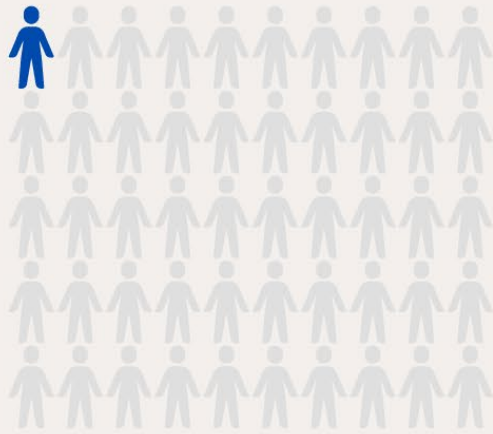# COVID-19 DESCRIPTIVE ANALYSIS

MINING THE DATASET OF COVID OUTBREAK AND POTENTIAL FEATURES IN THE US

BY
EZGI POLAT

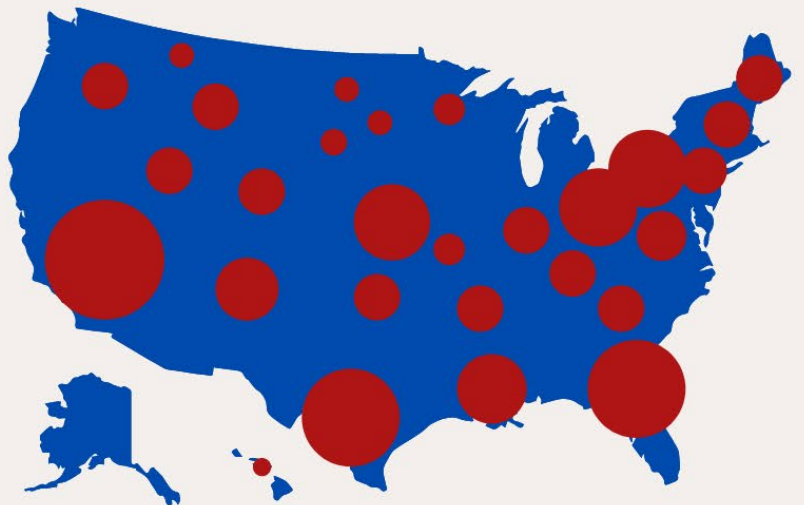# COVID-19:
## Coronavirus Disease

## 1.75%

of the people infected by COVID-19 in the USA lost their lives. Meaning that 2 out of every 100 infected Americans did not get the chance to survive. The numbers might make it look small until we realize that each number represents someone's loved one.

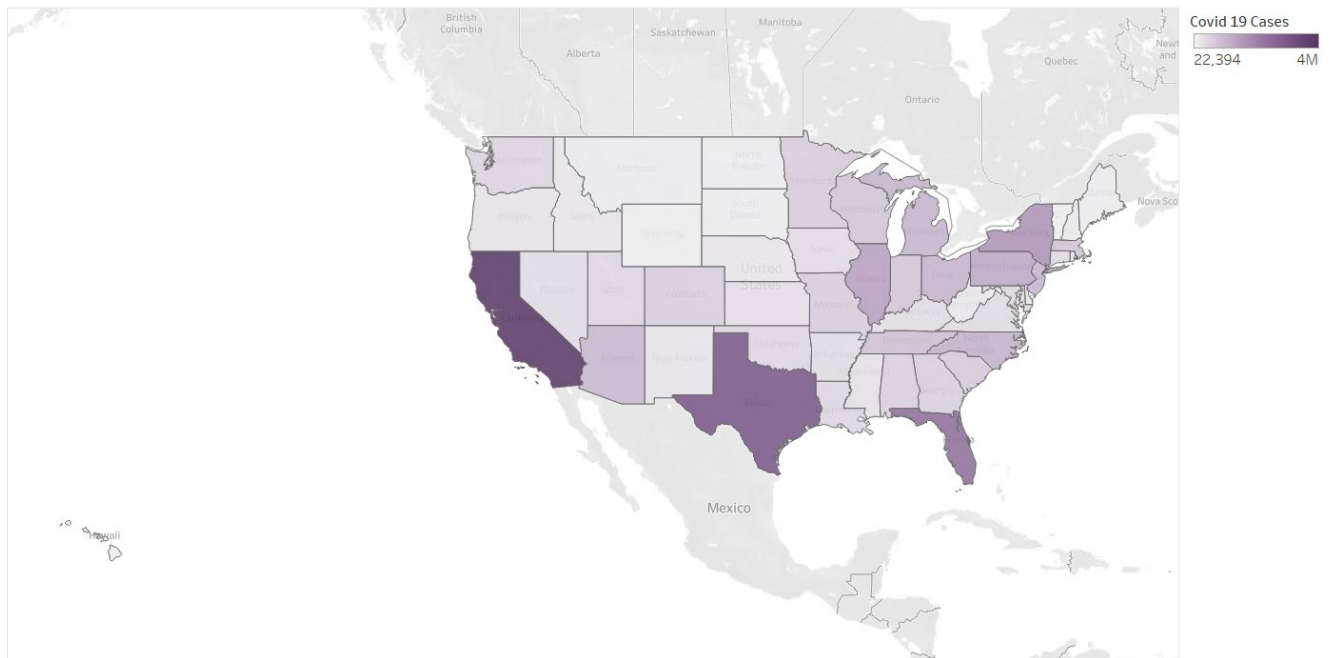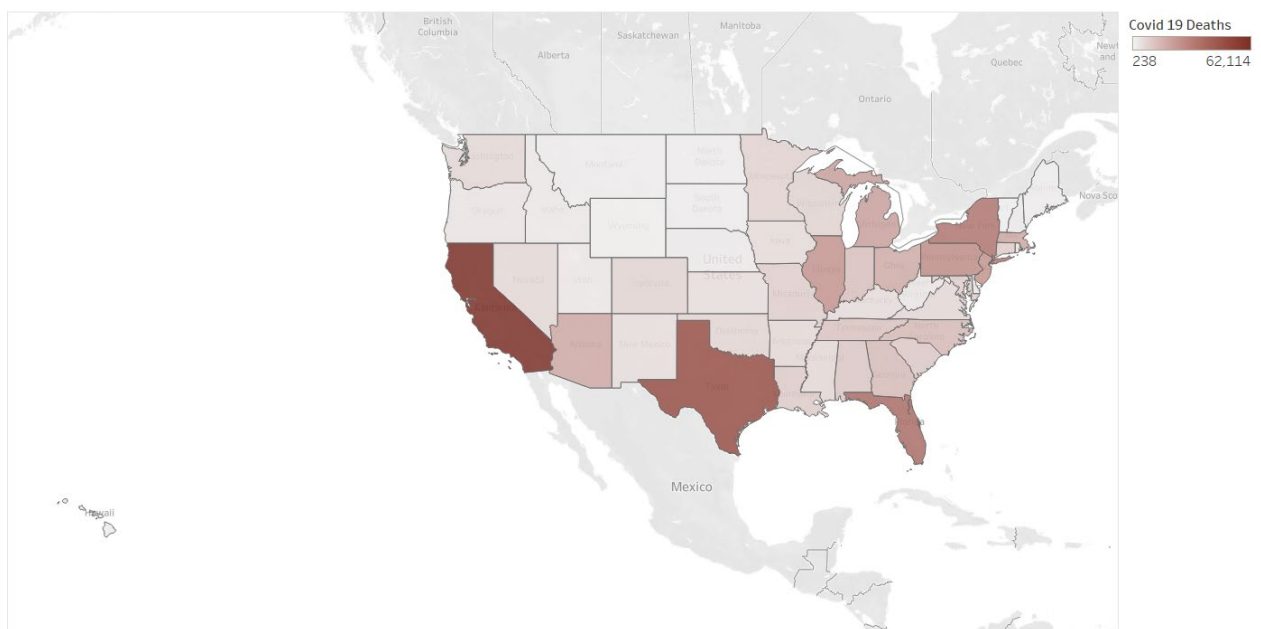**TOTAL CASES**

## 29,425,620

**TOTAL DEATHS**

## 517,539

## 156

in every 100K Americans lost their lives due to COVID-19. Making the United States one of the countries with the **highest mortality rates.**

# Geographical



**COVID-19 Cases Among the States**



**COVID-19 Deaths Among the States**

**The States with the Highest Number of Cases**

| State | Cases |
|---|---|
| California | 3 682 688 |
| Texas | 2 845 636 |
| Florida | 2 246 950 |
| New York | 1 506 657 |
| Illinois | 1 293 122 |

**The States with the Highest Number of Deaths**

| State | Deaths |
|---|---|
| California | 62 114 |
| Texas | 47 784 |
| Florida | 35 865 |
| New York | 32 974 |
| Pennsylvania | 25 818 |

**The States with the Lowest Number of Cases**

| State | Cases |
|---|---|
| Vermont | 22 394 |
| Hawaii | 35 522 |
| Wyoming | 56 152 |
| Maine | 68 448 |
| North Dakota | 85 322 |

**The States with the Lowest Number of Deaths**

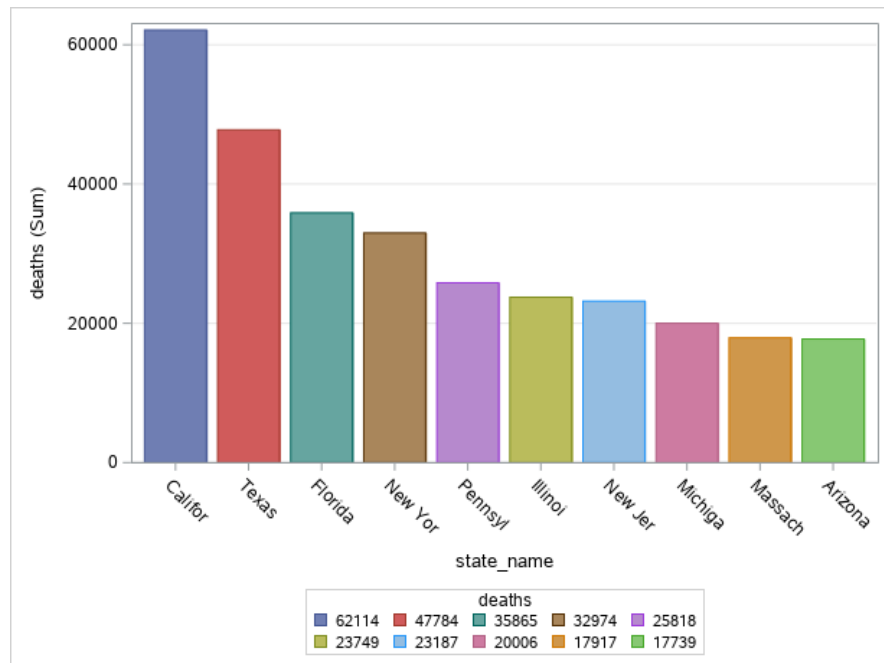| State | Deaths |
|---|---|
| Vermont | 238 |
| Hawaii | 502 |
| Wyoming | 662 |
| Maine | 843 |
| North Dakota | 1 011 |

These results show us that the first 5 states with the lowest number of COVID-19 cases are matching the first 5 states with the lowest number of COVID-19 deaths. This picture supports an association between the COVID-19 cases and COVID-19 deaths.

The case is almost the same for the states with the highest numbers as well. The first 4 states with the highest number of COVID-19 cases have a perfect match with the first 4 states with the highest number of COVID-19 deaths. This is also a strong support for the association between the COVID-19 cases and COVID-19 deaths but let's investigate the states with the highest and lowest numbers with bar charts as well and expand the samples to 10 states.
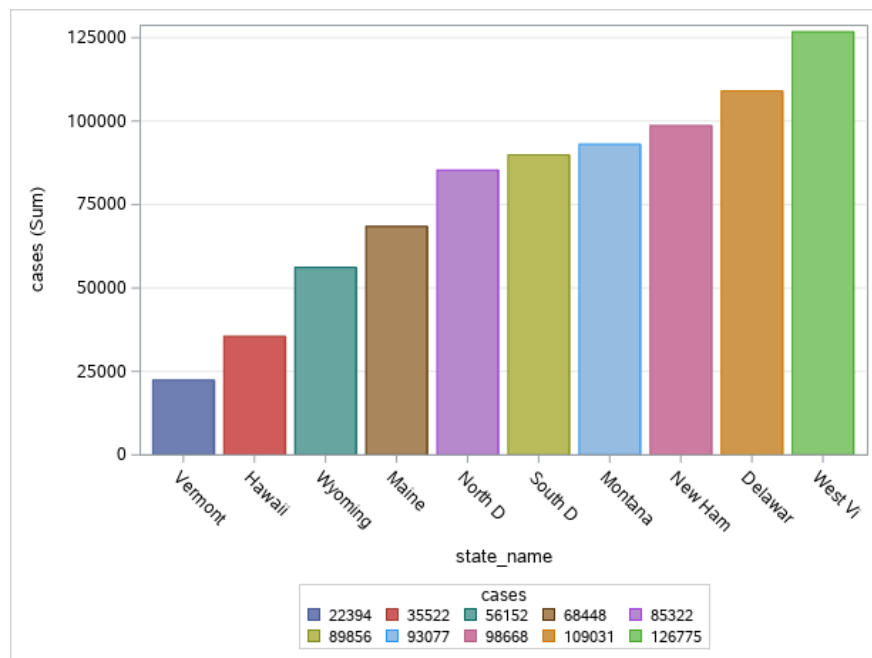
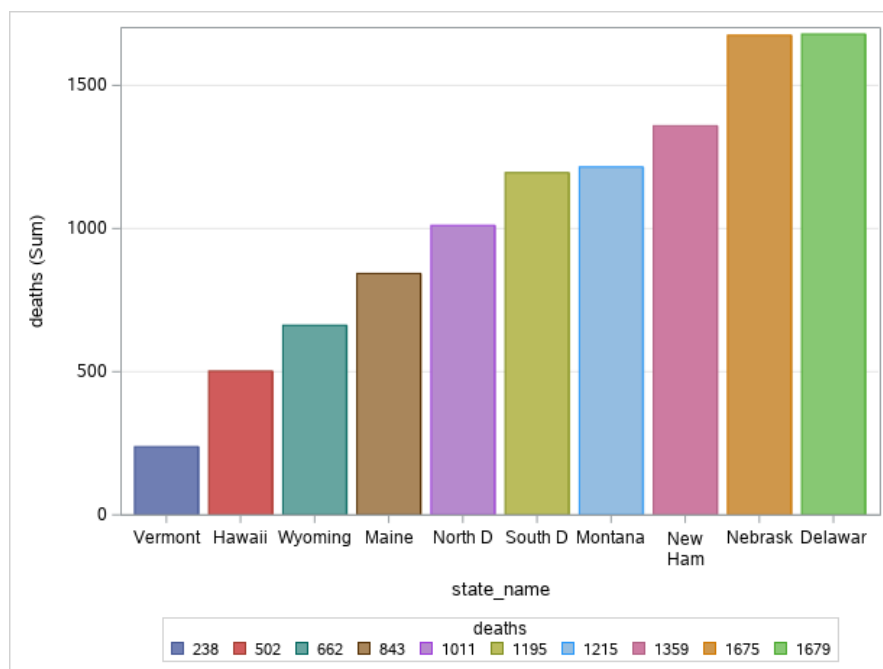# The States with the Highest Number of Cases
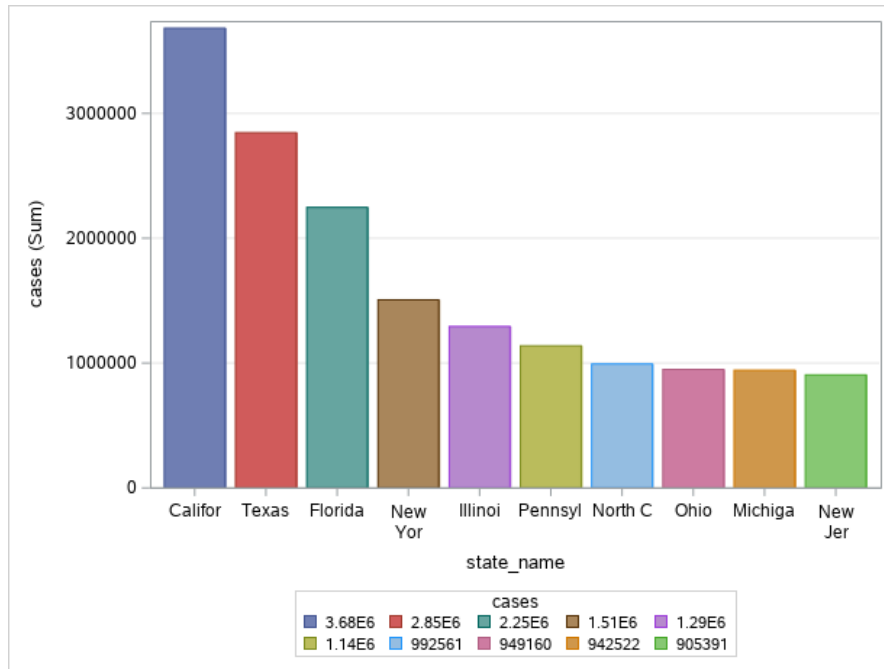


# The States with the Highest Number of Deaths

## The States with the Lowest Number of Cases



## The States with the Lowest Number of Deaths

We observed an **80%** match in both cases with the highest and lowest number of states. 8 states out of the first 10 states with the highest or lowest number of **cases** exist in the list of the highest or lowest number of **deaths**.

In the states with the lowest numbers table, even the ranking of these 8 states has a perfect match.

Up to this point, the situation provides strong support for the correlation coefficient of **0.98** between the number of COVID-19 **cases** and **deaths**.

## Understanding the Severity and Lethality of COVID-19

Although the number of cases and deaths are helpful to understand the COVID-19, we still need more metrics to understand how severe and deadly the virus is in the states of America. For that matter, we decided to use 2 more metrics; **Mortality Rate** and **Case Fatality Rate**.

Case Fatality Rate is vital because it helps us to understand how many of those with a confirmed diagnosis of COVID-19 die as a result of the disease. It is about how severe a disease is. It can measure the effectiveness of the treatment, quality of the health care or even effectiveness of COVID precautions in a state.
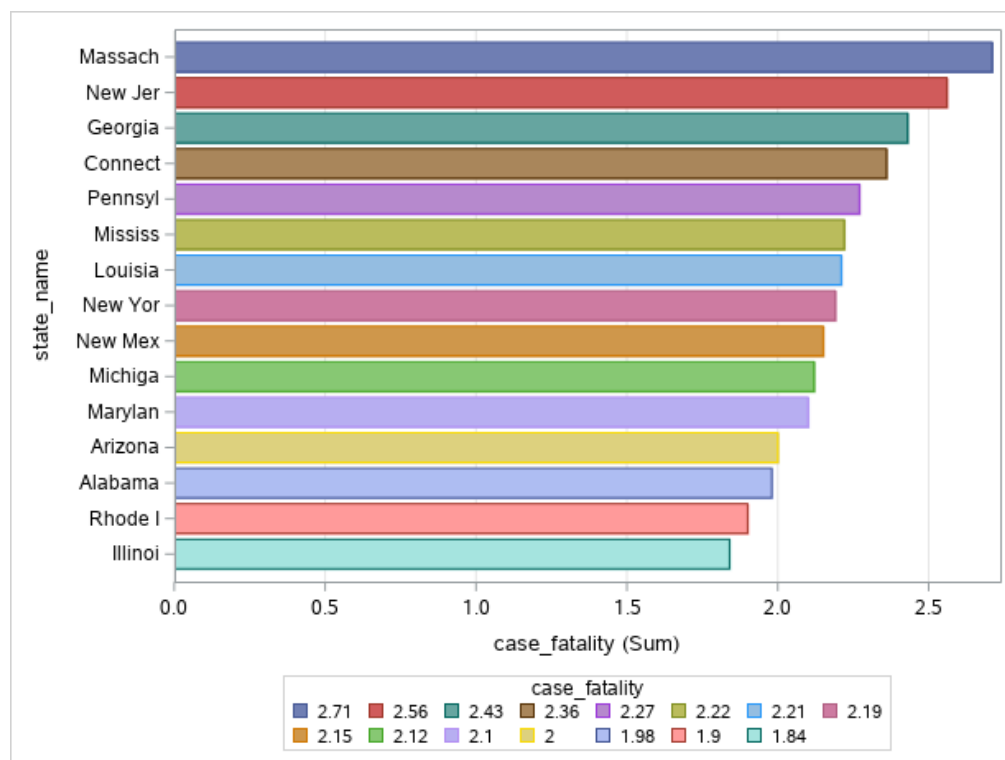
The states with highest number of cases & deaths such as California, Texas, Florida are not in this list. This provides a proof that the number of cases & deaths should not be the primary metrics to interpret the severity of the virus. While California, Texas and Florida have the highest numbers of cases & deaths, the severeness of COVID-19 is not as high as in the other states.

However, this does not mean the case fatality rate and the case/death numbers are not related to each other, the CFR is a metric calculated by these numbers after all, but it is just important to understand the difference between case/death numbers and the severity levels.

For example, **the states with the lowest case fatality rates** are mostly matching the states with the lowest numbers of cases & deaths. Although, the first 3 states are new and different; the rest of the states can be seen in the list of the states with the lowest case & death numbers as well.
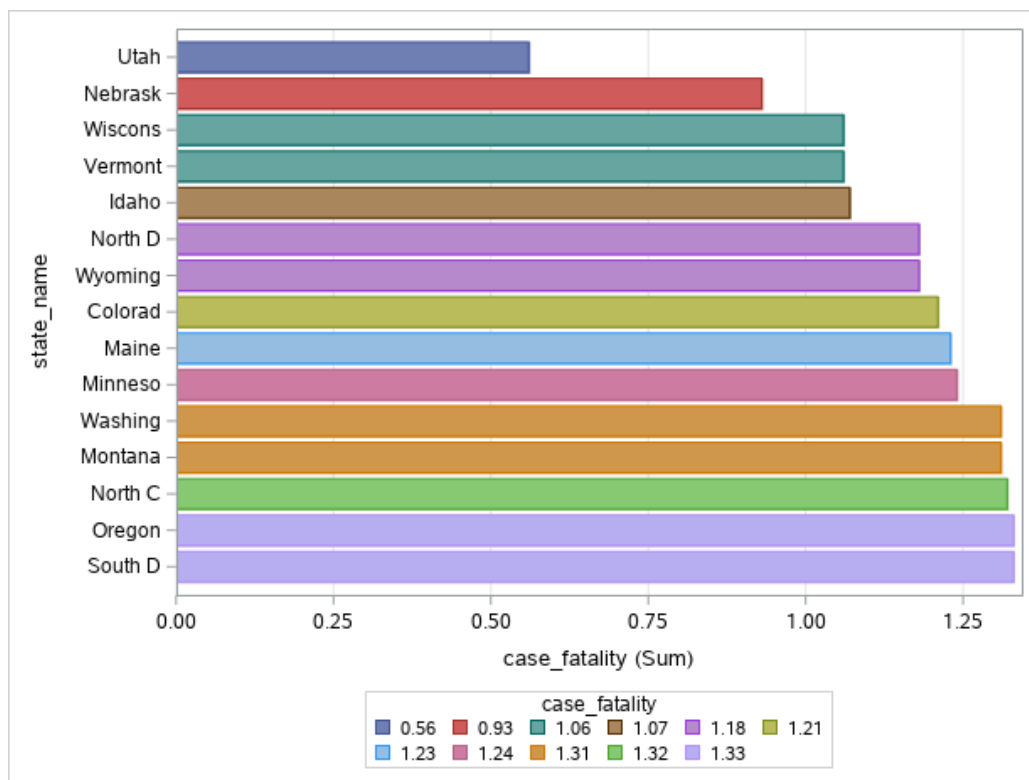


As much as the CFR demonstrates a better picture, it's important to differentiate case fatality rate from mortality rate. A **mortality rate (MR)** is a measure of the frequency of occurrence of death in a defined population during a specified interval. Mortality statistics provide a valuable measure for assessing community health status. The importance of mortality statistics derives

both from the significance of death in an individual's life as well as their potential to improve the public's health when used to systematically assess and monitor the health status of a whole community. Within the realm of public health, mortality statistics are often used as a cornerstone in formulating health plans and policies to prevent or reduce premature mortality and improve our quality of life. However, in our report, it is used as a metric to understand the **lethality** (or deadliness) of COVID-19 in states of America and demonstrated with the formula below.

$$\frac{\text{Deaths occurring during a given period}}{\text{Size of the population among which the deaths occurred}} * 10^n$$

We had the total population data of each county, but we needed to manipulate our data into getting the population data for each state. After reaching the population numbers for each state, we started calculating the mortality rate of each state with the formula given above.
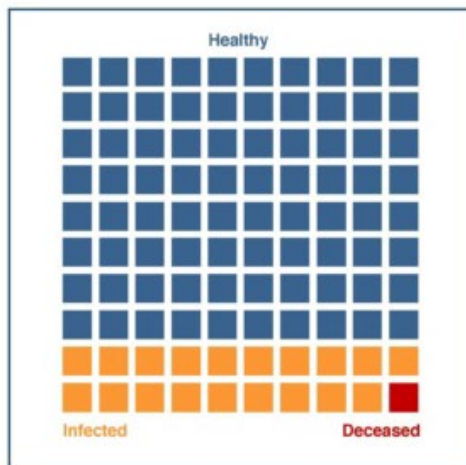
### The States with the Highest Mortality Rates
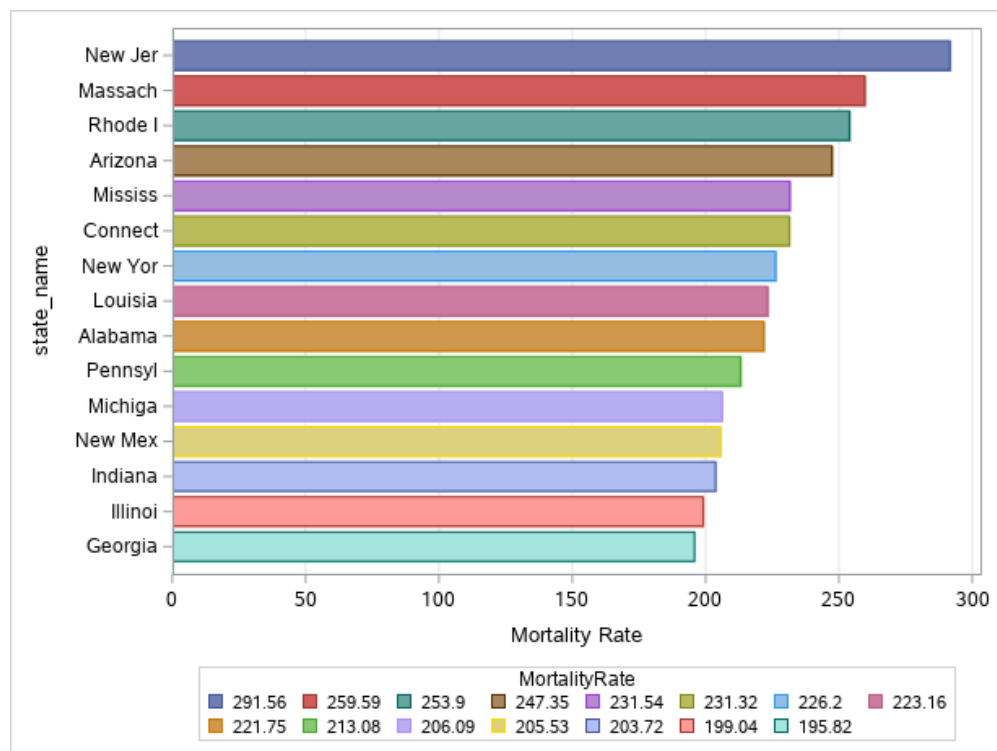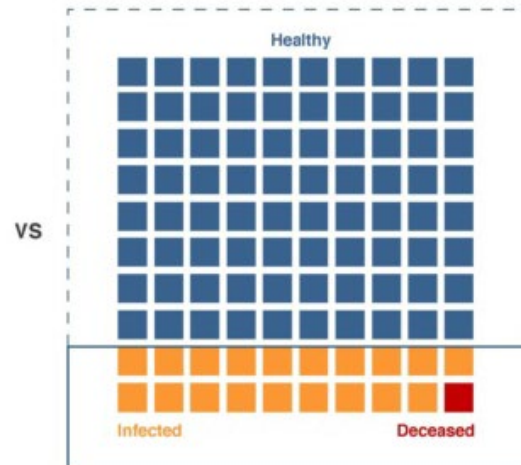
## The States with the Lowest Mortality Rates
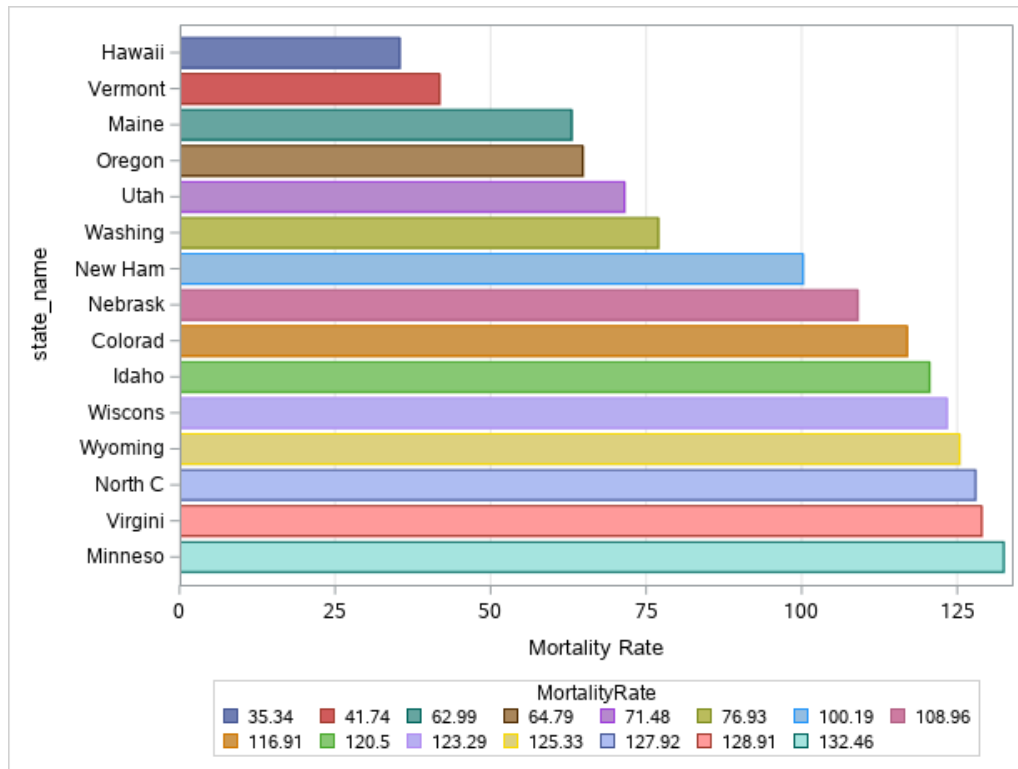


The mortality rate refers to the proportion of the population that dies due to a disease. Mortality rate differs from the case fatality rate, because not everyone in a population becomes infected with a disease.

Because it is challenging to measure both of these values with complete accuracy, it is crucial to consider the bias in the results. The number of patients with COVID-19 who have been diagnosed recently does not match the number of people who truly have it. And it's likely that we may not receive a reliable evaluation of that figure. Compared to the actual risk of dying from the disease, the present case fatality rate is probably higher. This is because those who have the most severe symptoms are most likely to get testing, which increases their risk of developing potentially fatal consequences.