Comparison of Effects of COVID-19 on Counties With and Without Coal Mines

https://github.com/tnh2/ENV872_Project

Thomas Hancock

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

The purpose of this analysis is to explore whether counties with coal mines are impacted differently by COVID-19. With the novel coronavirus rapidly spreading through the United States, different locations are likely to be affected in different ways. In order to perform the analysis, I combined datasets containing information about coal mine locations, populations, and COVID-19 cases and deaths at the county level. Because other factors such as socioeconomic levels, population densities, and even government responses to the COVID-19 crisis vary widely across states, I decided that it was best to compare counties within states rather than performing analysis on the entire country. In particular, I focused on Pennsylvania and West Virginia. Pennsylvania has had a relatively large number of confirmed COVID-19 cases, meaning there is a reasonable amount of data to analyze. West Virginia has had relatively few confirmed cases, but I included it as a comparison to Pennsylvania since West Virginia coal mines have continued operation whereas Pennsylvania mines were closed in March [1].

I am curious to see if there is any difference in COVID-19 outcomes between counties with coal mines and those without. In particular, I am interested in whether or not there will be a difference in mortality (the percentage of people who have the disease that die from it). The rational behind this exploration is that COVID-19 seems to be particularly lethal among people with existing conditions, especially respiratory issues. In turn, coal miners frequently suffer respiratory issues from the large amonts of dust that they encounter in their jobs. As such, I wondered if there might be a higher mortality rate in coal mining counties than non-coal mining counties (as a proxy for mortality amongst coal miners versus others). Likewise, coal mining involves sustained close-quarters operation and minimal opportunity for sanitization, hence my wanting to compare counties in Pennsylvania where the mines shut down in response to the outbreak to counties in West Virginia where mines continued operation.

My research questions are:

1. Do counties with coal mines have different mortality rates than similar counties without coal mines in the same state?
2. Do counties with coal mines in Pennsylvania have different mortality rates than West Virgnia counties with coal mines?

2 Dataset Information

To perform my analysis, I gathered together five datasets: US Census state and county boundaries (as shapefiles), US Census population data for US counties, coal mine locations from the US Department of Energy (as a shapefile), and county-level COVID-19 case and death counts from the Johns Hopkins University Center for Systems Science and Engineering (CSSE). The population data from the US Census Bureau are from 2019 estimates. The coal mine information inlcudes all operating surface and underground mines in the united states. While the dataset includes information on the mine type and output, I only used the location data for this analysis. The COVID-19 data from Johns Hopkins includes confirmed COVID-19 cases and reported deaths by county as of April 23rd, 2020.

I started wrangling my data by finding which counties had coal mines in them using the geospatial analysis capabilities of the sf package in R. I then joined the COVID-19 data with the census population data using the county FIPS codes and calculated metrics for COVID-19 case rates (per 100,000 people), mortality rate (deaths per confirmed COVID-19 case), and death rates per 100,000 people in the county. This data was then combined with the data above identifying whether a county had coal mines or not to complete the data joining. From this country-wide dataset, I filtered out the data specific to the two states I focused on for this study (Pennsylvania and West Virginia). The important data included in these data sets are outlined in Table 1 below.

Table 1: Important Data Fields

Field Name	Description	Units
STNAME	State Name	
CTYNAME	County Name	
POPESTIMATE2019	2019 Estimated Population	People
FIPS	5-digit county identifier	
Confirmed	Confirmed COVID-19 cases	Total Cases
Deaths	Reported COVID-19 deaths	Total Deaths
case_rate	Rate of COVID-19 cases	Cases per 100,000 people
death_rate_conf	Mortality	Deaths per confirmed case
death_rate_pop	Deaths rate in population	Deaths per 100,000 people
Mines.Label	Whether the County has mines	Mines or No Mines

3 Exploratory Analysis

3.1 Locational Verification

The first step of my analysis was to make sure that the counties with coal mines were properly identified. To do this verification, I made a map showing the counties that were identified as having a coal mine on one layer, and the point locations of the mines on another layer. If the counties were correctly identified, each of the mine points should be within one of the coal county boundaries. This map is shown in Figure 1. As the figure shows, the counties were correctly identified (i.e., no mines are in a non-coal county and all coal counties have at least one mine).

3.2 Pennsylvania Data

Turning my attention to Pennsylvania, I started by creating a histogram of the mortality rates of the coal counties and the non-coal counties. The first histogram I created is shown in Figure 1 below.

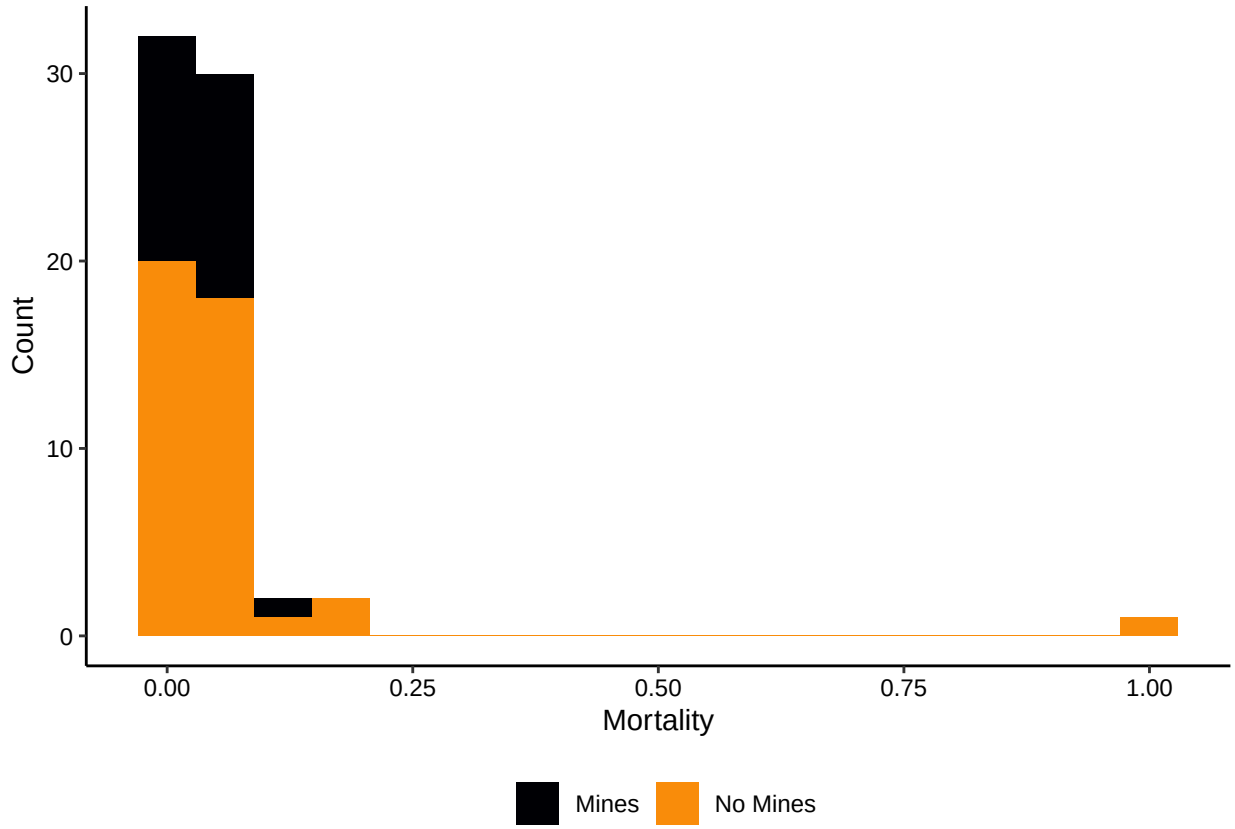

Figure 1: Histogram of Mortality in Pennsylvania counties

Two things are immediately evident: the data are not normally distributed, and there are some strange cases where the mortality is 1 (i.e., everyone confirmed case of COVID-19

resulted in a death). Based on general knowledge that COVID-19 is not likely to be lethal enough to kill even a majority of patients, we can assume that these counties had such limited testing that only the mortalities were tested. For the sake of this analysis, we will discard those counties and any others that have a mortality rate greater than 50%. We will assume that the remaining counties had testing that was sufficiently widespread to provide a reasonable approximation of the mortality. This new dataset yields the histogram in Figure 2. The distribution is still not normal, but it has a smoother shape for both the counties with mines and those without. There are still a few counties with suspiciously higher mortality rates, but without further data, there is not enough reason to remove them.

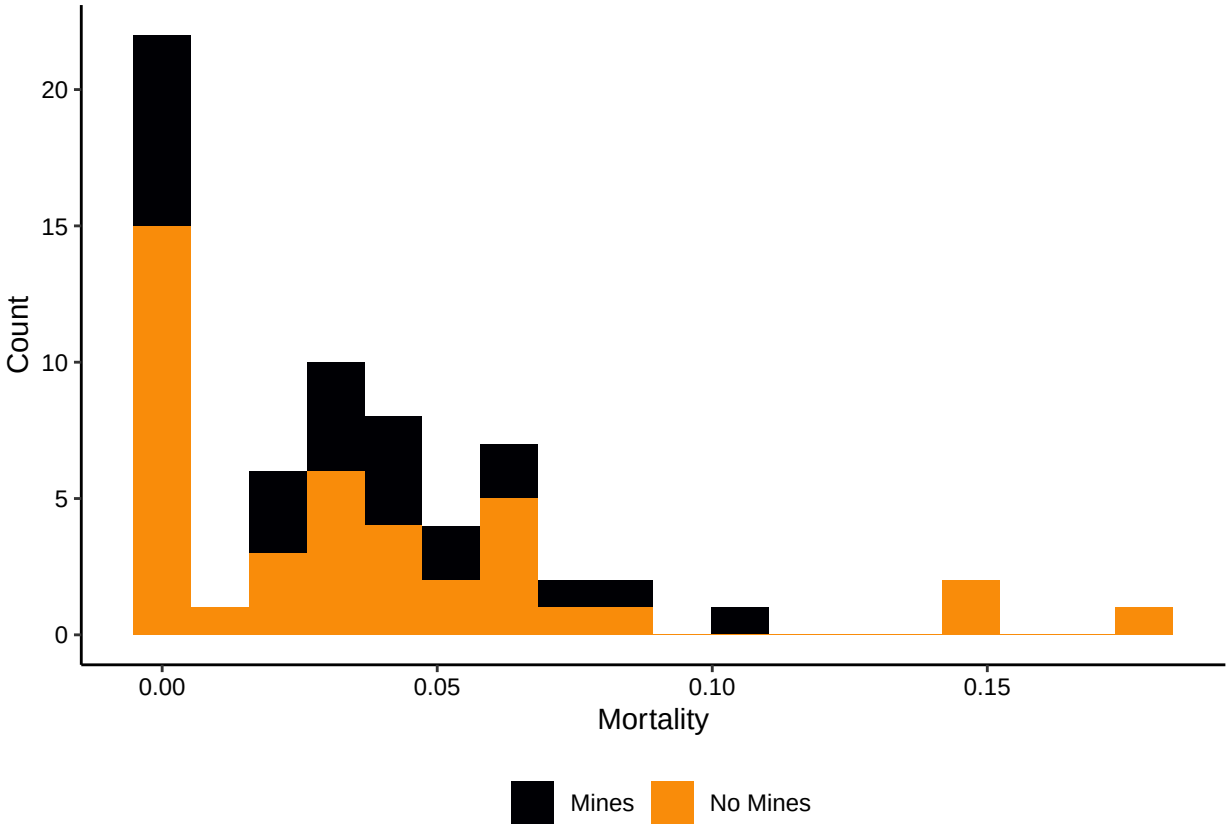

Figure 2: Histogram of Mortality in Pennsylvania counties (outliers removed)

I decided to explore the potential relationship between mortality and the amount of testing by plotting mortality against the number of confirmed cases per 100,000 people in each county. The results are presented in Figure 3. Even without the aforementioned counties that only tested fata cases, there appears to be a slight decreasing trend in mortality as the confirmed case rate increases. However, this trend is not statistically significant (linear regression; p-value = 0.395; F-statistic = 0.7344; degrees of freedom: 1 and 64).

- Map
- Histograms
- Scatterplot

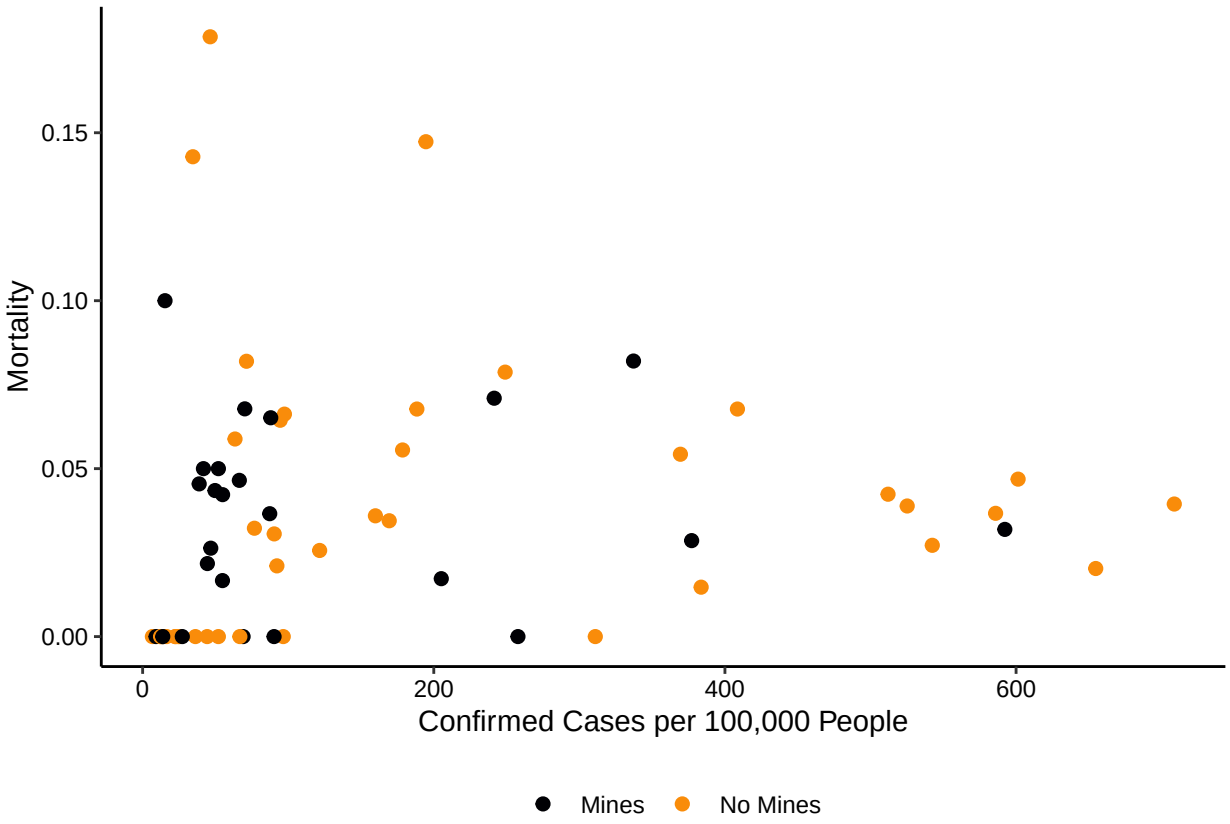

Figure 3: Mortality vs confirmed COVID-19 case rate for Pennsylvanian counties with and without coal mines.

3.3 West Virginia Data

As mentioned earlier, I am also interested in seeing if there is a difference between West Virginia, where coal mines stayed open despite the pandemice, and Pennsylvania, where mines were closed. In Figure 5, I plotted histograms for the West Virginia data, similar to what was done for Pennsylvania. From these histograms, we can see that there were a few West Virignia counties that tested mostly fatal cases of COVID-19, so it is reasonable to perform the same removal of ouliers with mortality rates above 50%. Likewise, the West Virginia data has a very non-normal distribution.

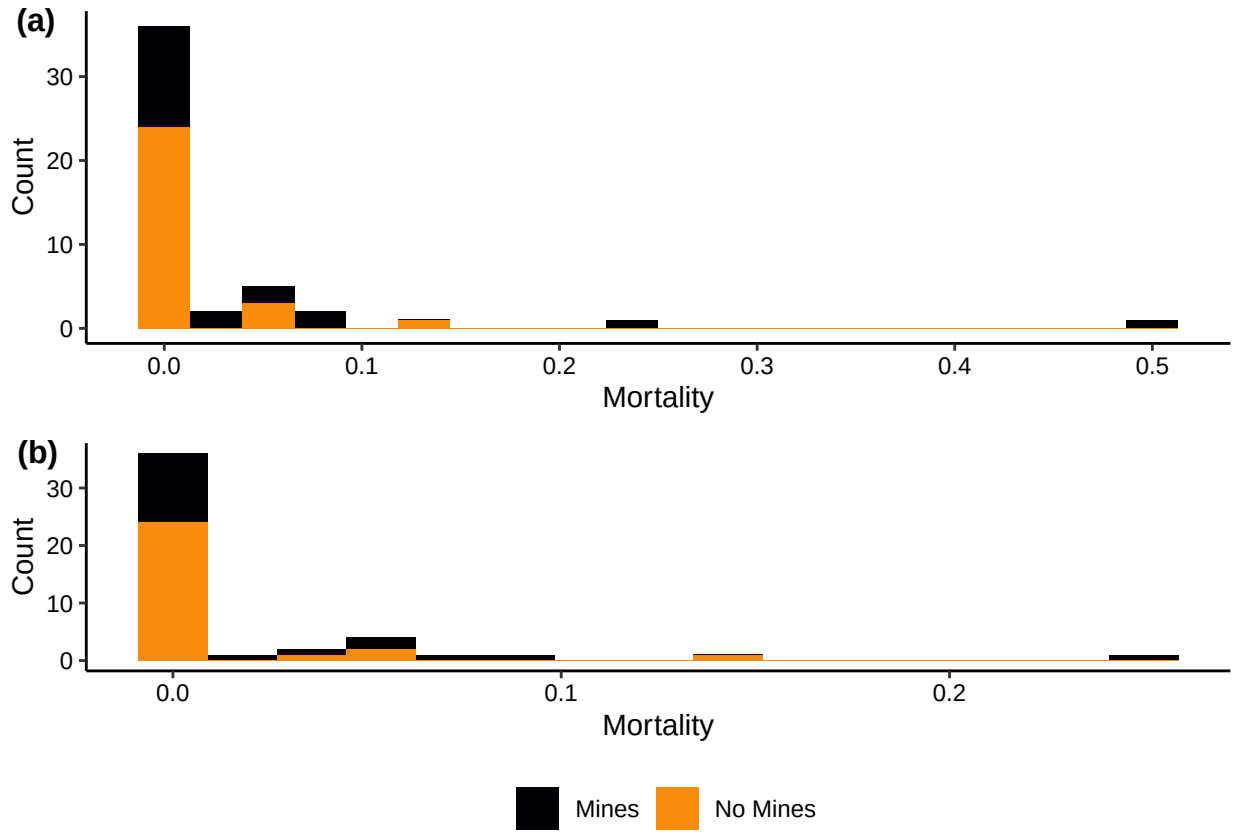

Figure 4: Histogram of Mortality in West Virginia Counties (a) with outliers and (b) without outliers.

Plotting the mortality against the confirmed case rate for West Virginia counties, as shown in Figure 6, reveals little structure to the data. Perhaps the most useful conclusion that can come from this exploration of the data is that West Virginia has a rather low reported case rate.

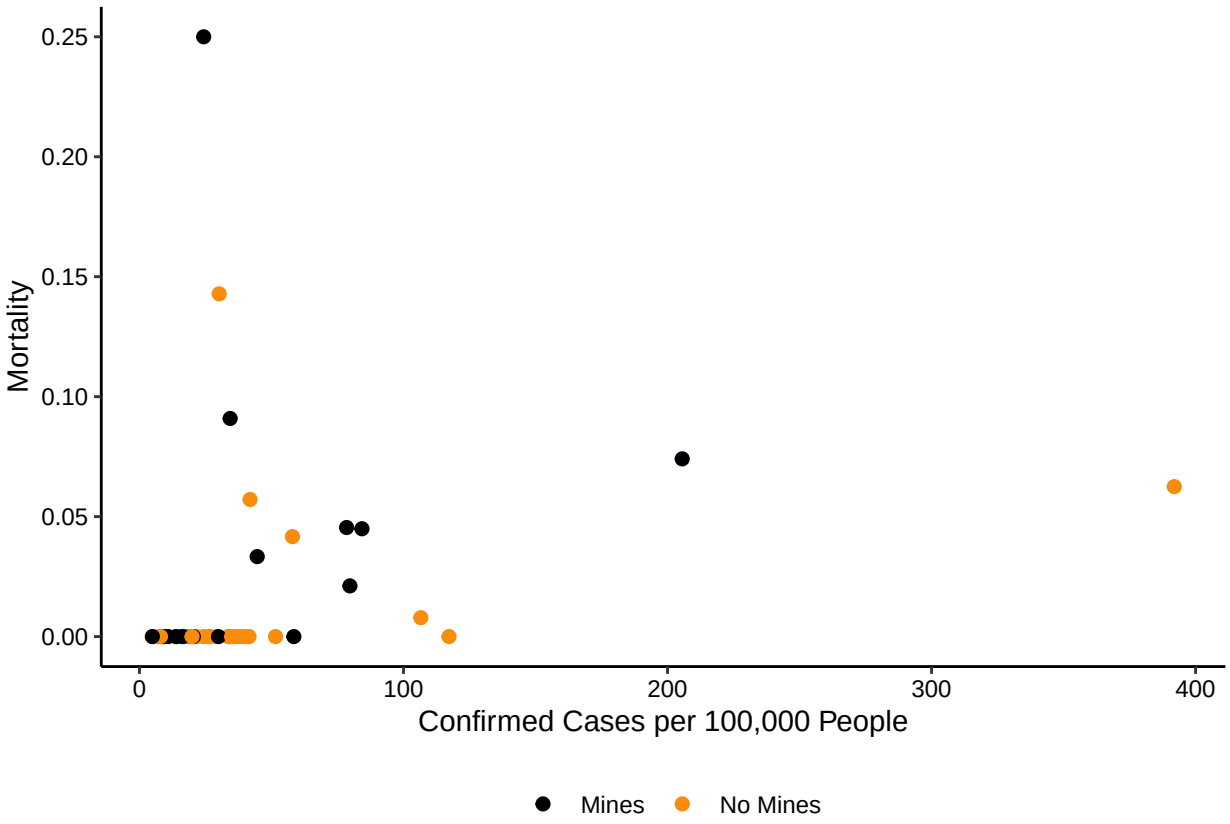

Figure 5: Mortality vs confirmed COVID-19 case rate for West Virginian counties with and without coal mines.

4 Analysis

- Boxplots
- Wilcoxon results

4.1 Question 1: Do counties with coal mines have different mortality rates than similar counties without coal mines in the same state?

To answer the first question comparing coal mining counties in Pennsylvania to non-mining counties, I performed a non-parametric Wilcoxon test. I chose to use a non-parametric test because the data are not distributed in a normal distribution (as seen in the Exploration section). The Wilcoxon uses a ranking system to see if two data sets have similar shapes.

4.2 Question 2: Do counties with coal mines in Pennsylvania have different mortality rates than West Virgnia counties with coal mines?

5 Summary and Conclusions

- No significant differences
- Reasons why this analysis isn't great

6 References

<add references here if relevant, otherwise delete this section>