Real Data Project:

Analyzing the Impact of COVID-19 on Minority Populations

Jessica Taylor

Indiana University – Purdue University Indianapolis

22 February 2021

**Abstract**

The purpose of this research was to analyze the relationship between COVID-19 cases in the United States and minorities within its population. Using data from The COVID Tracking Project, the study aimed to further test data related to the cases and deaths of racial/ethnic minorities within the United States. Linear regression models conducted on the data revealed that some minority groups experienced more cases and less deaths, and other minority groups experienced less cases and more deaths. However, when combined, it remained true that minorities, overall, fell victim to more cases and deaths than their population percentage, and whites experienced less.

Keywords: COVID, COVID-19, coronavirus, linear regression

**Introduction**

One study found the existence of a racial/ethnic disparity was likely for almost every state within the United States (Racial). However, it is not clear how these disparities affect every minority individually. The goal of this study is to reveal a more detailed look into the disparities in COVID-19 cases and COVID-19 related deaths; the study will show how concentrated or dispersed these disparities are among different groups.

**Methods**

In this study, there were multiple linear regression models to examine COVID-19 cases and COVID-19 related deaths. After establishing the groups to examine— white/Caucasian, black/African American, Hispanic/Latino, and Asian— the regression lines could be created. A regression line needed to be created that represented minorities, as well as multiple regression lines to represent each minority individually. By studying the variations of the regression lines between each minority, the way each minority suffers from COVID-19 will be revealed.

**Results**

Analysis of the linear regression models in *Figures 1.1 and 1.2* indicated that there was a larger 'gap', or disparity, between whites and minorities when it came to COVID-19 cases than it did for COVID-19 related deaths. The red line represents the mean population of the white community, and the black line represents the mean population of the selected minorities— black/African American, Hispanic/Latino, and Asian— combined.

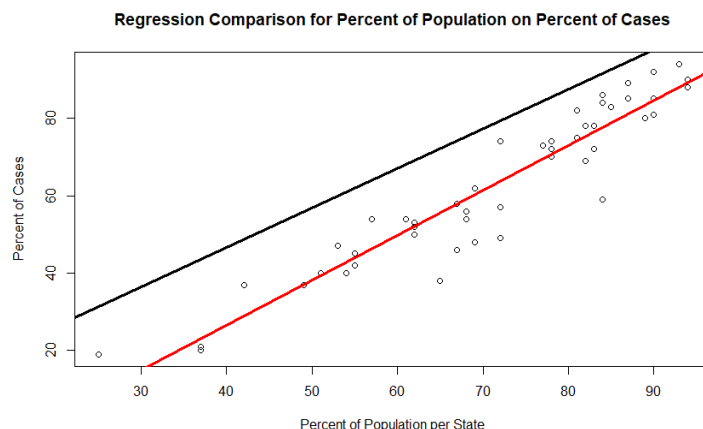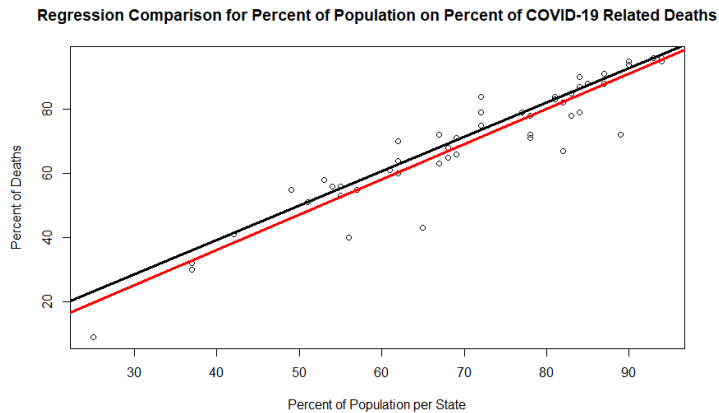Figure 1.1 - Regression Comparison for Percent of Population on Percent of Cases



Regression Comparison for Percent of Population on Percent of Cases

Figure 1.2 - Regression Comparison for Percent of Population on Percent of Deaths



Regression Comparison for Percent of Population on Percent of COVID-19 Related Deaths

Analysis of each individual group revealed variations in the percentage of cases and deaths. The white population had similar means in population and deaths, being 70.96 and 69.83673. However, their case mean was 62.79592, and had the weakest case correlation out of every other group— 0.9446668. The data suggests that there is little disparity in the white population's deaths, but a considerable difference in the number COVID-19 cases— which is less per percent population than expected.

Figure 2.1



Boxplot for White Population, Cases, and Deaths

1: Population,     2: Cases,     3: Deaths

Figure 2.2

**Regression for White Population on the Number of White COVID-19 Cases**



Figure 2.3

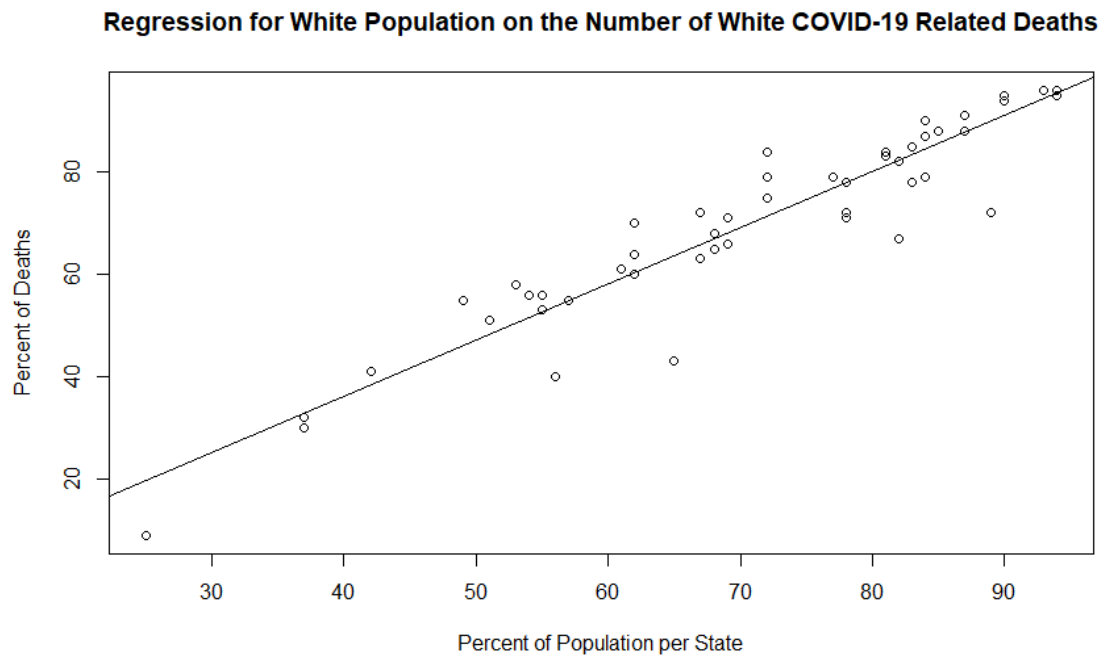**Regression for White Population on the Number of White COVID-19 Related Deaths**
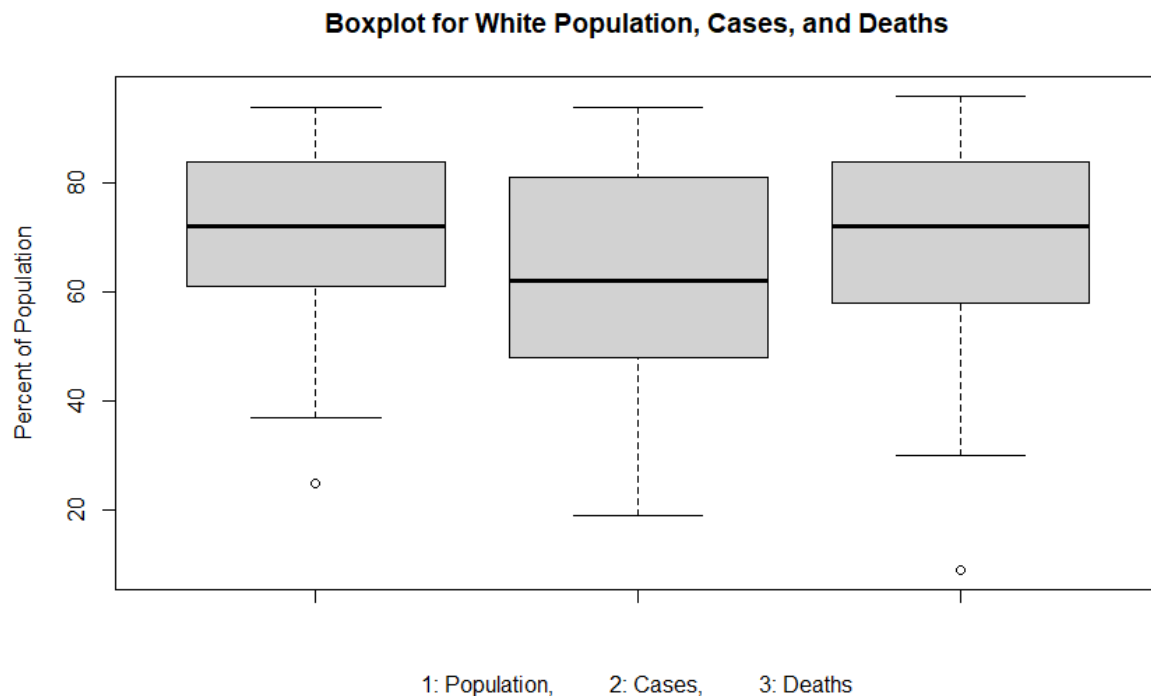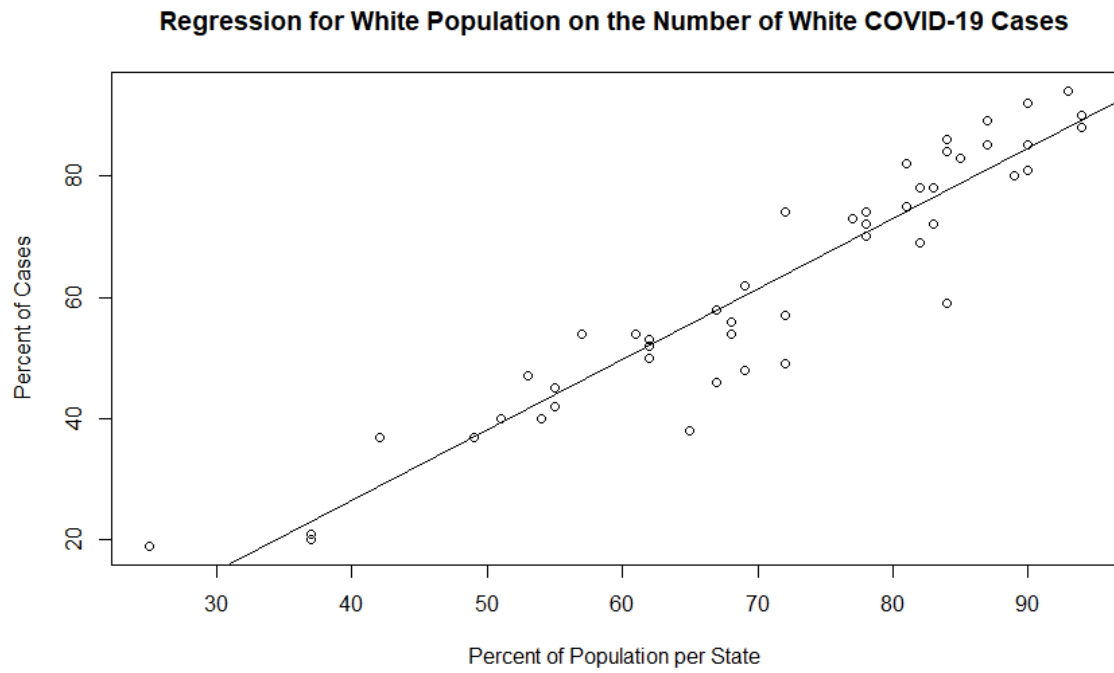
The black and African American community had slightly higher case and death means to their population; the population mean was 10.4, the case mean was 11.16327, and the death mean was 11.69388. The correlation between population-to-cases and population-to-deaths was also very high, being 0.983066 and 0.9836875. The data suggests that the black and African American community consistently experiences a slight disparity in both COVID-19 cases and COVID-19 related deaths.
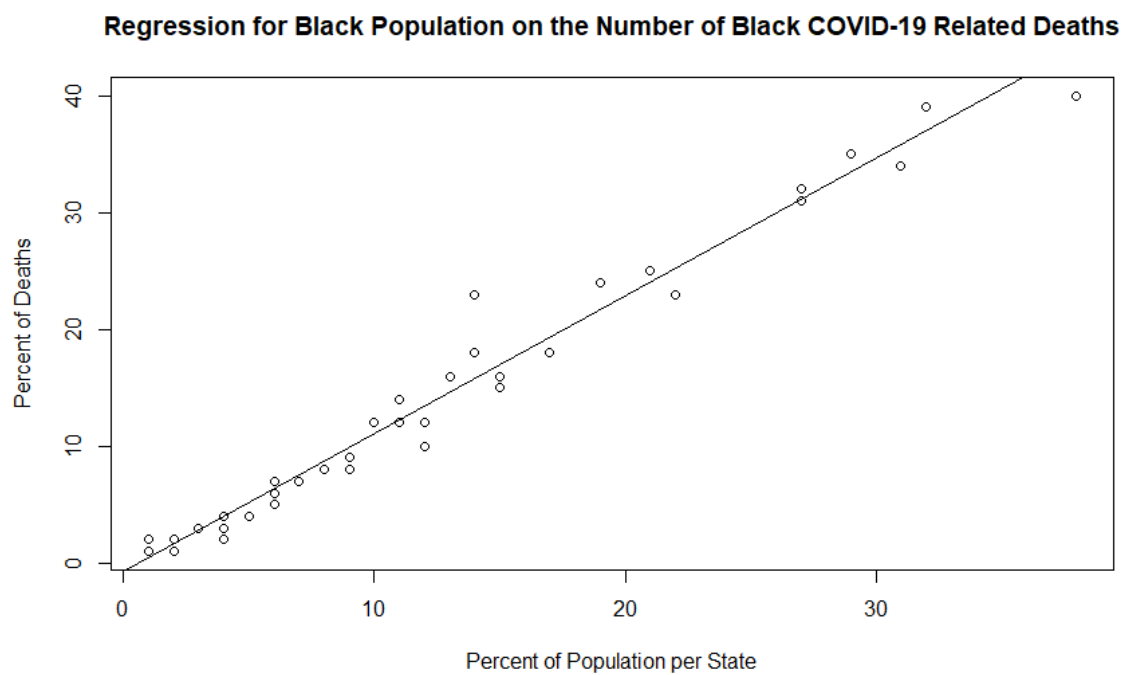
Figure 3.1



**Boxplot for Black Population, Cases, and Deaths**

1: Population,       2: Cases,       3: Deaths

Figure 3.2

**Regression for Black Population on the Number of Black COVID-19 Cases**



Figure 3.3

**Regression for Black Population on the Number of Black COVID-19 Related Deaths**
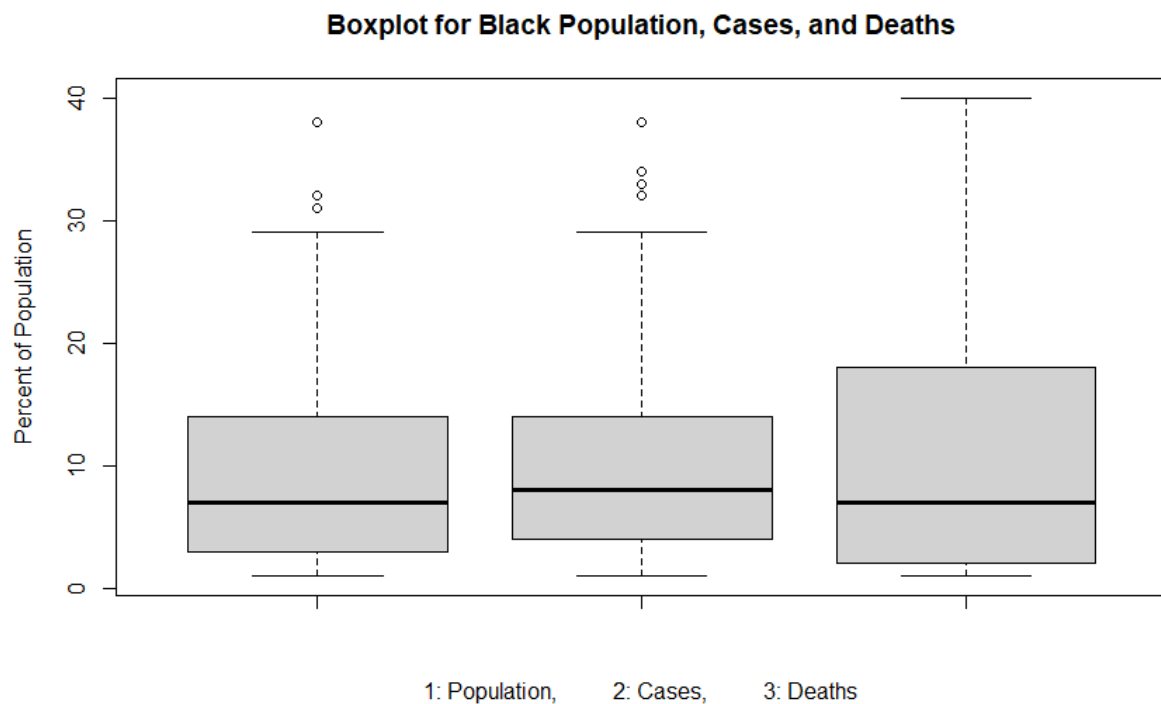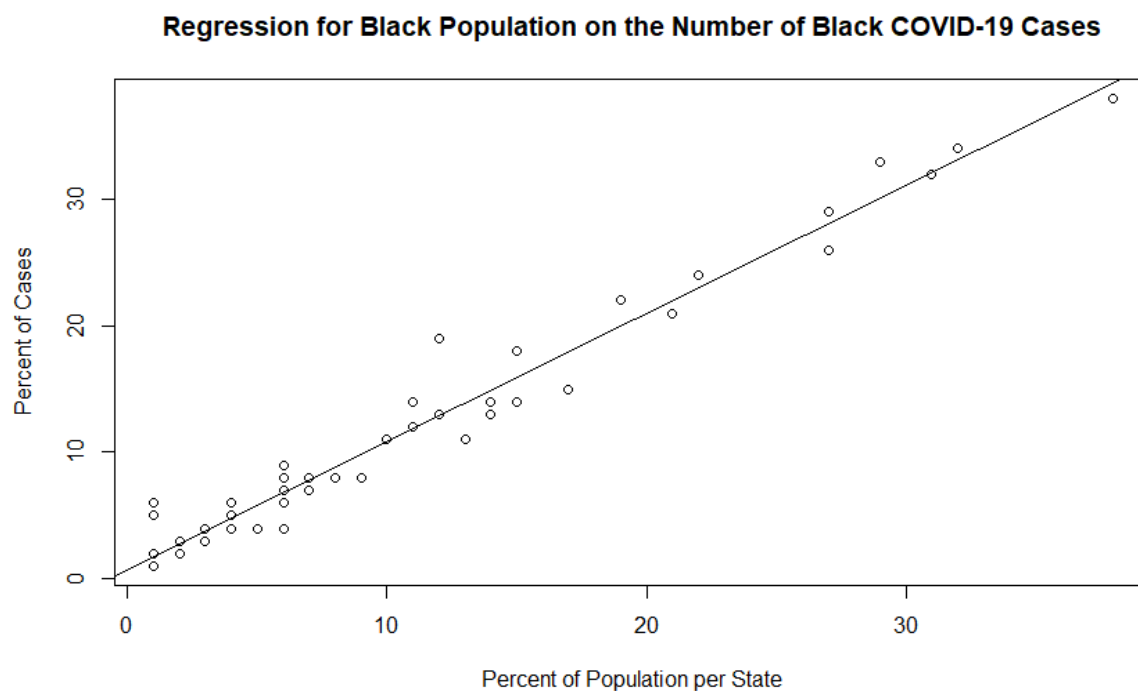
The Hispanic and Latino community had a significantly higher case mean, but a slightly lower death mean than their percent population. The population mean was 12.34043, the case mean was 18.86667, and the death mean was 10.67391; this means that the case mean was 53% higher than expected, but the death mean was 15% lower than expected. Overall, the data suggests that the Hispanic and Latino community experiences a high disparity in COVID-19 cases, but a slight difference in COVID-19 related deaths— which is less per percent population than expected.

Figure 4.1



Boxplot for Hispanic Population, Cases, and Deaths

Figure 4.2

**Regression for Hispanic Population on the Number of Hispanic COVID-19 Cases**



Figure 4.3

**Regression for Hispanic Population on the Number of Hispanic COVID-19 Related Deaths**
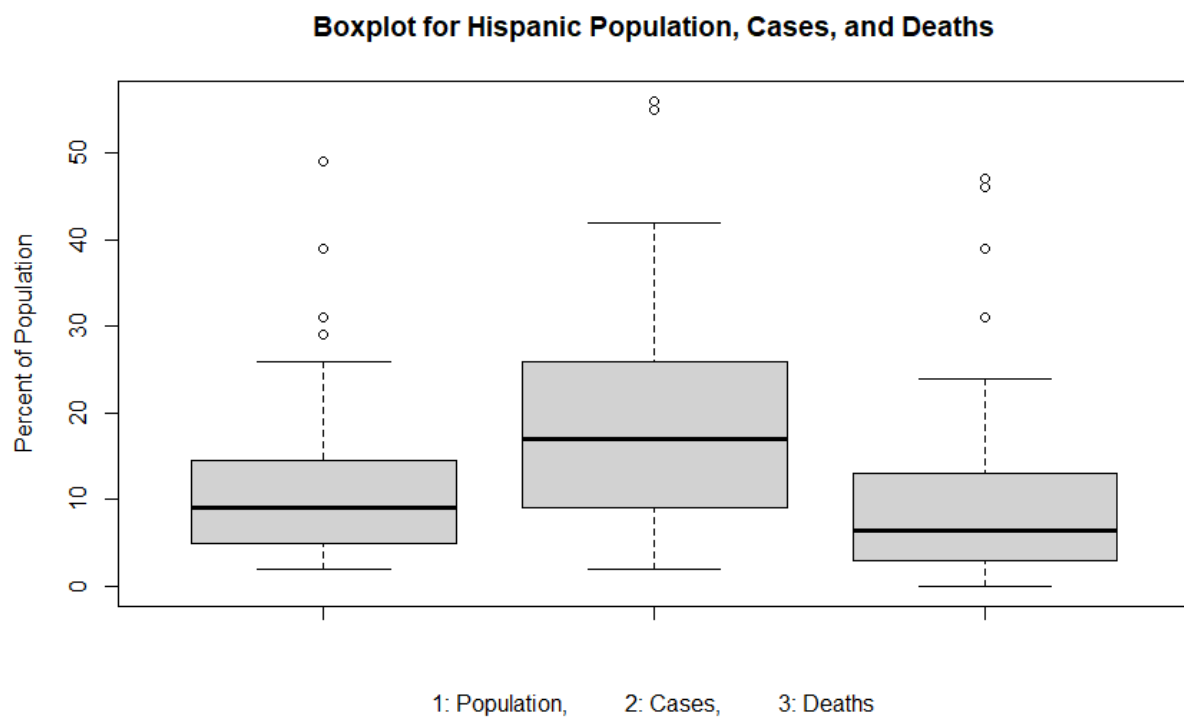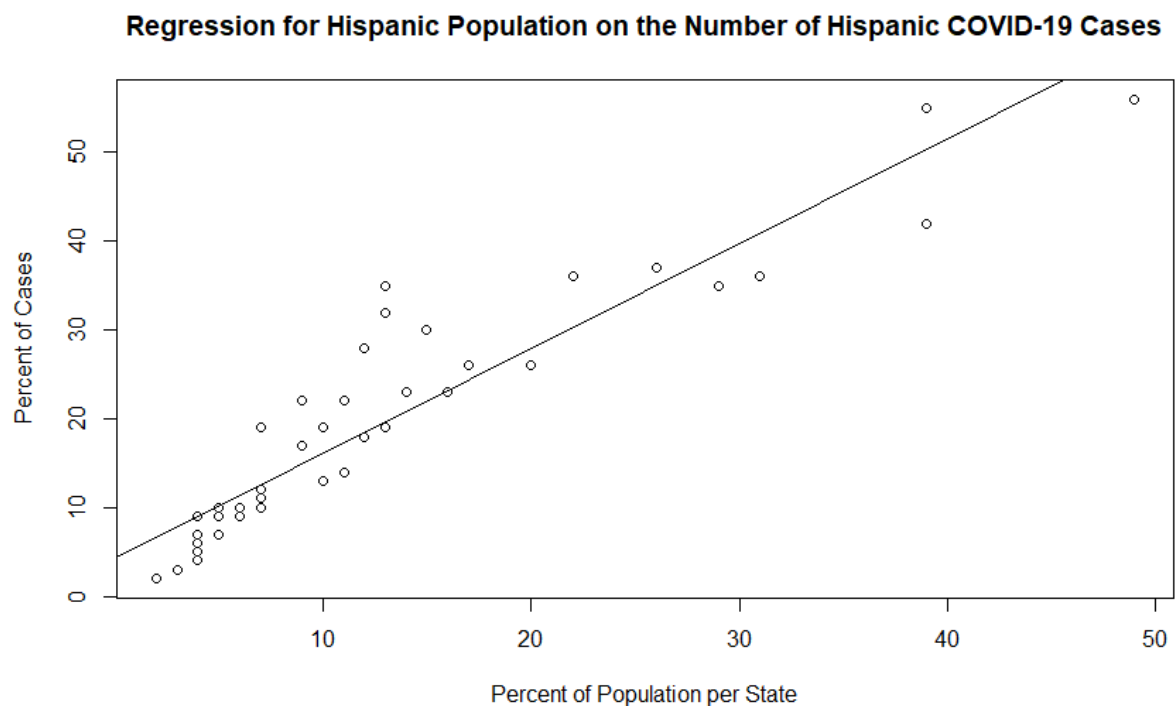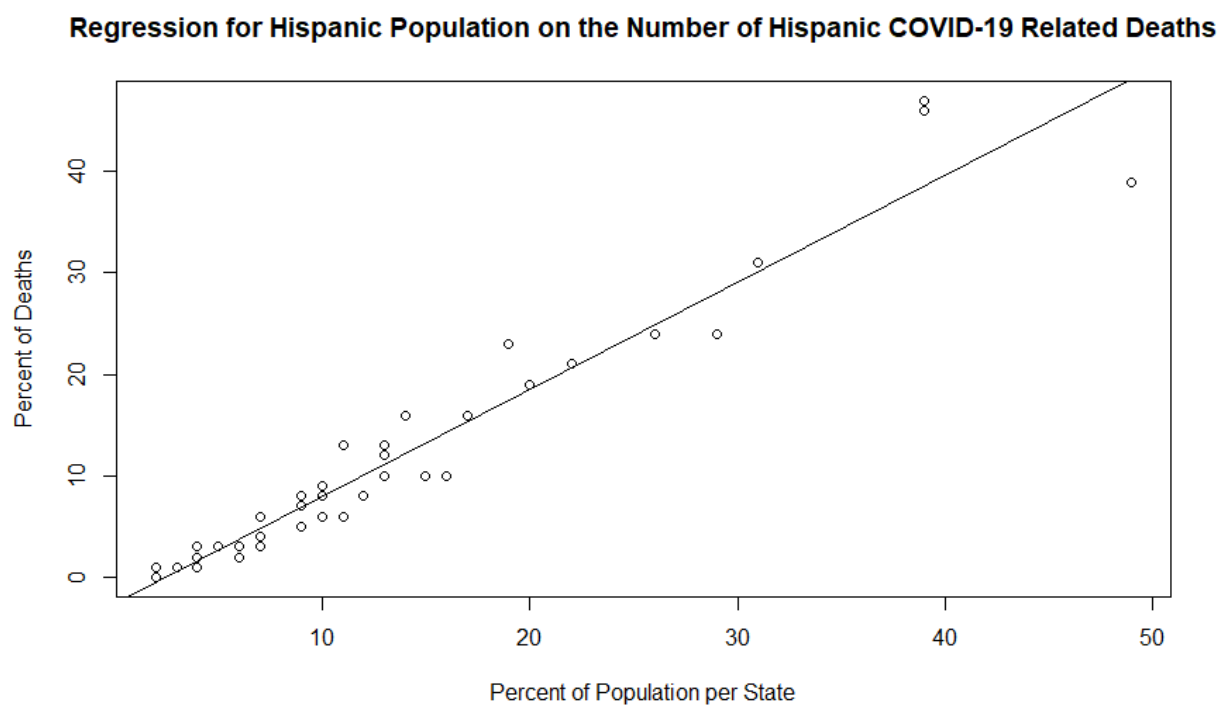
The Asian community had a lower case mean and a lower death mean than their population mean; the case mean was 3.061224, the death mean was 3.591837, and the population mean was 4.12. The data for the Asian community had one large outlier, which has slightly skewed the slopes of the linear graphs. With the exception of the Asian community in Hawaii, which experienced a 50% higher COVID-19 related death expectancy, the data suggests that the Asian community faces no disparity in COVID-19 cases or COVID-19 related deaths.

Figure 5.1



**Boxplot for Asian Population, Cases, and Deaths**

Figure 5.2



**Regression for Asian Population on the Number of Asian COVID-19 Cases**

Figure 5.3



**Regression for Asian Population on the Number of Asian COVID-19 Related Deaths**
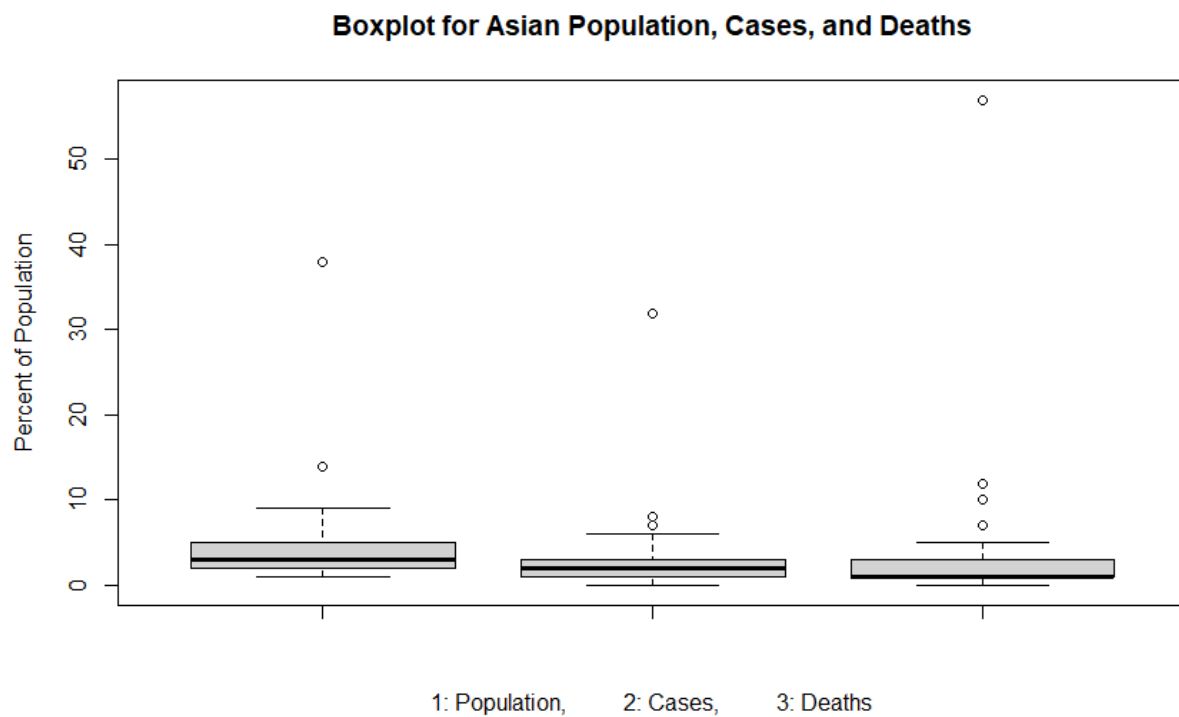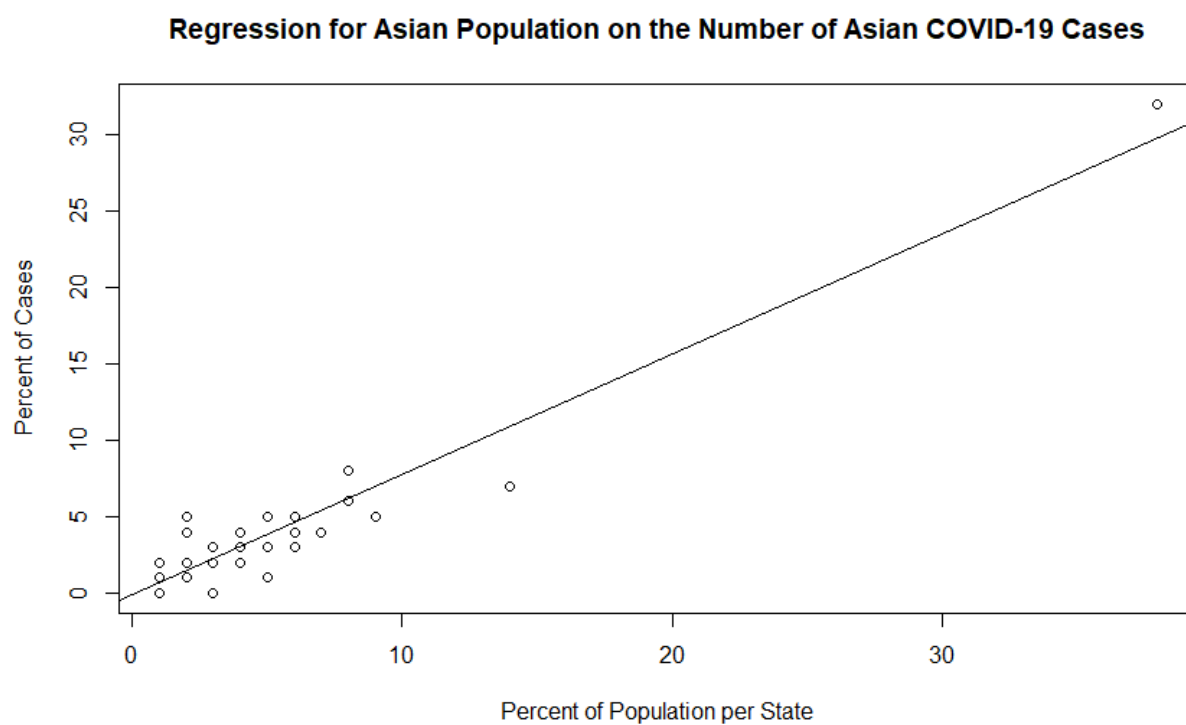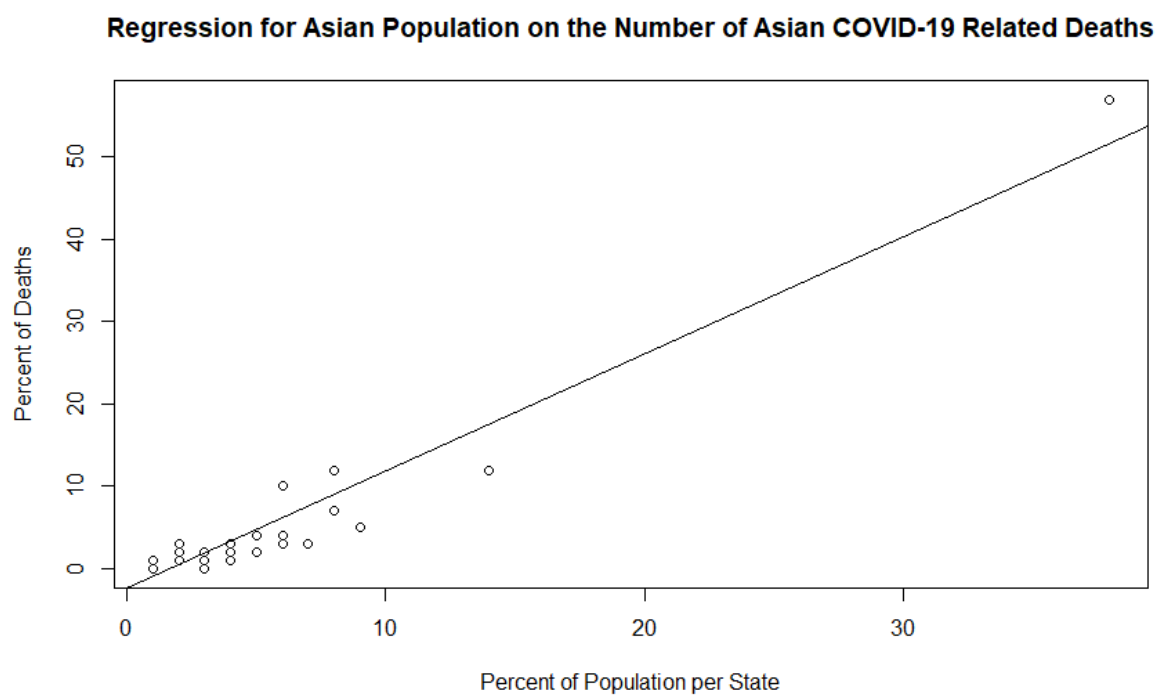
**Discussion**

Upon close analysis, it becomes clear that different minorities have suffered from COVID-19 in different ways. In the linear regression model below, red represents the white/Caucasian community, purple represents the black/African American community, orange represents the Hispanic/Latino community, and green represents the Asian community. With the multiple regression lines representing each group, it remained true that the white community experienced the least number of cases and deaths per percent population. However, it revealed how minorities compare to each other as well; for example, the Asian community had the lowest slope in COVID-19 cases, but it had the highest in COVID-19 related deaths. The Hispanic community, on the other hand, had the highest slope in COVID-19 cases, but had the second-lowest slope for COVID-19 related deaths. The black and African American community had relatively consistent slopes for both and remained second highest in both cases. Overall, the data suggests that some minorities experience greater disparities in cases, but not in deaths; some minorities experience greater disparities in deaths, but not in cases, and some minorities consistently face disparities. By understanding which community struggles with what, we can better understand how to fix the struggles they face.
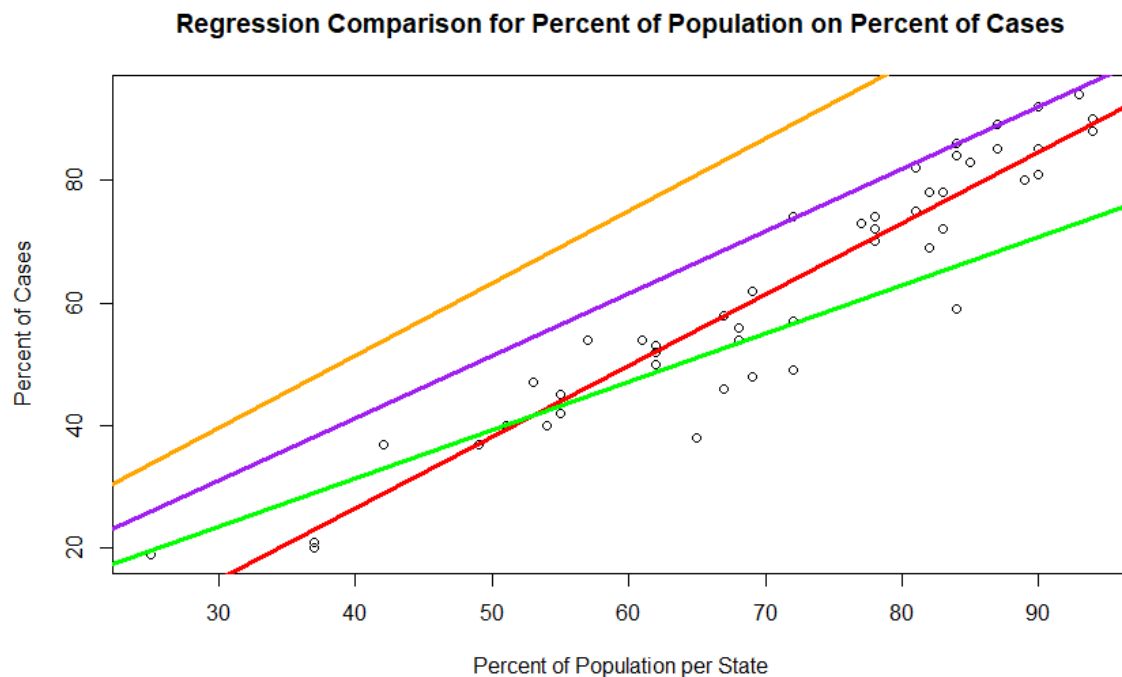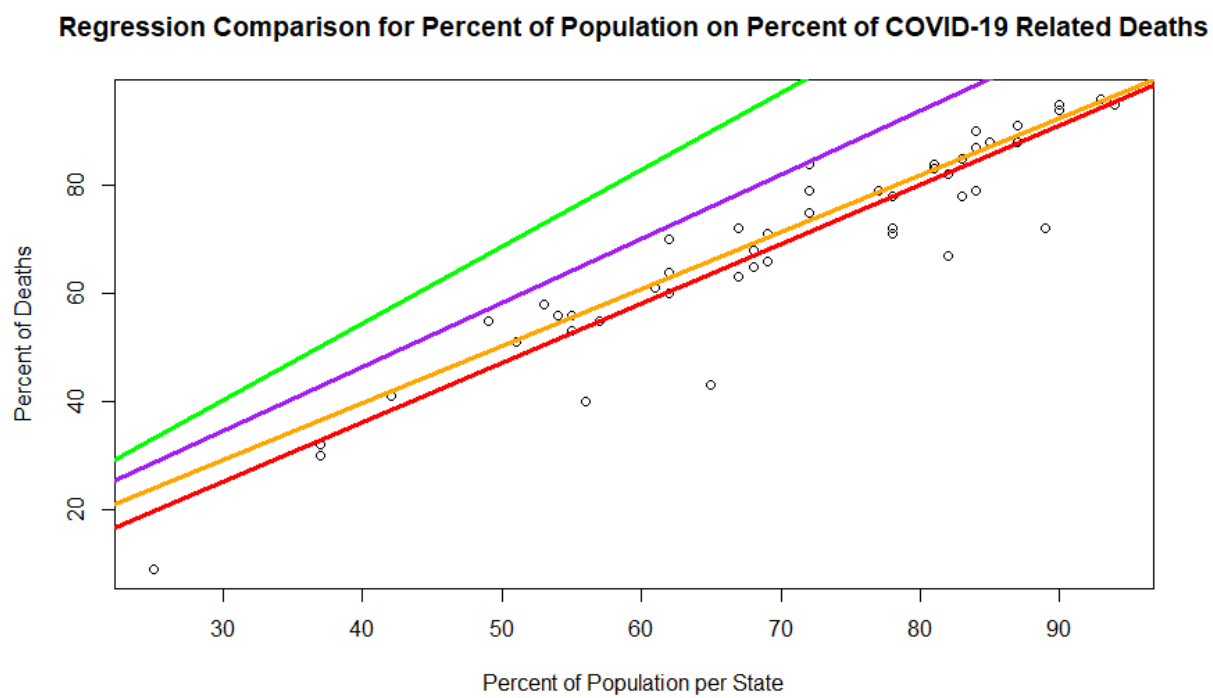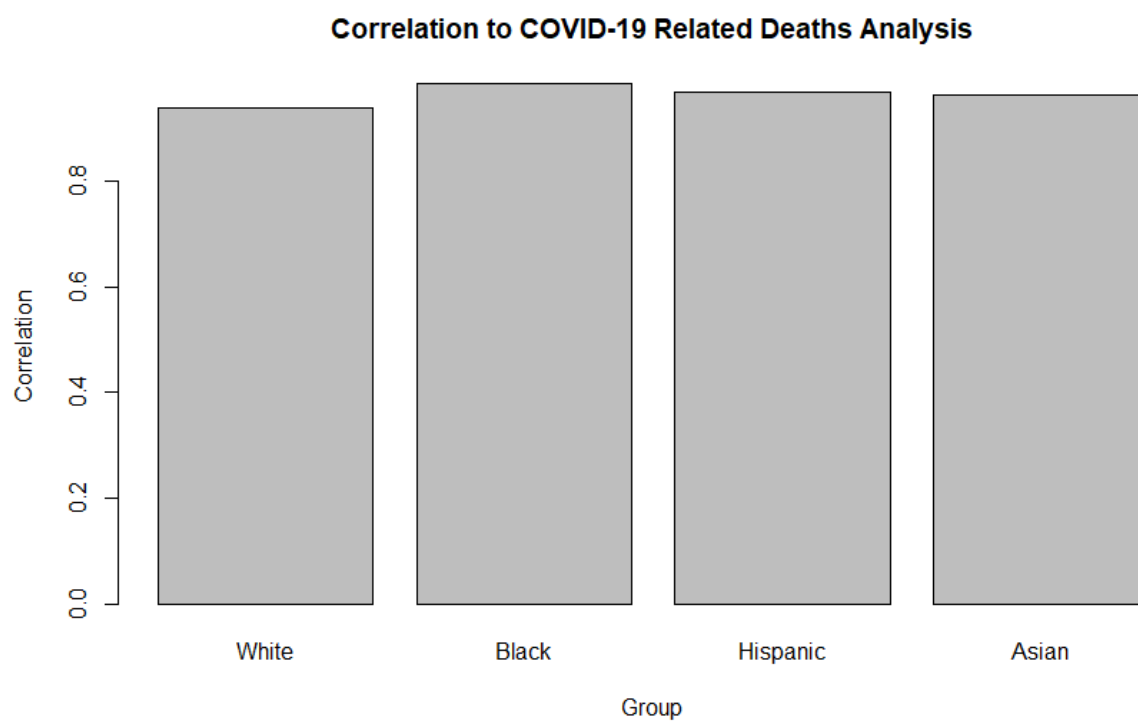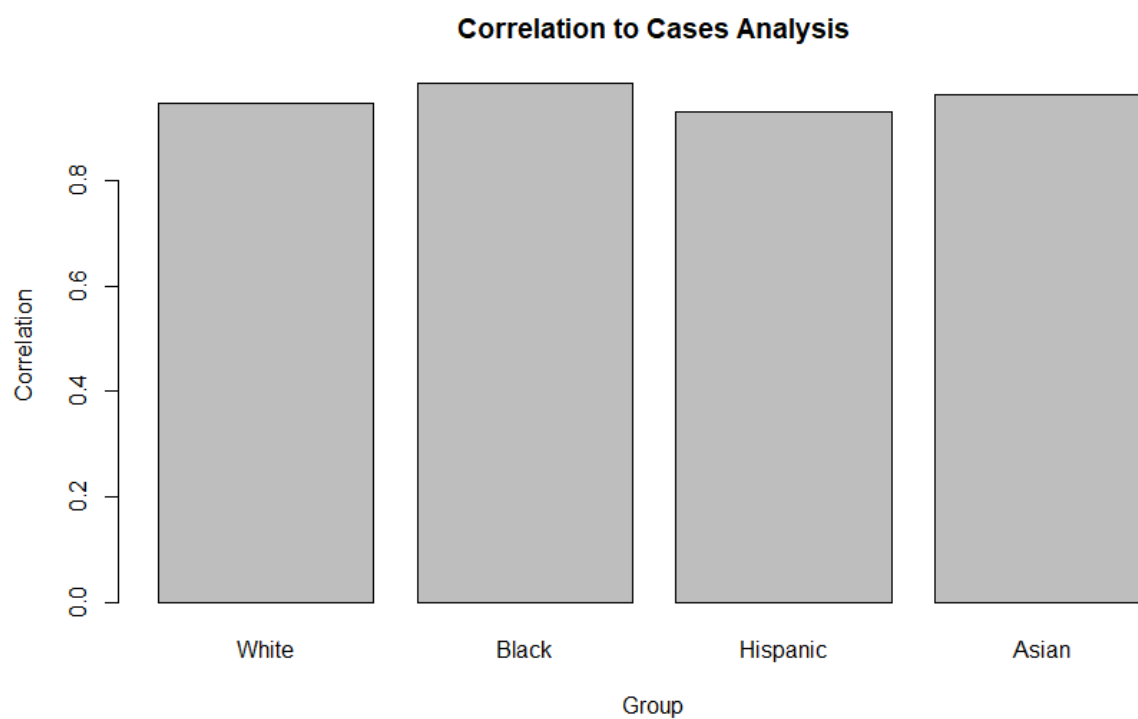
Figure 6.1

Figure 6.2

**Regression Comparison for Percent of Population on Percent of COVID-19 Related Deaths**
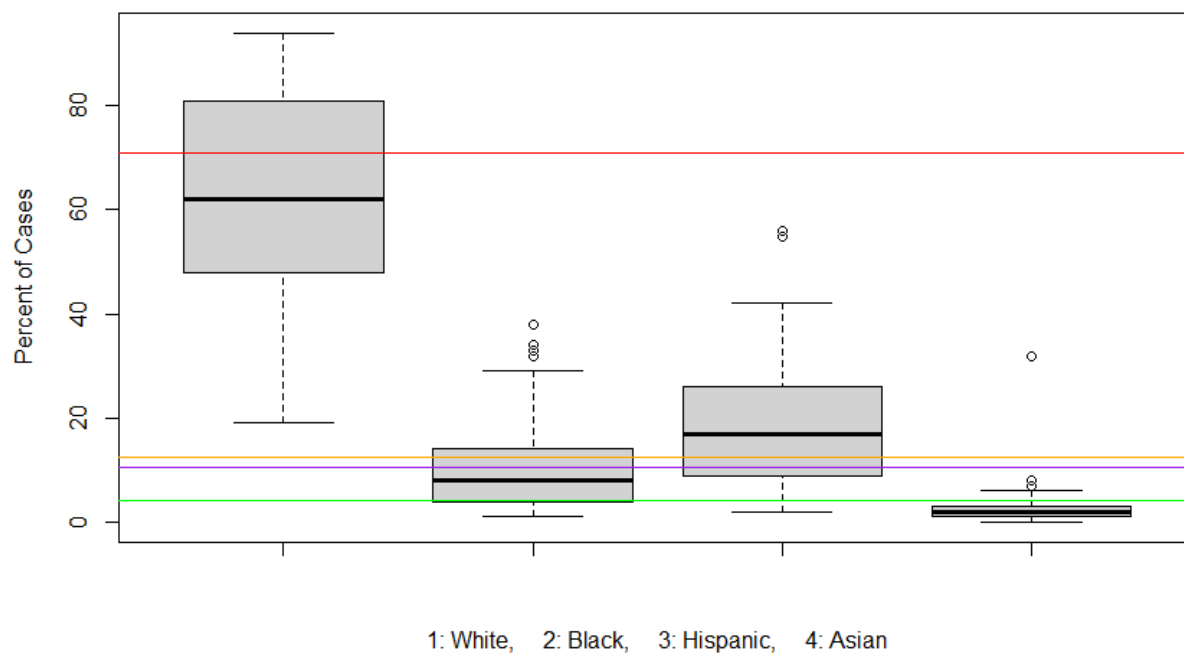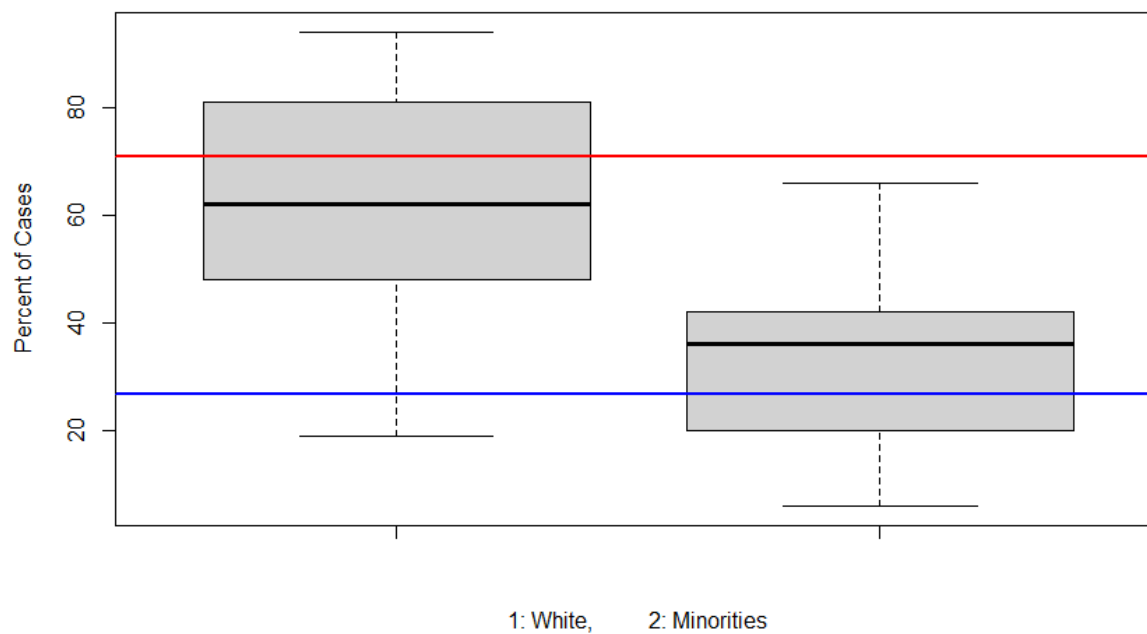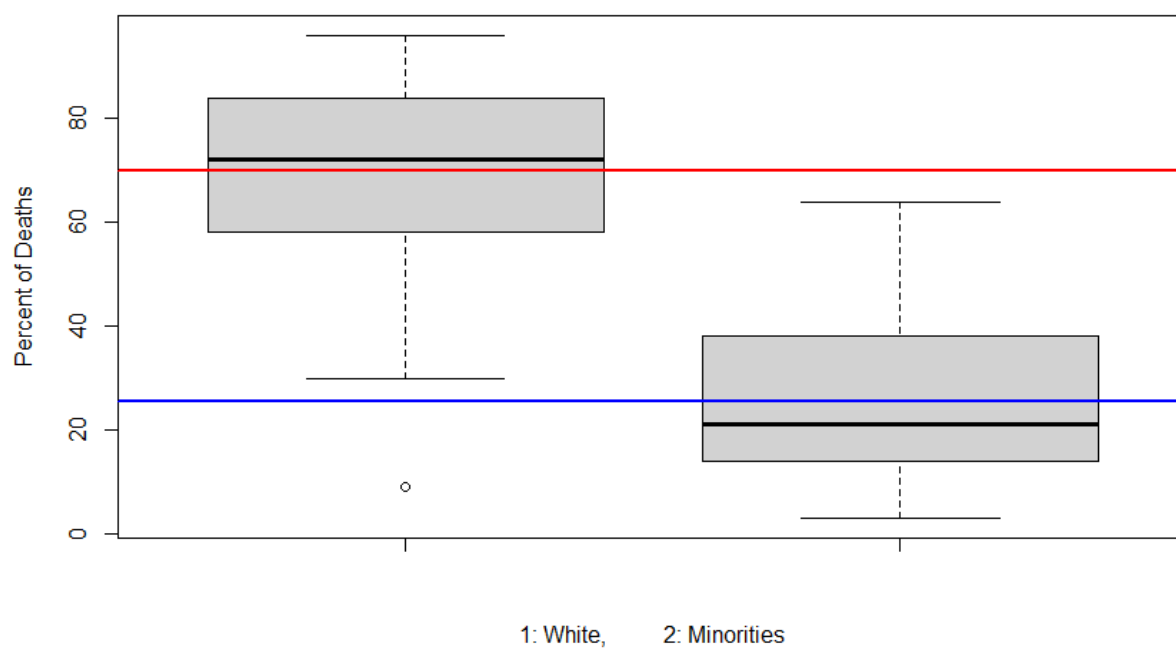
**References**

Alvarado, C., Umbelino, G., & Minnes, M. (2018). The persistent effect of pre-college computing

experience on college CS course grades. *ACM Inroads, 9*(2), 58–64.

https://doi.org/10.1145/3210551

**Appendix**

Additional Graphs:

**Correlation to Cases Analysis**



**Correlation to COVID-19 Related Deaths Analysis**

**Boxplot on Percent of Cases per Group**



1: White,    2: Black,    3: Hispanic,    4: Asian

**Boxplot on Percent of Cases per Group**



1: White,        2: Minorities

**Boxplot on Percent of COVID-19 Related Deaths per Group**



1: White,    2: Black,    3: Hispanic,    4: Asian

**Boxplot on Percent of COVID-19 Related Deaths per Group**



1: White,        2: Minorities
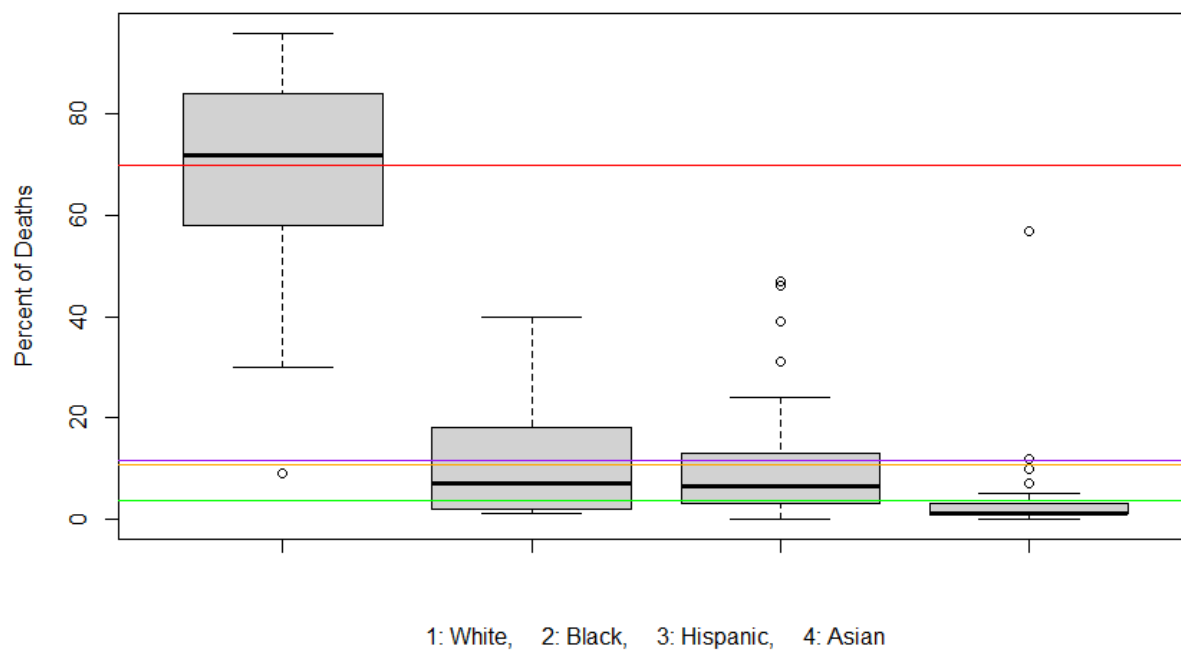
## Boxplot for White Population, Cases, and Deaths



1: Population,     2: Cases,     3: Deaths

## Boxplot for Black Population, Cases, and Deaths
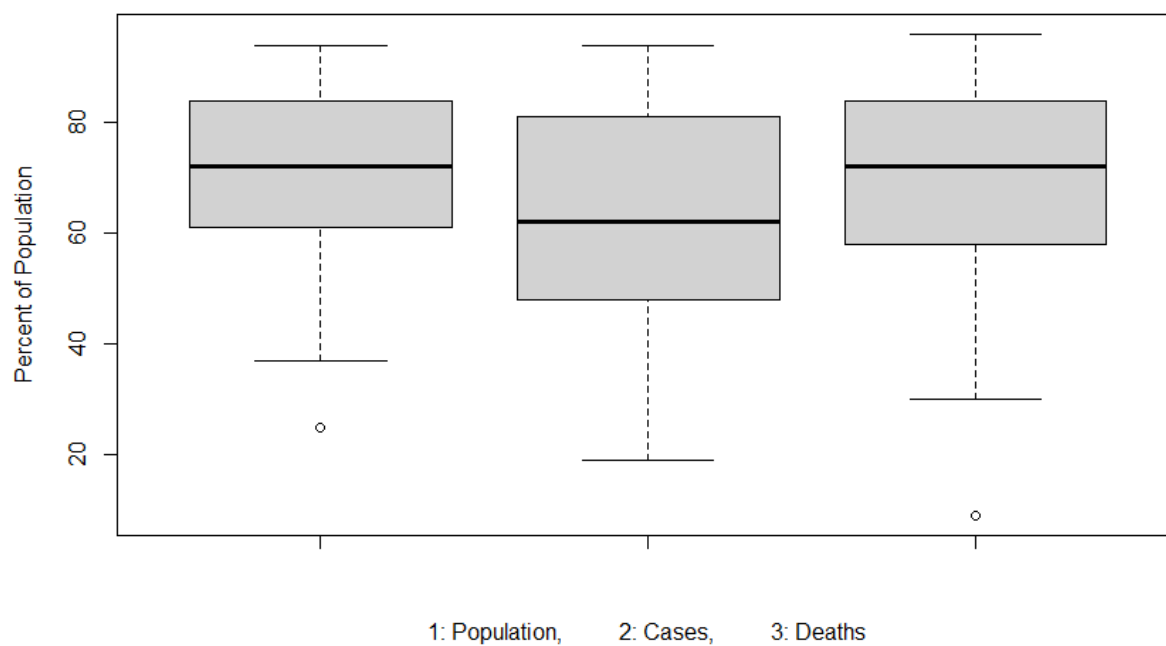


1: Population,     2: Cases,     3: Deaths

## Boxplot for Hispanic Population, Cases, and Deaths
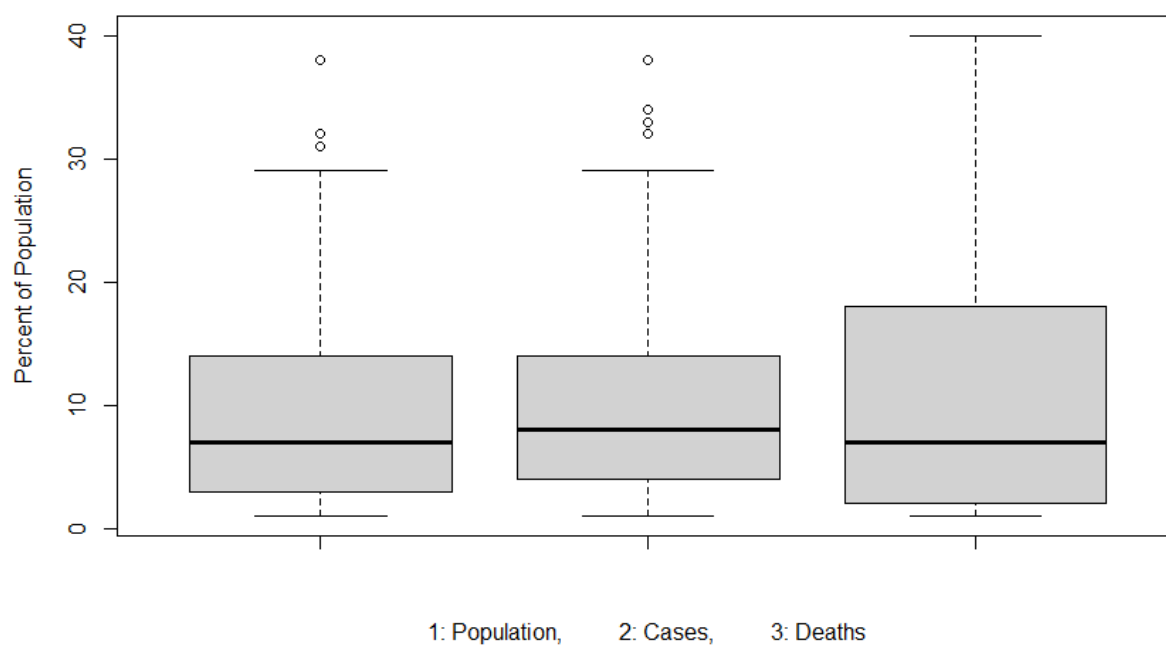


1: Population,     2: Cases,     3: Deaths

## Boxplot for Asian Population, Cases, and Deaths
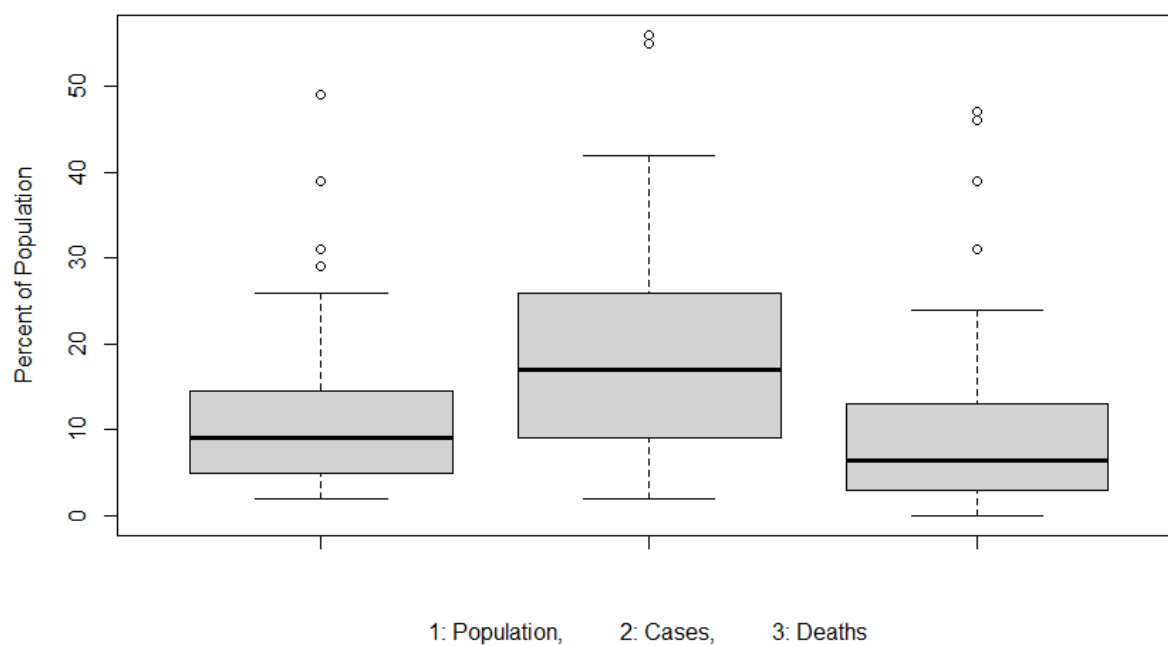


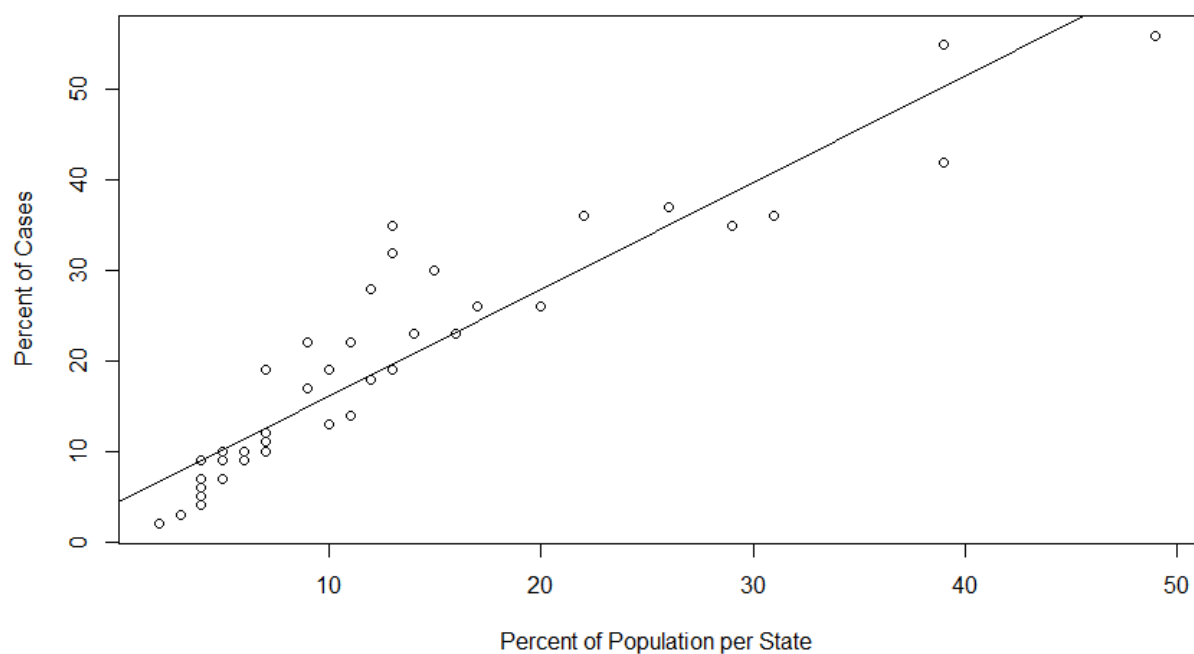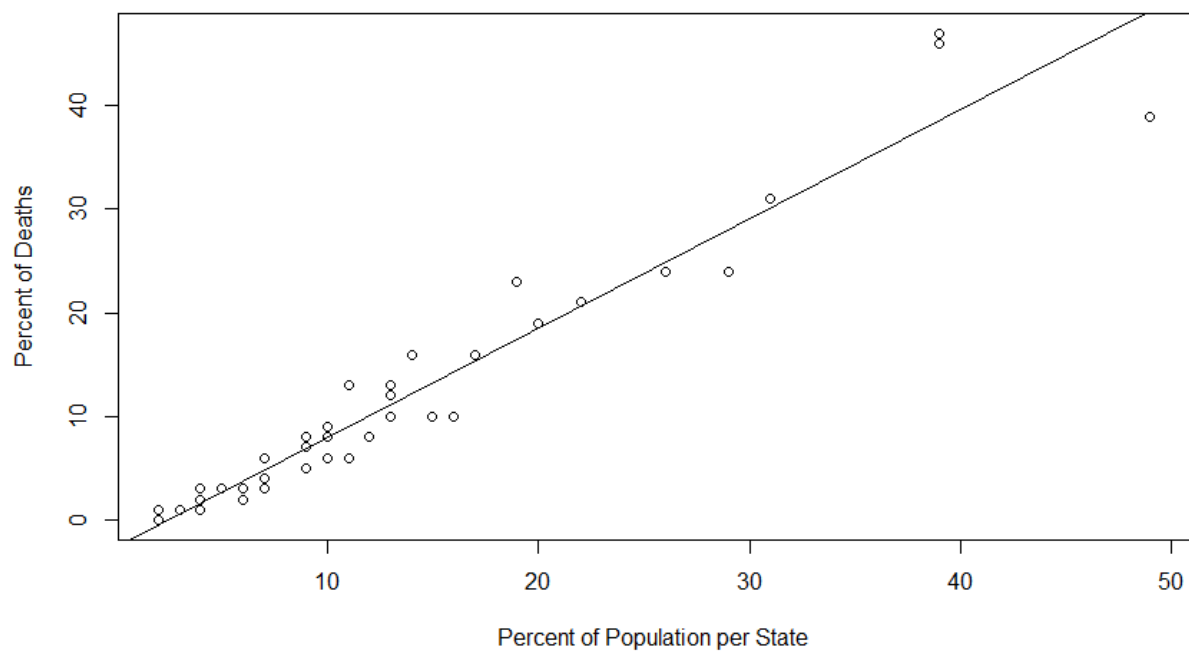1: Population,     2: Cases,     3: Deaths
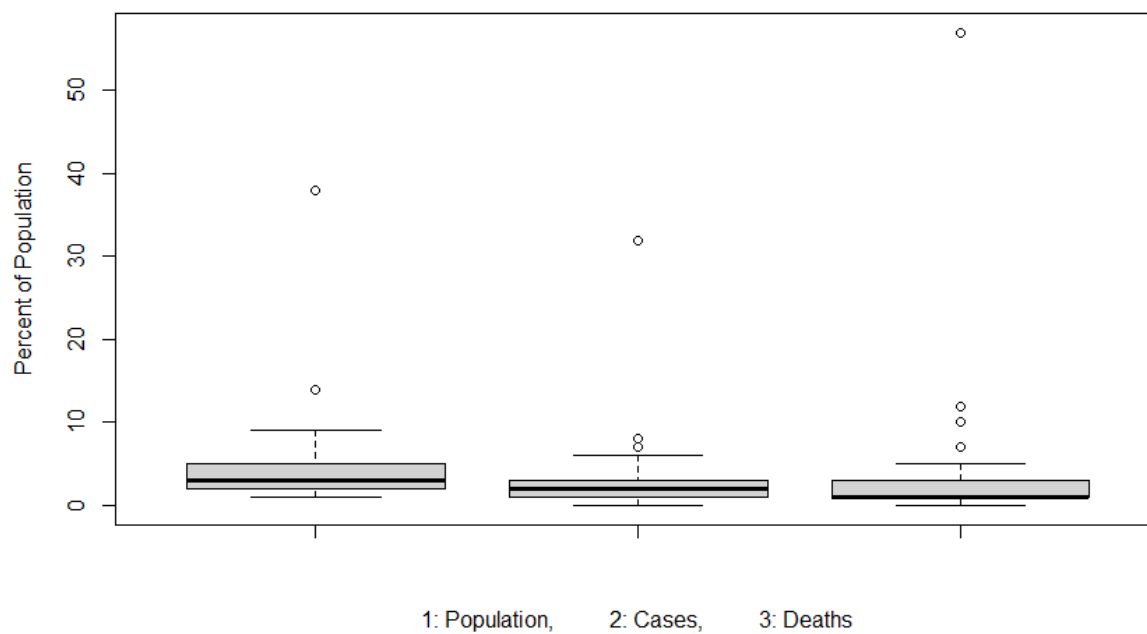
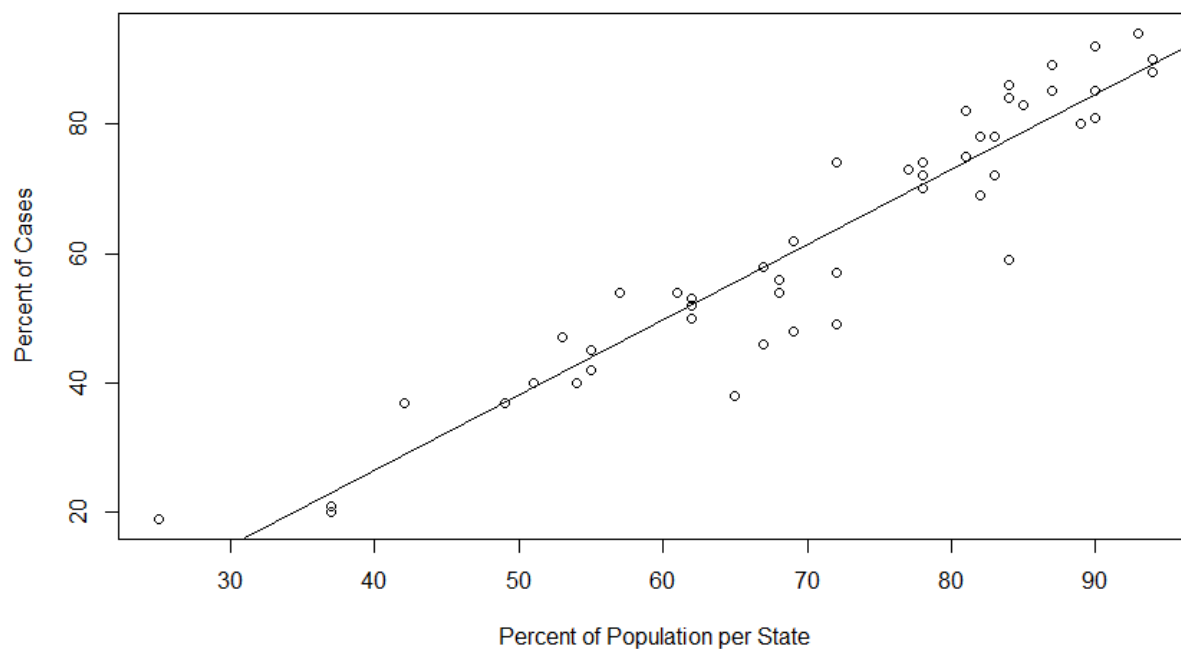**Regression for Asian Population on the Number of Asian COVID-19 Cases**



**Regression for Asian Population on the Number of Asian COVID-19 Related Deaths**

**Regression for Black Population on the Number of Black COVID-19 Cases**



Percent of Population per State

**Regression for Black Population on the Number of Black COVID-19 Related Deaths**



Percent of Population per State
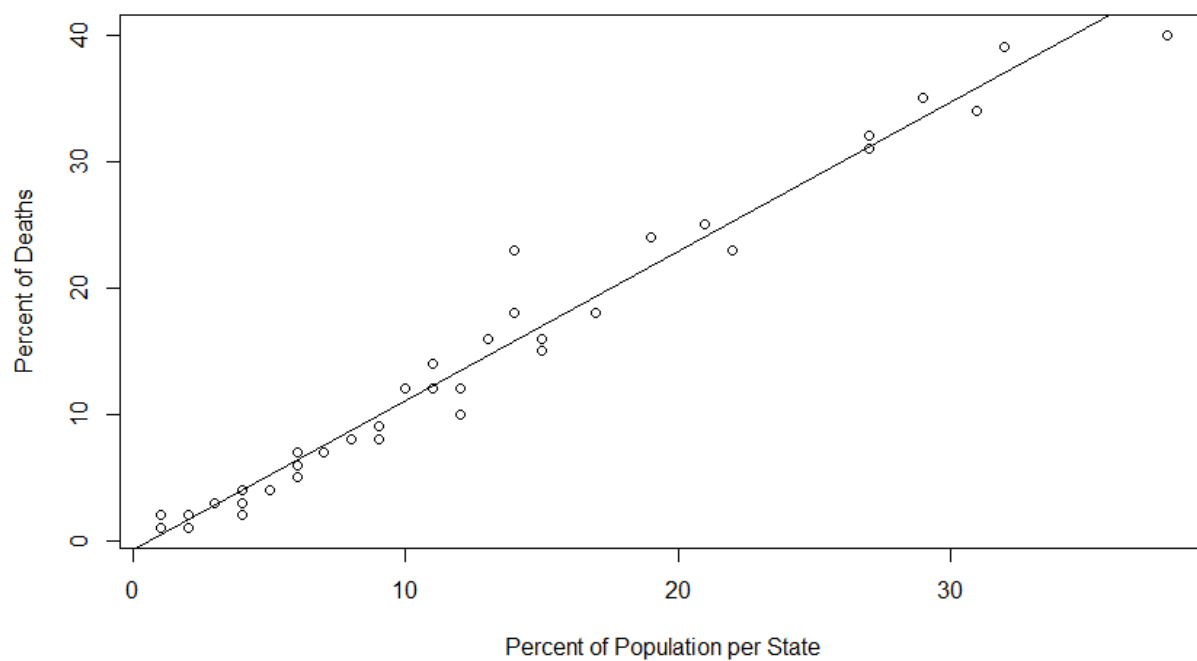
**Regression for Hispanic Population on the Number of Hispanic COVID-19 Cases**



**Regression for Hispanic Population on the Number of Hispanic COVID-19 Related Deaths**

**Regression for White Population on the Number of White COVID-19 Cases**



Percent of Population per State

**Regression for White Population on the Number of White COVID-19 Related Deaths**



Percent of Population per State