# Accessing the COVID-19 Death Rate in the US: While It Varies Across Different Income Levels, the Impact of Political Preference is Unignorable*

Yiliu Cao

November 3, 2023

This study utilizes the data on socio-economic factors, COVID-19 statistics, and election results sourced from the American Community Survey, CSSE at JHU and Fox News. Primary findings indicate that the death rate positively relates to voting for the Republican party but has an inverse relation to income levels. The research further identifies a trend of Republican party preference among low-income counties, and the counties voting for the Democratic exhibit lower mortality rates. Furthermore, the potential reasons contributing to the inequality in mortality rate might be the high education and insurance system. Future research should focus on more correlated variables, such as hospital capacity.

## Table of contents



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

# 1 Introduction

On May 23rd, the WHO's Director-General, Dr. Tedros Adhanom Ghebreyesus, announced the end of the COVID-19 pandemic (United Nations 2023), declaring that the virus was no longer considered a global threat. This marked the official end of the pandemic that spanned three years. Over this period, nearly 7.6 billion people got infected, and approximately 6.9 million succumbed to the disease. The US seems to be one of the countries that have been influenced by COVID-19 the most, with about 1.1 million dead and 104 million infections (Al Jazeera 2023). The mortality rate of COVID-19 in the US is about 3.41 per 1,000, which is significantly higher than in other western developed countries. Interestingly, the pandemic began around the time of the 2020 Federal Election and unfolded entirely during President Biden's term. As another Federal Election looms next year, it is worth knowing how the COVID deaths are related to socio-economic indicators. There's also speculation about the role of political preferences. Analyzing these connections could provide insights to better prepare for and prevent future health crises, especially as we navigate the post-pandemic period.

To understand and explore the COVID death rate, I will use five data sets from three different sources. American Community Survey (Bureau 2023b) collects the socio-economic information for each county, such as the mean household income. The Center for Systems Science and

Engineering (CSSE) at John Hopkins University (CSSE 2023) presents data on COVID-19 cases and deaths by county. Regarding electoral data, Fox News have comprehensive records of the 2020 US Federal Election, providing the vote counts for each party across all counties. These data can help us to assess the inequality of mortality caused by COVID across different countries.

My analysis will be focused on the variations of death rates across different counties to see how the counties with different socio-economic factors will affect the COVID-19 deaths. Moreover, I will incorporate a political dimension, especially the voting rate in support of the Republican party in each county, to see whether a higher proportion of Republican voters in a county correlates with the COVID-19 death rate. Through this paper, I aim to construct a model where the COVID-19 death rate is the dependent variable and includes only those predictors that are statistically significant. This study can help US citizens understand the factors affecting the death rate and inform their voting decisions in 2024.

There will be five main parts in this paper. In the Data part, I will introduce the data used in this paper and highlight the key variables. The Methods and Results parts will introduce the methods used in this paper and the corresponding results. Moreover, the results will be interpreted and provide insights about the COVID-19 death rate in the Discussion part. Lastly, I will conclude the entire paper and discuss the limitations and drawbacks.

# 2 Data

As introduced in the introduction, there are three primary sources of data in the data: the American Community Survey (ACS) (Bureau 2023b), The Center for Systems Science and Engineering (CSSE) at John Hopkins University (CSSE 2023), and Fox News (Network n.d.). Each source provides distinct data dimensions that contribute to predicting the COVID-19 death rate in each US county.

## 2.1 American Community Survey (ACS)

The data from ACS contains the socio-economic information for each county. While there are many tables from ACS, the three tables I will use in this paper are DP02, DP03, and DP05. All the tables are the 2021 five-year ACS estimates.

DP02 contains data regarding social characteristics. In this paper, I will extract information related to educational attainment, specifically focusing on the percentage of residents in each county who have attained at least a bachelor's degree. In contrast, DP03 contains information about the economic characteristics. I will take the information about the proportion of residents with private insurance, the mean household income, and the unemployment rate in each county. Lastly, DP05 contains the demographic data and housing estimates. In this paper, I will take the total population, the proportions of children and individuals above 85,

Table 1: The summary of important variabls from ACS

| | Unique (#) | Missing (%) | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|---|
| no_insurance | 252 | 0 | 9.6 | 5.1 | 0.0 | 8.5 | 44.9 |
| age_85 | 73 | 0 | 2.3 | 1.0 | 0.0 | 2.1 | 12.0 |
| high_education | 449 | 0 | 22.7 | 9.6 | 0.0 | 20.3 | 76.3 |

Table 2: The summary of infection and death rate of COVID in US

| | Unique (#) | Missing (%) | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|---|
| infection_rate | 2867 | 0 | 306.9 | 107.6 | 48.7 | 302.4 | 4855.4 |
| death_rate | 2847 | 0 | 4.3 | 1.7 | 0.0 | 4.2 | 13.7 |

and the percentages of white and black residents in each county. The Table 1 summarizes the important variables.

### 2.2 Center for Systems Science and Engineering at JHU

The CSSE at JHU has been aggregating global COVID-19 data from 2020 to 2023 for each country in the world. All the raw data is readily accessible on their CSSE GitHub repository. Notably, they have provided daily COVID-19 reports for each US county since 2020. However, they stopped updating the data after March 9th, 2023, as there were no significant changes. The COVID data for the US used in the paper was collected on March 9th.

The data on March 9th contains the number of confirmed cases and deaths for each county in the US, as well as the incident rate. For the subsequent data analyses, I will only use the COVID cases and deaths in each county. The Table 2 summarized them.

#### 2.2.1 Fox News

Fox News has posted the 2020 Election data, which was collected by Tony McGovern (McGovern n.d.). It contains the county-level election results during the 2020 Federal Election. For each county, it includes the number of votes for the demographic and republican party, as well as the total votes.

### 2.3 Data Cleaning

All data was collected and cleaned using the statistical programming language `R` (R Core Team 2022). The packages used are `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et

Table 3: Summary of important variables in the model

| Variable | Encoded Name | Description |
|---|---|---|
| High education attainment | high_education | The proportion of local residents with at least Bachelor degree |
| Income Percentile | IncomePctile | The mean household income percentile representing the income level for each county |
| People with no insurance | no_insurance | The proportion of local residents without insurance |
| People with private insurance | private_insurance | The proportion of local residents with private insurance |
| Males | males | The proportion of male residents |
| People aged above 85 | age_85 | The proportion of people aged above 85 |
| White people | white | The proportion of white people |
| Black people | black | The proportion of black people |
| People supporting Republican | rep_rate | Number of voters supporting Republican by 1000 |

al. 2023), `stringr` (Wickham 2022), and `janitor` (Firke 2023). Besides, `here` (Müller 2020), `modelsummary` (Arel-Bundock 2022), `girdExtra` (Auguie 2017), `usmap` (Di Lorenzo 2023), `ggpot2` (Wickham 2016), `knitr` (Xie 2023), `kableExtra` (Zhu 2021) are used to analyze the data.

The way to clean each data set is different. For all the ACS data, I downloaded them using ACS API ("Available APIs" n.d.) and only extracted the necessary variables using the codebook published by ACS (Bureau 2023a). Additionally, I standardized the county names and transformed state names into their abbreviated forms. Regarding the COVID data sourced from JHU, I only select the cases and death numbers of US counties. Lastly, for the election data, I not only extracted the county name and transformed the state name to its abbreviation form but also calculated the number of people voting for the Republican party per 1000 voters.

After cleaning each data set, I merge the five data sets by county and state. Based on the merged data set, I calculate the infection and death number per 1000 residents in each county. Therefore, each row in the merged data represents a US county and provides demographic information such as the total population. In addition, the data also contains the voting pattern in each county. The Table 3 summarizes all the important variables in this paper.

To understand the data clearly, Figure 1 presents the distribution of death rates for counties based on their voting preference for the two major parties. It seems that political prefer-

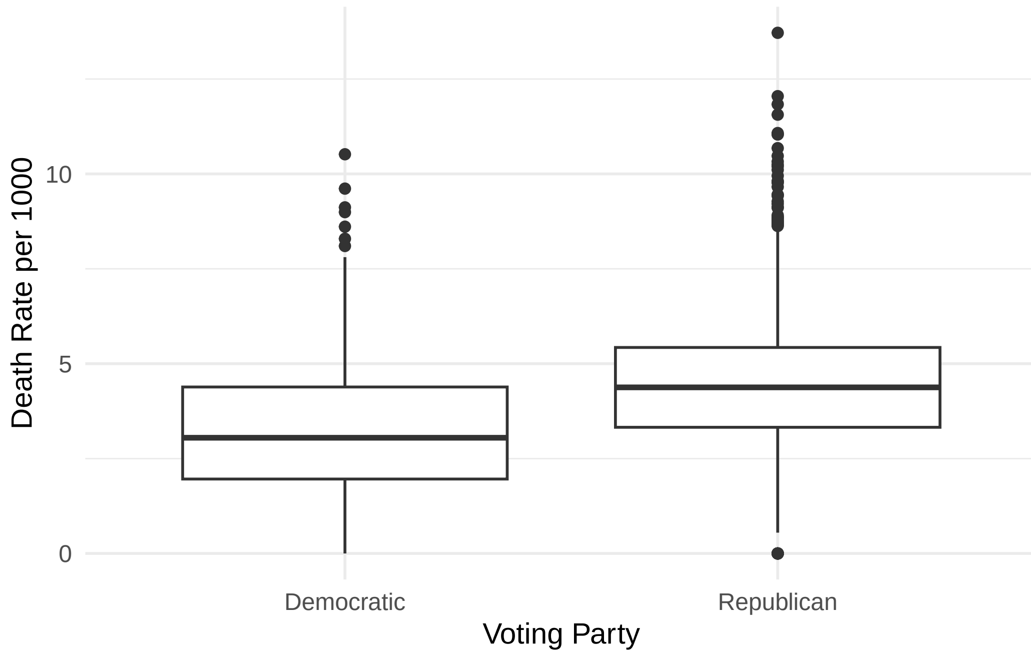


Figure 1: Summary of deaths rate for counties voting for Democratic and Republican

ence corresponds with variations in death rates. Specifically, counties that leaned Republican exhibited a relatively higher death rate compared to those favoring the Democratic party. Notably, outliers are present in both groups.

To better understand the distribution of death rate from a geography perspective, Figure 2 compares the support for Republicans with the corresponding death rate by plotting them onto maps. From the graph, we can observe that the dark color corresponds to the dark blue, meaning that the counties voting for the Republican seem to have a higher death rate than those voting for the Democratic Party. This pattern is especially significant for the central states.

## 2.4 Data limitations

Our data combines five data sets from three sources. Nonetheless, there are still limitations that could influence our results, potentially reducing the accuracy of our predictions.

First, there is a lack of COVID information in certain states, especially in "Utah." This can be observed from Figure 2, where the counties are grey, indicating the data for this county is missing. Due to this lack of one data set, the merged data set will also lack information on these counties. The accuracy of the subsequent analyses may also be decreased.

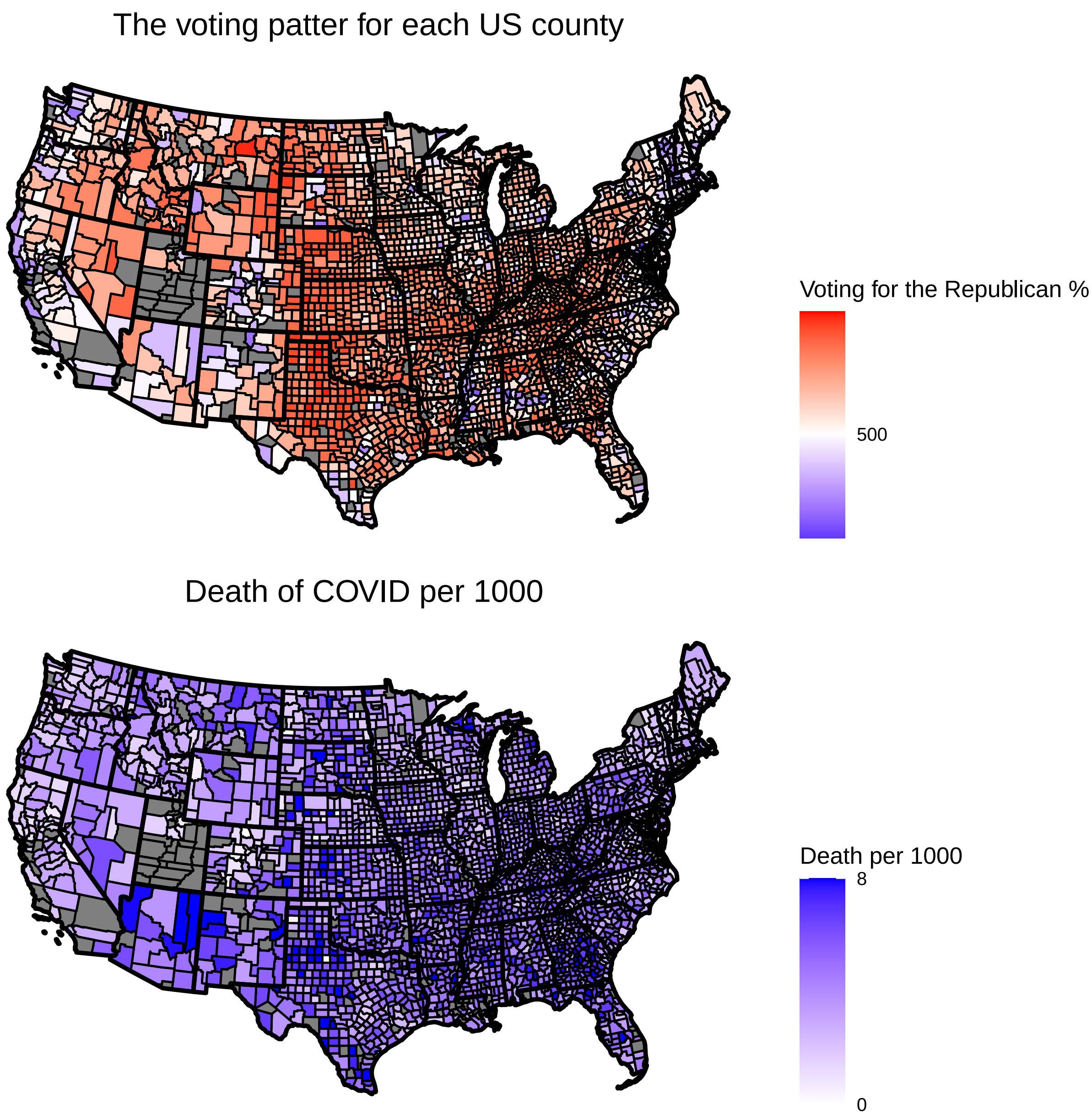


Figure 2: Election results of 2020 Federal Election and the corresponding death rate for each US county

# 3 Methods

As I introduced, this paper aims to predict the death rate of COVID in each country of the US and to detect whether this correlation is related to their political preferences. Based on the merged data above, I will perform a series of models, from simple to complicated, to detect which predictors are significant in predicting the death rate.

## 3.1 Model Specifics

Before we fit the models, I will split the data into training and testing data. The training data is used to fit all the models, and they are tested on the testing model to see which one performs better.

In each OLS regression model, the dependent variable will consistently be the COVID-19 death rate for each county. Given that death rates can be influenced by many factors, including demographics and economics, and given our interest in the potential impact of political party affiliation from the last Federal Election, I will establish two sets of models. The first set will not incorporate political preferences, while the second set will. Within each of these sets, there will be three models, each focusing exclusively on one category of predictors. Additionally, there will be a comprehensive model encompassing all predictors and a best-fit model determined by $R^2_{adj}$ and $RMSE$. In total, this approach will yield ten models, including the two best-fit models. To decide which model is the best, I will evaluate them using testing data to see which one has a lower testing $RMSE$.

The general form of the regression models is

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

where - $Y$ represents the response variables, which is the death rate. - $\beta_0$ represents the intercept, which is the death rate holding all variables to zero - $\beta_j$ represents the change of death rates with one unit increase in $X_j$ - $\epsilon$ captures measurement errors and other discrepancies

## 3.2 Assumptions of the regression models

After we fit the linear regression models, it is essential to check the assumptions to ensure the accuracy of our predictions. There are four main assumptions, which are linearity, uncorrelated error, constant variance, and normality. In this paper, I will assume the models satisfy both of the four assumptions.

Table 4: Summary of the Best Model Without Politics Preference

| Predictor | Estimate | Standard Error | Statistics | P-value |
|---|---|---|---|---|
| (Intercept) | 7.521 | 0.793 | 9.479 | 0.000 |
| high_education | -0.051 | 0.005 | -10.899 | 0.000 |
| pctile | -0.012 | 0.002 | -6.792 | 0.000 |
| no_insurance | 0.066 | 0.008 | 8.242 | 0.000 |
| private_insurance | -0.014 | 0.005 | -2.642 | 0.008 |
| males | -0.055 | 0.013 | -4.299 | 0.000 |
| age_85 | 0.303 | 0.034 | 9.012 | 0.000 |
| white | 0.011 | 0.004 | 2.766 | 0.006 |
| black | 0.013 | 0.004 | 3.149 | 0.002 |

### 3.3 Expected MSE and Bias-Variance Trade-off

The rationale for partitioning the data into training and testing sets is to validate our model. That said, by training a model using training data and then evaluating it using test data, we can determine which model produces the lowest expected MSE and better performance.

$$\mathbb{E}\left[(Y - \hat{f}(x))^2 \mid X = x\right] = \mathrm{Var}(\hat{f}(x)) + (\mathbb{E}[\hat{f}(x)] - f(x))^2 + \sigma^2$$

The above equation shows the expected MSE, which underscores the bias-variance trade-off. It tells us the trade-off between the model's complexity and the interpretability. An intricate model may increase the bias, but its variance will be higher if we vary the data, and vice-versa.

Hence, by evaluating our models using the testing data, we can know whether we should fit a relatively simple or complicated model.

## 4 Results

### 4.1 Model Excluding Political Preferences

Table 4 shows the summary table for the model without political references. Using the above information, we can write the equation model:

$$
\begin{aligned}
\text{Death Rate} = 7.521 \\
& - 0.051 \times \text{prop\_higher\_education} \\
& - 0.012 \times \text{IncomePctile} \\
& + 0.066 \times \text{no\_insurance} \\
& - 0.014 \times \text{private\_insurance} \\
& - 0.055 \times \text{males} \\
& + 0.303 \times \text{old\_85} \\
& + 0.011 \times \text{white\_pct} \\
& + 0.013 \times \text{black\_pct}
\end{aligned}
$$

The above equation shows the model excluding the political preferences. When all variables are set to zero, the anticipated death rate stands at approximately 7.2 per 1,000. Keeping other variables constant, an increase of one percent in the proportion of individuals having at least a bachelor's degree is associated with an expected rise of 0.05 in the death rate. Furthermore, for every percentile increase in income, the death rate is projected to drop by 0.02. Notably, there's a negative correlation between the proportion of uninsured individuals and the death rate; an uptick of 10% in the ratio of uninsured residents in a county may lead to an additional 0.7 deaths per 1,000.

Regarding demographic factors, the number of deaths is inversely correlated with the number of males in a county. One percent increase in the males will result in 0.06 less death. Moreover, it seems that older people are more likely to die from COVID. That said, there will be 0.29 more deaths if there is a one percent increase in the proportion of people aged above 85. Lastly, one percent increase in the proportion of white and black people is expected to have 0.01 more deaths.

### 4.2 Model Including Plotical Preference

Table 5 shows the summary of the best model without and with the political preferences. Based on them, we can write the regression question.

Table 5: Summary of the Best Model With Politics Preference

| Predictor | Estimate | Standard Error | Statistics | P-value |
|---|---|---|---|---|
| (Intercept) | 7.596 | 0.774 | 9.816 | 0.000 |
| rep_rate | 0.003 | 0.000 | 10.190 | 0.000 |
| high_education | -0.024 | 0.005 | -4.517 | 0.000 |
| pctile | -0.010 | 0.002 | -5.797 | 0.000 |
| private_insurance | -0.028 | 0.005 | -5.130 | 0.000 |
| no_insurance | 0.027 | 0.009 | 3.155 | 0.002 |
| males | -0.059 | 0.013 | -4.722 | 0.000 |
| age_85 | 0.308 | 0.033 | 9.391 | 0.000 |
| white | -0.008 | 0.004 | -1.999 | 0.046 |
| black | 0.013 | 0.004 | 3.277 | 0.001 |

$$\begin{aligned}\text{Death Rate} = 7.596 &\\ &+ 0.003 \times \text{Rep\_Rate}\\ &- 0.024 \times \text{High\_Education}\\ &- 0.010 \times \text{Pctile}\\ &+ 0.027 \times \text{No\_Insurance}\\ &- 0.028 \times \text{Private\_Insurance}\\ &- 0.059 \times \text{Males}\\ &+ 0.308 \times \text{Age\_85}\\ &- 0.008 \times \text{White}\\ &+ 0.013 \times \text{Black}\end{aligned}$$

The difference between the second and the first model is that the second model includes the political preference for each county as one predictor to predict the death rate due to COVID-19. Holding other variables, we can see that for every 100 more votes per 1000 for the Republican party, 0.3 more people are expected to die.

### 4.3 Comparing the testing error

The above two models are fitted using the training data. Comparing the training RMSE from Table 7 and Table 8, we can find that the model containing the political preference has a lower training RMSE. As introduced in the Methods part, to test the prediction performance of the models, I will calculate the test error for the two models to see which is better.

Table 6: The testing error of the above two models

| Model | Testinf Error |
|---|---|
| Model 1 | 1.769733 |
| Model 2 | 1.675969 |

Table 6 shows the testing error of the above two models. We can observe that the errors for the two models are 1.769733 and 1.675969. Therefore, the model containing the political preferences has a lower testing MSE; it has better predictions than the another one.

# 5 Discussion

To sum up, the previous results show that it is essential to consider the political reference when predicting the COVID mortality rate. However, combining political preference with other demographic and economic factors will be more convincing and interpretable. The key takeaways are as follows.

## 5.1 Although income levels are typically associated with variations in COVID-19 mortality, the influence of political preferences appears to be more significant than anticipated.

From the Table 5, both income percentile and the support of Republicans are significant in predicting the death rate of COVID-19, but with opposing effects. More people voting for the Republican party correlates with a rise in mortality; conversely, higher-income counties are expected to have a lower death rate. In terms of the values of coefficients, the coefficient for the income percentile is three times as large as the one for voting for Republican, meaning that for every additional three out of a thousand individuals voting Republican, the increase in death rate is equivalent to the decrease in the death rate by a one-percentile rise in income level. This suggests that the lowest-income counties may suffer disproportionately from COVID-19 fatalities, particularly where Republican support is vital.

Figure 3 corroborates our assertions. It shows the spread of the COVID-19 death rate for each state in the order of ascending income levels. The colour represents the party each state supported during the 2020 US Federal Election. Notably, twelve out of the top 15 poorest counties exhibited a preference for Donald Trump, while, in stark contrast, only one among the fifteen wealthiest states favoured the Republican Party. Regarding mortality, the Figure 3 also shows that the death rate decreases as income level increases, and the majority of counties voting for the Democratic party have a death rate lower than the national average level.

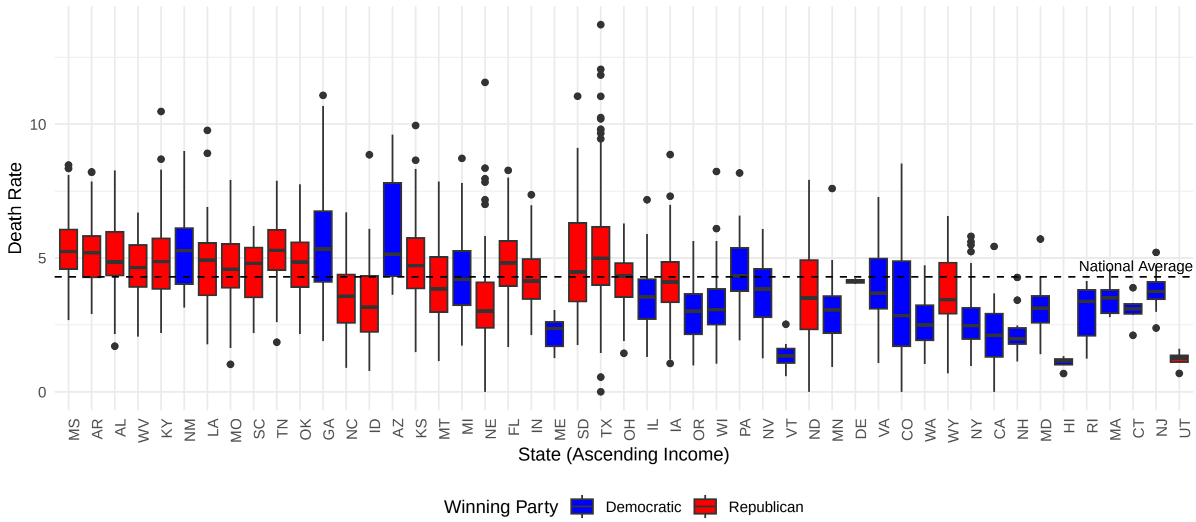


Figure 3: Variation in COVID-19 Mortality Rates Across States, Categorized by Income Levels and Dominant Political Preferences

Compared with Figure 3, Figure 4 examines the number of voters across the income percentile. It shows that Republicans are favoured among lower-income groups, while Democratic voters have grown significantly, corresponding to the wealthiest percentile.

Combining the Figure 3 and Figure 4, they yield an intriguing yet concerning observation. They suggest that lower-income voters, who disproportionately voted for Donald Trump, may inadvertently face higher mortality risks. This observation is underpinned by evidence indicating that COVID-19 mortality rates among low-income populations are nearly double those in higher-income brackets (Thomson Reuters 2022). While it is critical to approach causality with caution, these findings seem to suggest that Trump has an inescapable responsibility for the high death under his misleading during the pandemic.

## 5.2 The difference in high-education attainment and insurance structure may be potential factors causing the above phenomenon.

The inequality observed in COVID-19 mortality rates between low and high-income areas in the US may be potentially explained by the differences in educational attainment and insurance coverage, as suggested in Table 5. The model shows an inverse correlation between mortality rates and the percentage of individuals with high education levels: for every 10% increase in the population attaining a bachelor's degree or higher, there is a corresponding 0.24 decrease in the number of deaths.

Figure 5 shows the average death rate across different income percentiles and the corresponding average proportion of the population attaining a bachelor's degree or higher. The blue line

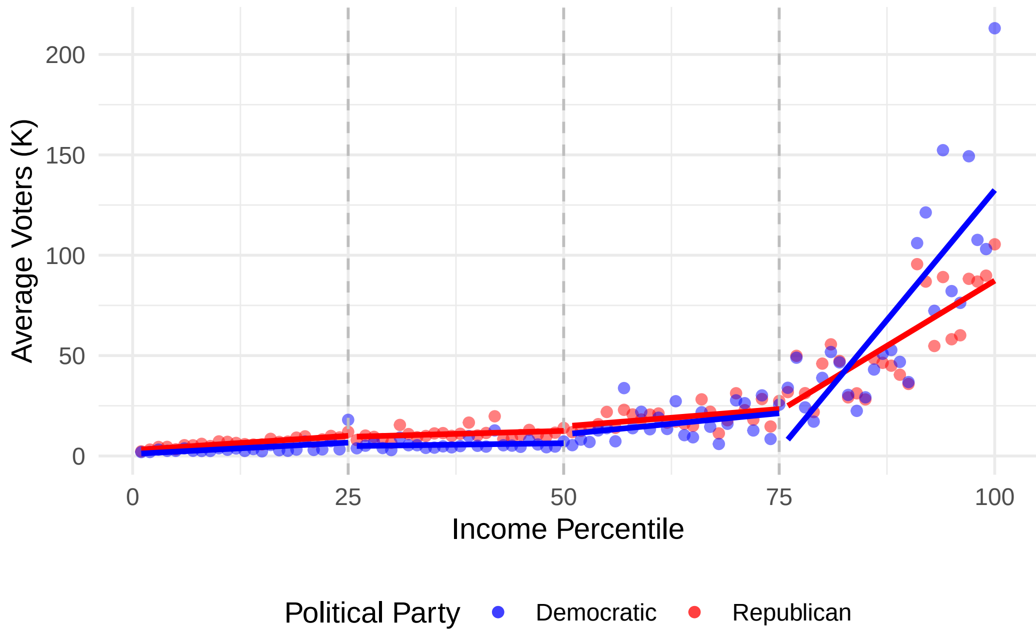


Figure 4: The average votes for the two party at each income percentile

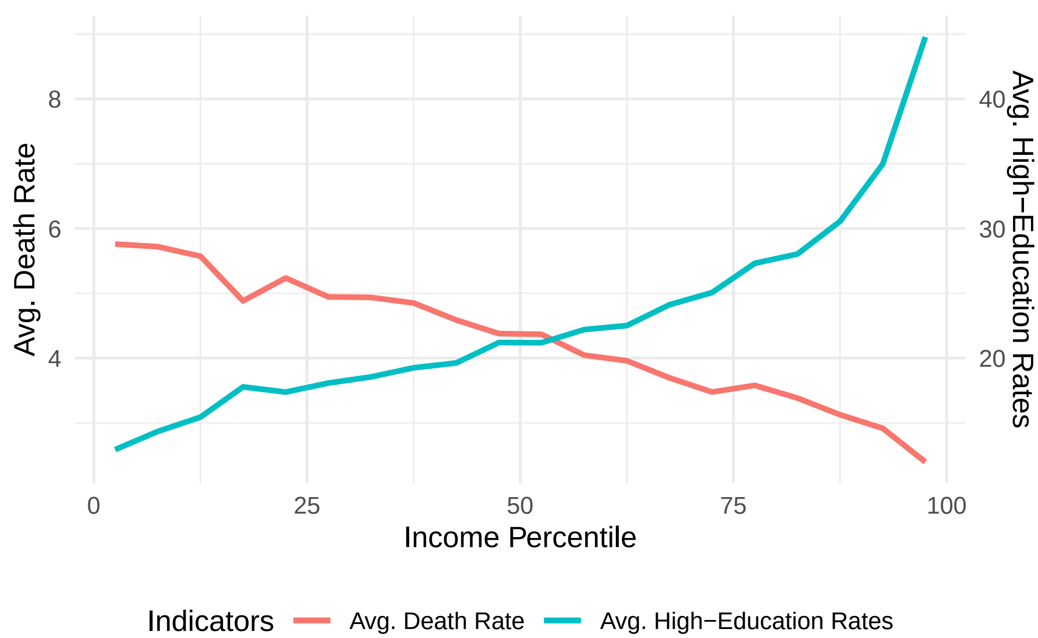


Figure 5: The change of average death rate across different education levels

represents the average high education rates, and the red line is the average death rate. From the graph, as the income levels increase, more people have at least a bachelor's degree and the average death rate decreases.

Therefore, one possible reason explaining the inequality of COVID mortality is that individuals with higher degrees may prefer to live in more affluent areas. These populations are generally more informed and may have differing perspectives on the effectiveness of the policies during the Trump administration, which may explain the lower support for Trump in wealthier counties or states.

In addition to education level, healthcare insurance is another significant in predicting the death rate for each county, as suggested on Table 5. The variables `no_insurance` and `private_insurance` demonstrate substantial inverse correlations with death rates. Interestingly, the magnitudes of their coefficients are nearly identical, suggesting that the presence of private insurance potentially counterbalances the impact on mortality rates from the lack of insurance.

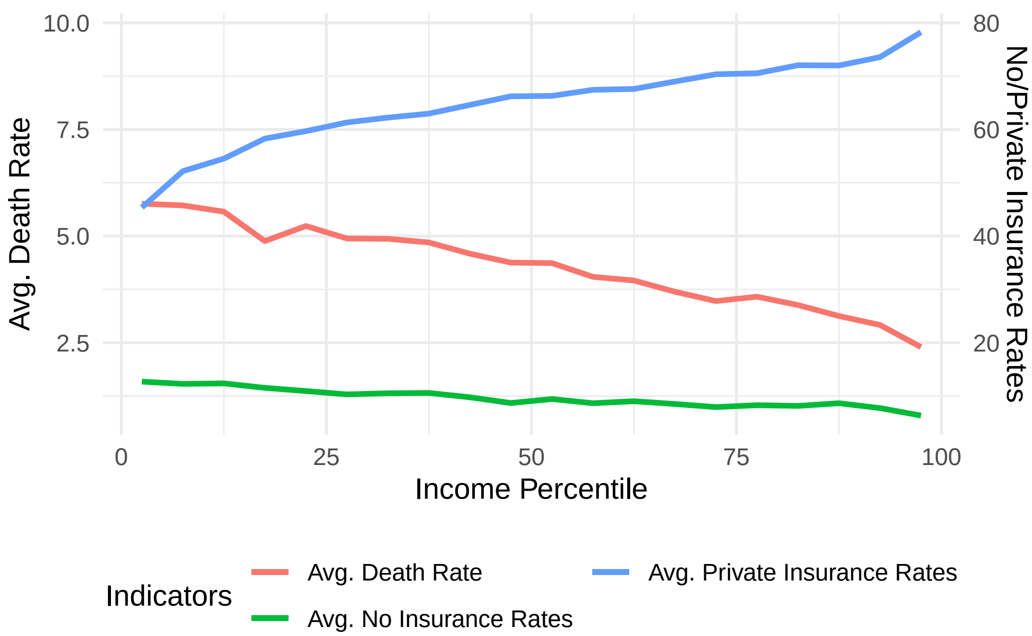


Figure 6: The change of average death rate across different education levels

Figure 6 delineates the correlation between income levels and the proportion of individuals with varying types of health insurance. Unsurprisingly, as income percentiles increase, more people have private insurance increases, and fewer people have no insurance. This pattern underscores the pivotal role of insurance during the pandemic, as people with insurance, especially private insurance,e may be more likely to have medicine or vaccines. Hence, the counties with more people having insurance are relatively lower compared to the counties with fewer people having insurance.

When considering both education and insurance, there is a convergent pattern among the diverging trends in poorer and wealthier countries. Residents of more affluent counties not only tend to be better educated but also exhibit a higher awareness about their health. This is possible because people living in wealthier counties care more about public health issues, and they can also afford the cost of private insurance. This composite of socio-economic advantages elucidates the observed discrepancy in COVID-19 mortality rates, with wealthier counties experiencing fewer fatalities than their less affluent counterparts.

# 6 Conclusion

This paper reflects on the socio-economic dynamics, COVID-19 mortality, and electoral outcomes in the United States over the past three years of the pandemic, examining variations in death rates across counties. The primary findings suggest that political preferences may have a more pronounced correlation with COVID-19 death rates than income levels. By illustrating the interplay between these two factors, it suggests a trend where lower-income counties, which showed a higher propensity to vote for Donald Trump, also experienced higher mortality rates. Throughout the model, education and insurance coverage seem to be the two key factors explaining the inequality of mortality rate between the wealthier and poorer counties. More educated people are richer, and they can afford the private insurance cost, meaning that they are more accessible to those healthcare facilities and hence have a lower death rate.

Since the political data used in this paper was the last US Federal Election results, this study could offer insights into the electoral landscape for the forthcoming Federal Election. By observing the death rate in the counties or states voting for Joe Biden, we can see that they have a lower death rate than the counties voting for Donald Trump. It may suggest that the Democratic party may be able to implement more effective public health measures. This can bring insights regarding the next federal election, which Joe Biden may be more likely to reelect.

## 6.1 Weaknesses

This paper has several limitations. First, the data on COVID and elections are lost for certain counties, especially the counties in Utah. In addition, I assume the models do not have any violations of their assumptions. Besides, the way that I choose the best model is to find the model in which all predictors are significant. There should be a more complicated procedure when selecting the models. With these drawbacks, it may lead to a bad performance of the models, and hence, the predictions might be inaccurate.

## 6.2 Future research

For future research, one primary thing to do is to use a more complicated model, such as Multilevel with Post-Stratification (MR.P). In addition, it would also be necessary to delve deeper into other factors that may affect the COVID-19 death rate, such as the healthcare capacity in each county. Furthermore, it will also be necessary to find the causal inference between the Electrol outcome and COVID.

# Appendix

## Model details

Table 7: Summary of the set of models with political preferences

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| (Intercept) | 6.524*** | 1.942+ | −0.810 | 11.840*** | 7.521*** |
| | (0.083) | (1.115) | (1.033) | (1.845) | (0.793) |
| high_education | −0.097*** | | | | −0.051*** |
| | (0.003) | | | | (0.005) |
| pctile | | −0.024*** | | −0.010*** | −0.012*** |
| | | (0.002) | | (0.002) | (0.002) |
| unemployment | | −0.078*** | | −0.020 | |
| | | (0.015) | | (0.016) | |
| no_insurance | | 0.100*** | | 0.027+ | 0.066*** |
| | | (0.014) | | (0.015) | (0.008) |
| private_insurance | | 0.017+ | | −0.031** | −0.014** |
| | | (0.010) | | (0.012) | (0.005) |
| public_insurance | | 0.048*** | | −0.002 | |
| | | (0.009) | | (0.012) | |
| males | | | 0.032* | −0.064*** | −0.055*** |
| | | | (0.016) | (0.013) | (0.013) |
| age_85 | | | 0.381*** | | 0.303*** |
| | | | (0.041) | | (0.034) |
| children | | | 0.073*** | −0.011 | |
| | | | (0.012) | (0.012) | |
| white | | | 0.009* | | 0.011** |
| | | | (0.004) | | (0.004) |
| black | | | 0.033*** | | 0.013** |
| | | | (0.005) | | (0.004) |
| prop_higher_education | | | | −0.027*** | |
| | | | | (0.006) | |
| old_85 | | | | 0.300*** | |
| | | | | (0.037) | |
| white_pct | | | | −0.009* | |
| | | | | (0.005) | |
| black_pct | | | | 0.012** | |
| | | | | (0.004) | |
| dem_rate | | | | −0.003*** | |
| | | | | (0.000) | |
| Num.Obs. | 2013 | 2013 | 2013 | 2013 | 2013 |
| R2 | 0.290 | 0.333 | 0.074 | 0.424 | 0.397 |
| R2 Adj. | 0.289 | 0.331 | 0.071 | 0.420 | 0.395 |
| AIC | 7222.9 | 7105.9 | 7765.5 | 6823.4 | 6907.5 |
| BIC | 7239.7 | 7145.1 | 7804.7 | 6901.9 | 6963.6 |
| Log.Lik. | −3608.428 | −3545.940 | −3875.735 | −3397.718 | −3443.764 |
| RMSE | 1.45 | 1.41 | 1.66 | 1.31 | 1.34 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Summary of the set of models with political preferences

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| (Intercept) | 5.336*** | 2.313* | 6.182*** | 8.772*** | 7.596*** |
| | (0.225) | (1.080) | (0.937) | (1.778) | (0.774) |
| rep_rate | 0.001*** | 0.003*** | 0.007*** | 0.003*** | 0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| high_education | −0.085*** | | | −0.026*** | −0.024*** |
| | (0.004) | | | (0.006) | (0.005) |
| pctile | | −0.017*** | | −0.010*** | −0.010*** |
| | | (0.002) | | (0.002) | (0.002) |
| unemployment | | −0.019 | | −0.019 | |
| | | (0.016) | | (0.016) | |
| no_insurance | | 0.061*** | | 0.024 | 0.027** |
| | | (0.014) | | (0.015) | (0.009) |
| private_insurance | | −0.008 | | −0.032** | −0.028*** |
| | | (0.010) | | (0.012) | (0.005) |
| public_insurance | | 0.028** | | −0.002 | |
| | | (0.009) | | (0.011) | |
| males | | | −0.062*** | −0.064*** | −0.059*** |
| | | | (0.014) | (0.013) | (0.013) |
| age_85 | | | 0.300*** | 0.299*** | 0.308*** |
| | | | (0.036) | (0.037) | (0.033) |
| children | | | −0.034** | −0.011 | |
| | | | (0.011) | (0.012) | |
| white | | | −0.041*** | −0.011* | −0.008* |
| | | | (0.004) | (0.005) | (0.004) |
| black | | | 0.014*** | 0.011** | 0.013** |
| | | | (0.004) | (0.004) | (0.004) |
| Num.Obs. | 2013 | 2013 | 2013 | 2013 | 2013 |
| R2 | 0.301 | 0.376 | 0.303 | 0.427 | 0.427 |
| R2 Adj. | 0.300 | 0.374 | 0.301 | 0.424 | 0.424 |
| AIC | 7192.9 | 6973.7 | 7195.5 | 6811.7 | 6807.8 |
| BIC | 7215.3 | 7018.6 | 7240.3 | 6890.2 | 6869.5 |
| Log.Lik. | −3592.435 | −3478.864 | −3589.739 | −3391.850 | −3392.893 |
| RMSE | 1.44 | 1.36 | 1.44 | 1.30 | 1.31 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## References


Al Jazeera. 2023. "Three Years, 1.1 Million Deaths: Covid Emergency Ending in US." May 11, 2023. https://www.aljazeera.com/news/2023/5/11/three-years-1-1-million-deaths-covid-emergency-ending-in-us.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

"Available APIs." n.d. Census.gov. n.d. https://www.census.gov/data/developers/data-sets.html.

Bureau, U. S. Census. 2023a. "American Community Survey 5-Year Data (2009-2021)." 2023. https://www.census.gov/data/developers/data-sets/acs-5year.html.

———. 2023b. "American Community Survey (ACS)." 2023. https://www.census.gov/programs-surveys/acs.

CSSE, JHU. 2023. "COVID-19 Content Portal." 2023. https://systems.jhu.edu/research/public-health/ncov/.

Di Lorenzo, Paolo. 2023. *Usmap: US Maps Including Alaska and Hawaii*. https://CRAN.R-project.org/package=usmap.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

McGovern, T. n.d. "US_COUNTY_LEVEL_ELECTION_RESULTS_08-20." GitHub. n.d. https://github.com/tonmcg/US_County_Level_Election_Results_08-20.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Network, FOX News. n.d. "Elections 2020." Fox News. n.d. https://www.foxnews.com/elections/2020/general-results.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Thomson Reuters. 2022. "U.s. Poor Died at Much Higher Rate from Covid Than Rich, Report Says." Reuters. 2022. https://www.reuters.com/world/us/us-poor-died-much-higher-rate-covid-than-rich-report-2022-04-04/.

United Nations. 2023. "WHO Chief Declares End to COVID-19 as a Global Health Emergency | UN News." https://news.un.org/en/story/2023/05/1136367.

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

———. 2022. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr.

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 François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=

dplyr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.