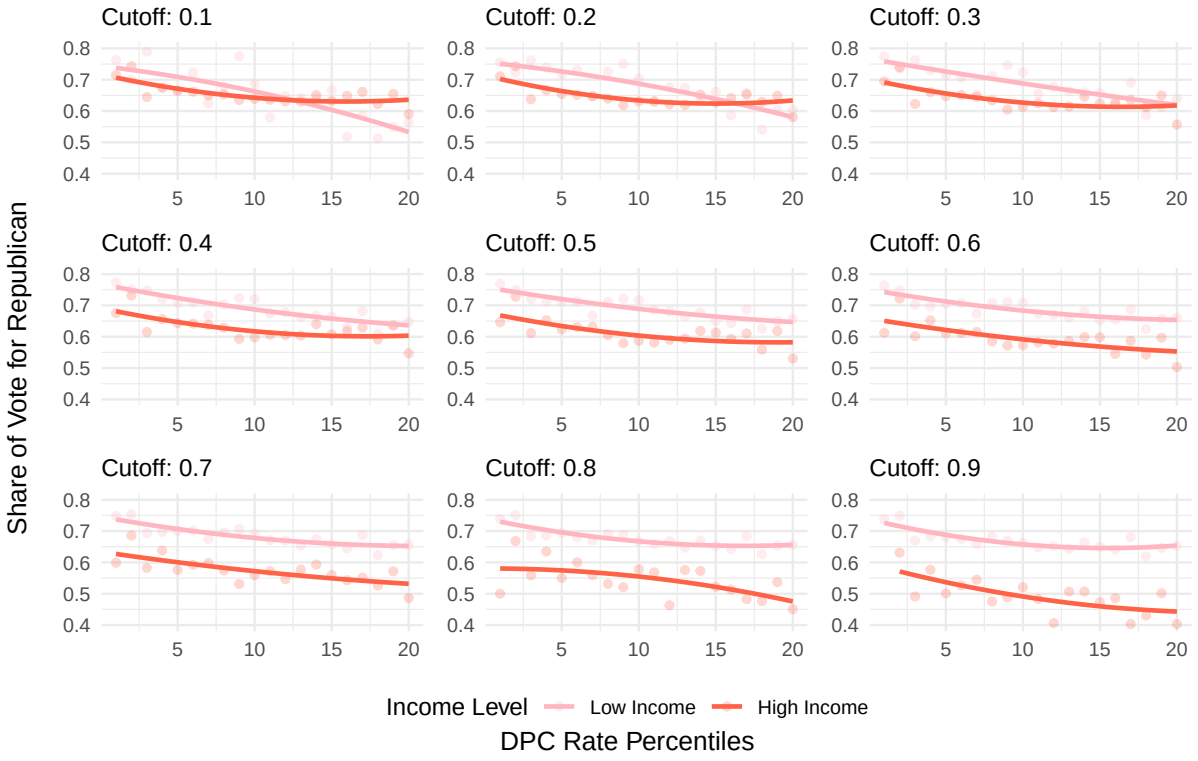


Figure 6: The Correlation Between Voting for Trump and DPC Rate, Categorized by Income Levels at Varied Quantile Cutoffs (e.g., Top 40% as High Income and bottom 60% as Low Income for "Cutoff: 0.6")

most significant pattern is observed at the 0.9 income cutoff, contradicted to the patterns identified in Figure 5.

So, should we include all the income levels and DPC in the settings of treatment? The answer is NO and we should only include the DPC. From Figure 5, a consistent quadratic relationship between income and voting patterns for Trump is evident, regardless of how the cutoff for defining high DPC is adjusted. This pattern demonstrates that poorer counties (in the first half of the income spectrum) may show a positive or a positive correlation with voting for Trump depending on the cutoff we choose, whereas richer counties will only exhibit a negative correlation. This quadratic pattern can be verifies from Figure 2, which illustrates a similar relationship between votes and income. Conversely, Figure 6 presents a different narrative. Here, all regression lines, irrespective of the DPC cutoff, display a linear or nearly linear relationship. Besides, unlike the Figure 5, all the lines in Figure 6 shows a negative slope, meaning that the choice of cutoff for "High DPC" group will not change the general voting patterns for Trump. The optimal cutoff we found previously, 0.4, is the one that the "High DPC" and "Low DPC" have the maximum difference. This finding can be also verified from Figure 2, which shows a monotonically decreasing cubic correlation between votes and DPC. The Figure 7 shows a simple visualization of why is the case.

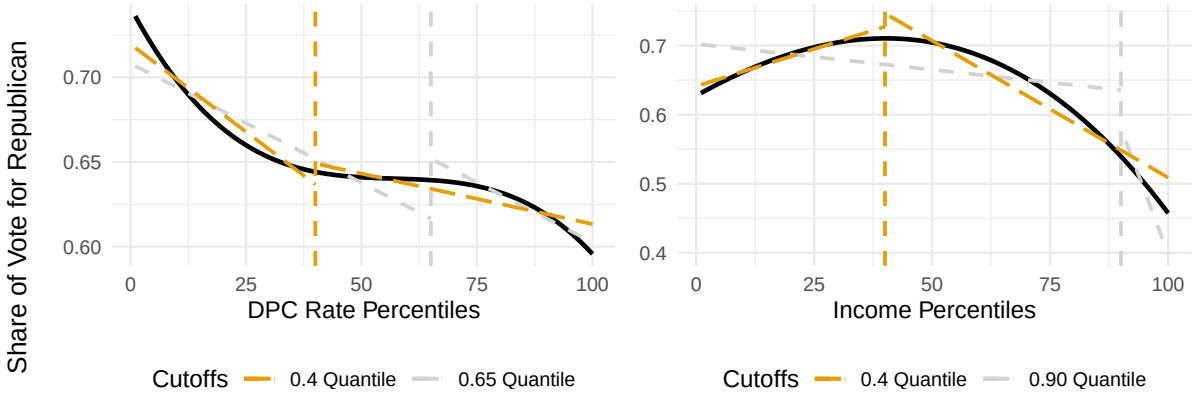


Figure 7: **?(caption)**

Figure 7 shows a simple example when we test different cutoffs for DPC rate and income. From the graph, we can see that no matter which cutoff we set, the low DPC and high DPC counties always have the a negative value of slope. The only difference is the absolute values of the slope. However, for the income, we can see that for the cutoff 0.4 quantile, the low income and high income groups show a opposite patterns for voting for Trump. But when we increase the cutoff to 0.9 quantile. The two groups are both negatively related to the change. This indicates that it is not appropriate to incorporate income in treatment as the choice of cutoff will significant change the

no matter which cutoff we set, the two groups will always have opposite altitude for voting for Trump. Suppose we have enough samples, when we include income into the treatment, then treatment effect will always be more negative when we increase the cutoff. The treatment

Table 3: The summary of Propensity Score Matching

		Original		Matched	
Treatment Effect	P value	Obs.	Treat. Obs.	Treat. Obs	Treat. Obs. Unw.
-0.018342	0.021404	3115	1869	1869	5415

effect will be more and more significant when we increase the cutoff. However, the number of counties in treatment group will decrease and hence it will be unnecessary to have income into treatment settings.

From Figure 5, we already verify that cutoff 0.4 is the optimal threshold for defining high and low DPC group, therefore the treatment group will be the counties having DPC rate larger than 0.4 quantile. Furthermore, to balance the effect of income, I will incorporate income into the ways to calculate the propensity scores for each county.

4.2 Propensity Score Matching

In my previous paper, I already show that proportion of residence with at least a bachelor degree, income, proportion of people with private insurance and withou health insurance, ratio of males, ratio of black and black residence are statistically in predicting the probability of a county has a high mortality rate or not. In this paper, I will directly use the results and only consider these variables when predicting the propensity scores.

$$\begin{aligned}\log\left(\frac{P(T=1|X)}{1-P(T=1|X)}\right) = 8.06 \\ &- 0.02 \times \text{prop\_higher\_education} \\ &+ 0.01 \times \text{pctile} \\ &+ 0.01 \times \text{no\_insurance} \\ &- 0.03 \times \text{private\_insurance} \\ &- 0.13 \times \text{males} \\ &+ 0.00 \times \text{white\_pct} \\ &+ 0.06 \times \text{black\_pct}\end{aligned}$$

The above equation shows the logistic regression model to predict the propensity score for each county. It seems that males is the most critical factor when predict the propensity score. Perhaps the majority of COVID deaths are males. Interestingly, the counties with a higher income levels will have a higher death per case rate. This may explain why richer counties are less likely to vote for Trump.

Table 4: The balance of each covariates before and after the mathcing

	Before Matching		After Matching	
Variables	Mean Contr	Std Mean Diff	Mean Contr.	Std Mean Diff.
prop_higher_education	23.2460	-10.772	21.6330	5.6159
pctile	51.7740	-10.389	47.8040	2.9221
no_insurance	8.7107	24.449	9.7494	4.6766
private_insurance	67.6560	-35.913	63.4090	5.9550
males	50.5120	-32.492	49.5520	9.6593
white_pct	87.1370	-47.207	79.6360	-5.5169
black_pct	4.1685	49.083	11.4560	5.4646

Table 3 summarize the results of Propensity Score Matching. The treatment effect is about -0.18 means that the counties with DPC rate higher than 0.4 will in average vote 0.018 less for Trump, compared to the counties with lower DPC rate. This value is statistically significant as its p value is about 0.02 which is lower than 0.05. In addition, there are 1869 treatment observations in both original and matched data, meaning that there some counties in control group that are matched more than once.

Table 4 summarize each variable's mean values in the control group and standard mean difference before and after the matching. We can see that the mean values in the control group approximately remain same after the matching. Besides, there is a significant decrease in the standard mean difference for each variable. This indicates that the Propensity Score Matching reduced the imbalance between the treatment and control group considerablely.

4.3 Counterfactual Analysis

Table 5 shows the summary of vote for the eleven swing states in 2020 if the difference in voting for Trump between the high DPC and low DPC eliminates. This equivalent to on average In 2020, the truth is that the Trump only won three of eleven swing states and he only won 232 electoral votes. However, with counterfactual votes, Trump will win five more states and hence bring him 52 more electoral votes. Therefore the total electoral votes for Trump will be 284. Since the cutoff to win is 270 votes, this indicate that Trump could re-elect if he can eliminate the disparities in DPC rate.

Table 5: The summary of counterfactual votes for the eleven swing staes in 2020

		Actual Results			Simulated Results*		
States	Electoral Votes	Trump	Biden	Margin	Trump*	Biden*	Margin*
NH	4	365660	424937	7.5% D	378054.6	412542.4	4.36% D
MN	10	1484065	1717077	7.28% D	1514486.2	1686655.8	5.38% D
MI	15	2649852	2804040	2.83% D	2733394.7	2720497.3	0.24% R
NV	6	669608	703314	2.46% D	693512.9	679409.1	1.03% R
PA	19	3377674	3458229	1.18% D	3497988.6	3337914.4	2.34% R
WI	10	1610065	1630673	0.64% D	1613541.4	1627196.6	0.42% D
AZ	11	1661686	1672143	0.31% D	1723779.1	1610049.9	3.41% R
GA	16	2461837	2474507	0.26% D	2547370.9	2388973.1	3.21% R
NC	16	2758773	2684292	1.37% R	2810662.1	2632402.9	3.27% R
FL	30	5668731	5297045	3.39% R	5847431.0	5118345.0	6.65% R
IA	6	902009	759061	8.61% R	917420.3	743649.7	10.46% R

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

6 References