COVID-19 Data Findings Report

The data supplied by this report can be found on Kaggle at this link:

[Covid-19 Case Surveillance Public Use Dataset | Kaggle]

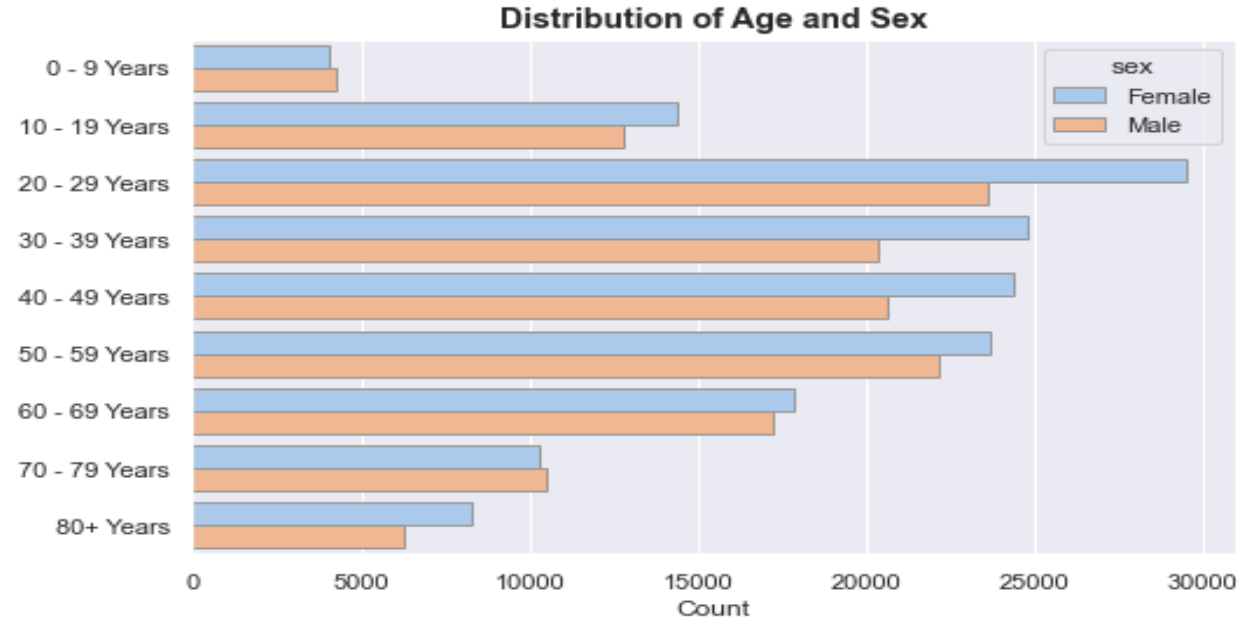
More information can be found at this link from the CDC website:

https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf

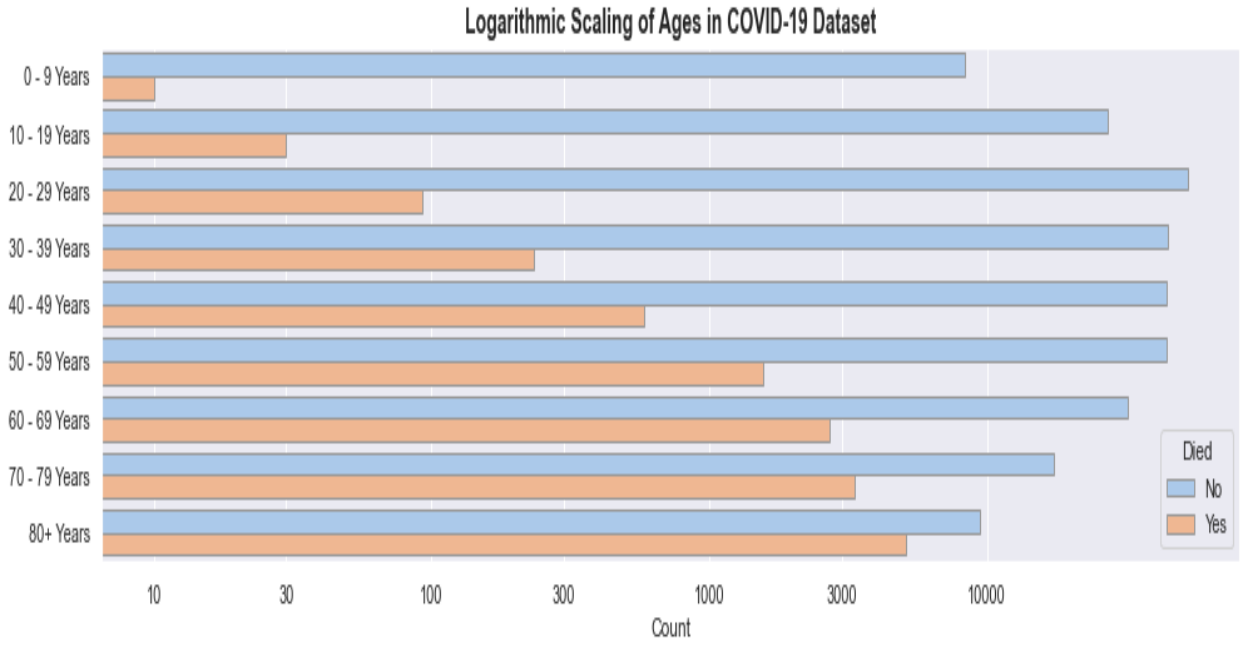
While the Kaggle site gives details on how the dataset is structured, this report will detail the dataset that was wrangled removing all missing, unknown and null values. This means every row on this dataset had information for every column within the scope of the dataset. This resulted in a more condensed dataset that contains 294,946 rows. One additional column was added to this dataset, which includes death with preexisting medical condition(s) when person was tested positive for COVID-19.

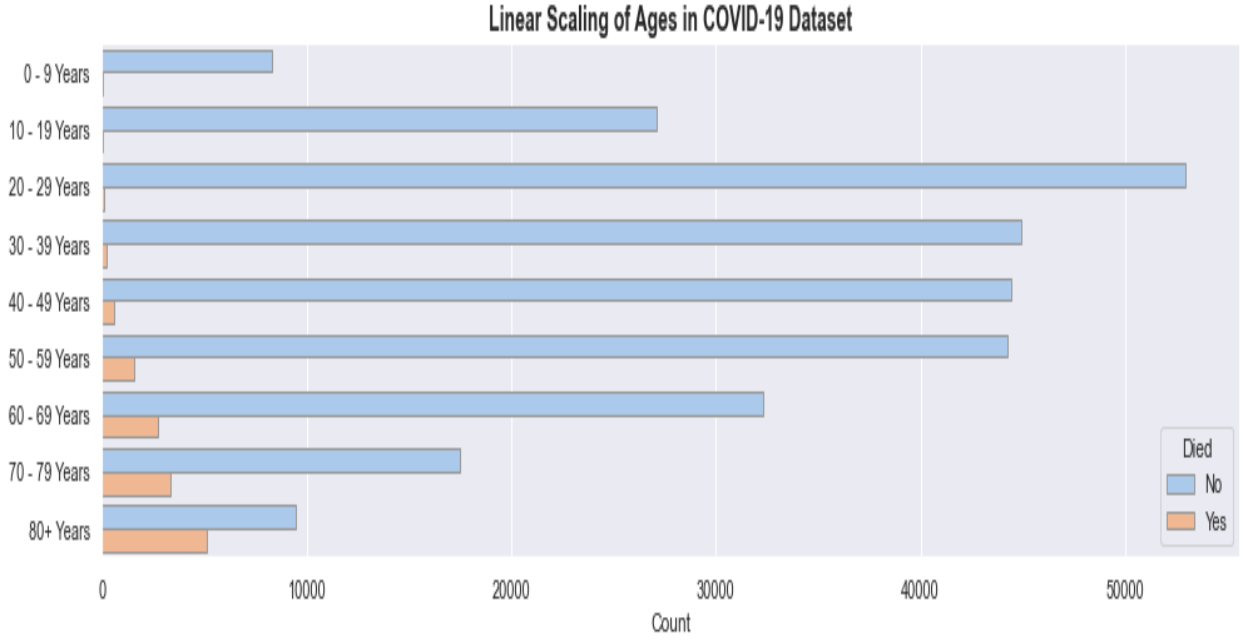
The dataset dates from 1/1/2020 to 11/19/2020, so it covers the majority of 2020 before and during when COVID-19 was stated as a pandemic. The focus of the analysis done is to determine the death rate from COVID-19 based on race, age, and if person had a preexisting condition(s). The following visuals will illustrate the distribution of each of these demographics. The hypothesis is that most deaths "caused" by COVID-19 was not by the virus, itself, but possibly if someone had a preexisting medical condition or by the age of the person that contracted the virus.

The first visual will show the distribution of the 294,946 individuals that were in the dataset by sex and age, most falling in the range between 20 and 70 years old:

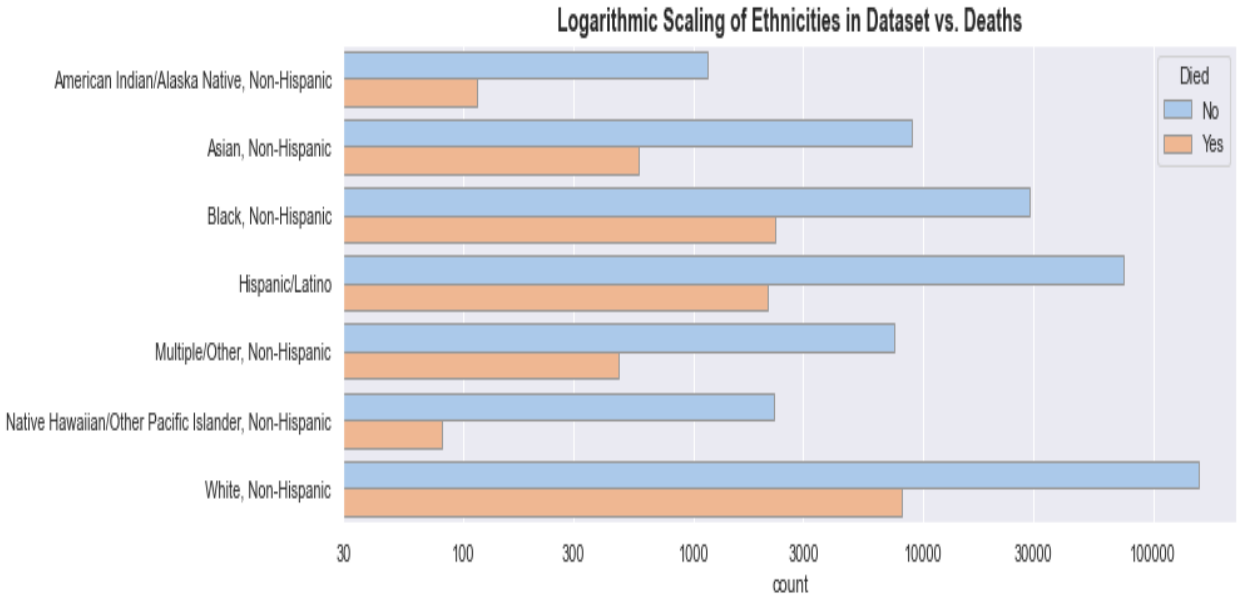


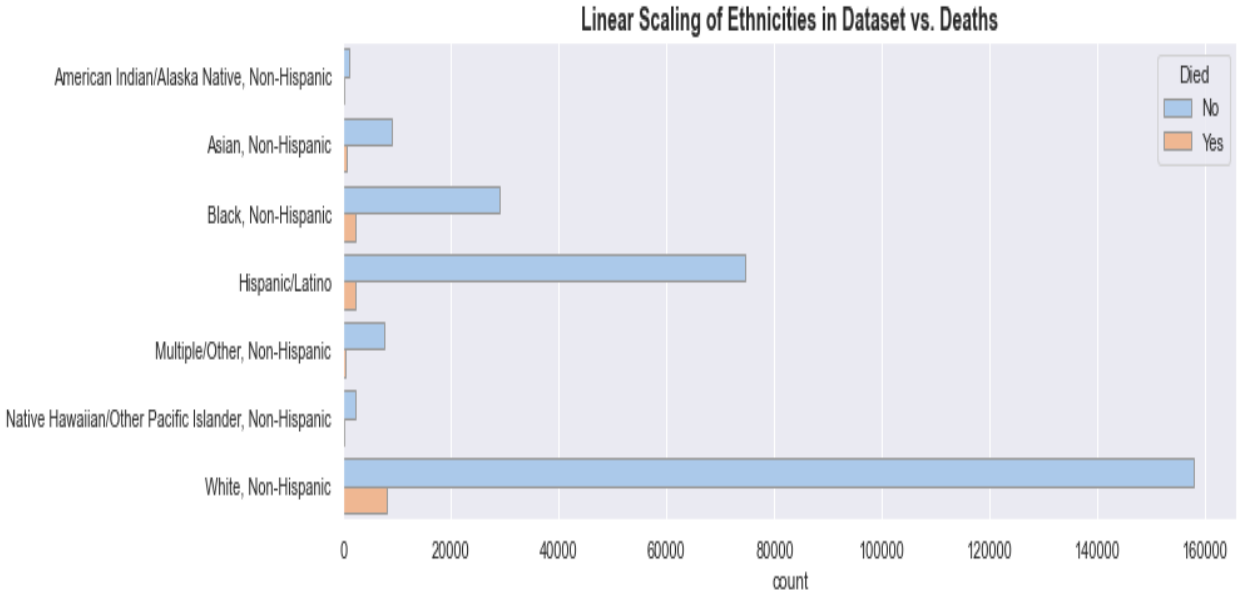
This next visual will compare the individuals by age group that survived vs. died from contracting the coronavirus both in a logarithmic and a linear scale. The logarithmic scale can be misleading; however, it helps illustrates the death rate in the younger age groups. The linear scale is a better representation of the actual proportion of survived vs died.





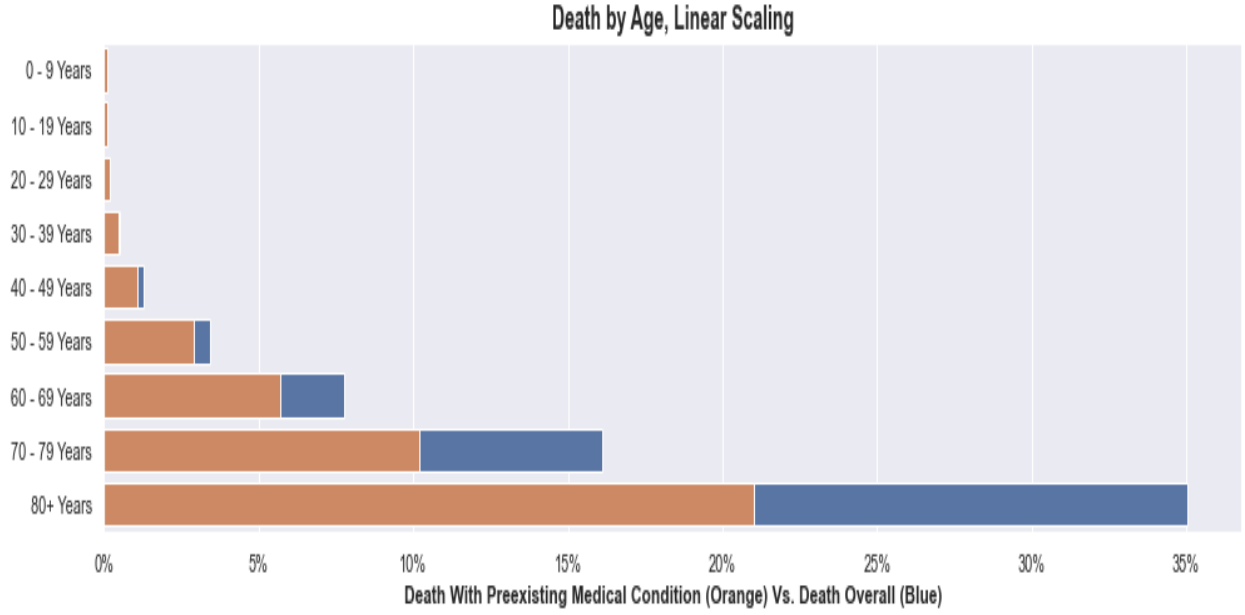
The next visual is distribution of ethnicities that survived or died from contracting COVID-19. Like the previous visual, it also has logarithmic and linear scaling. Logarithmic to highlight deaths by race and linear for actual proportion of deaths.

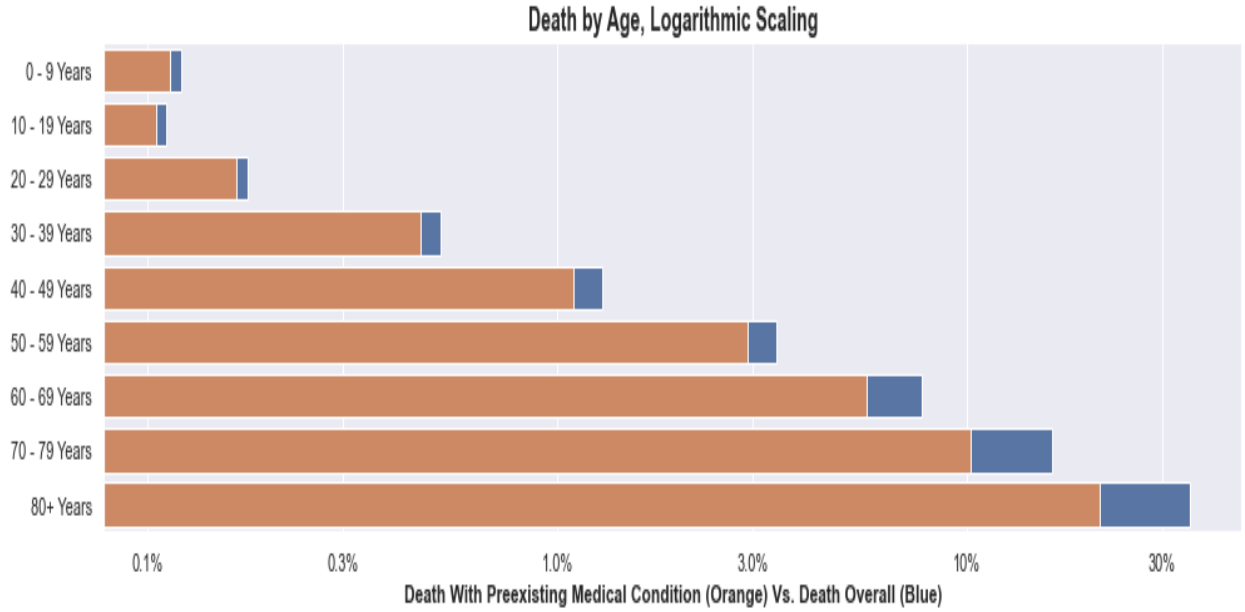




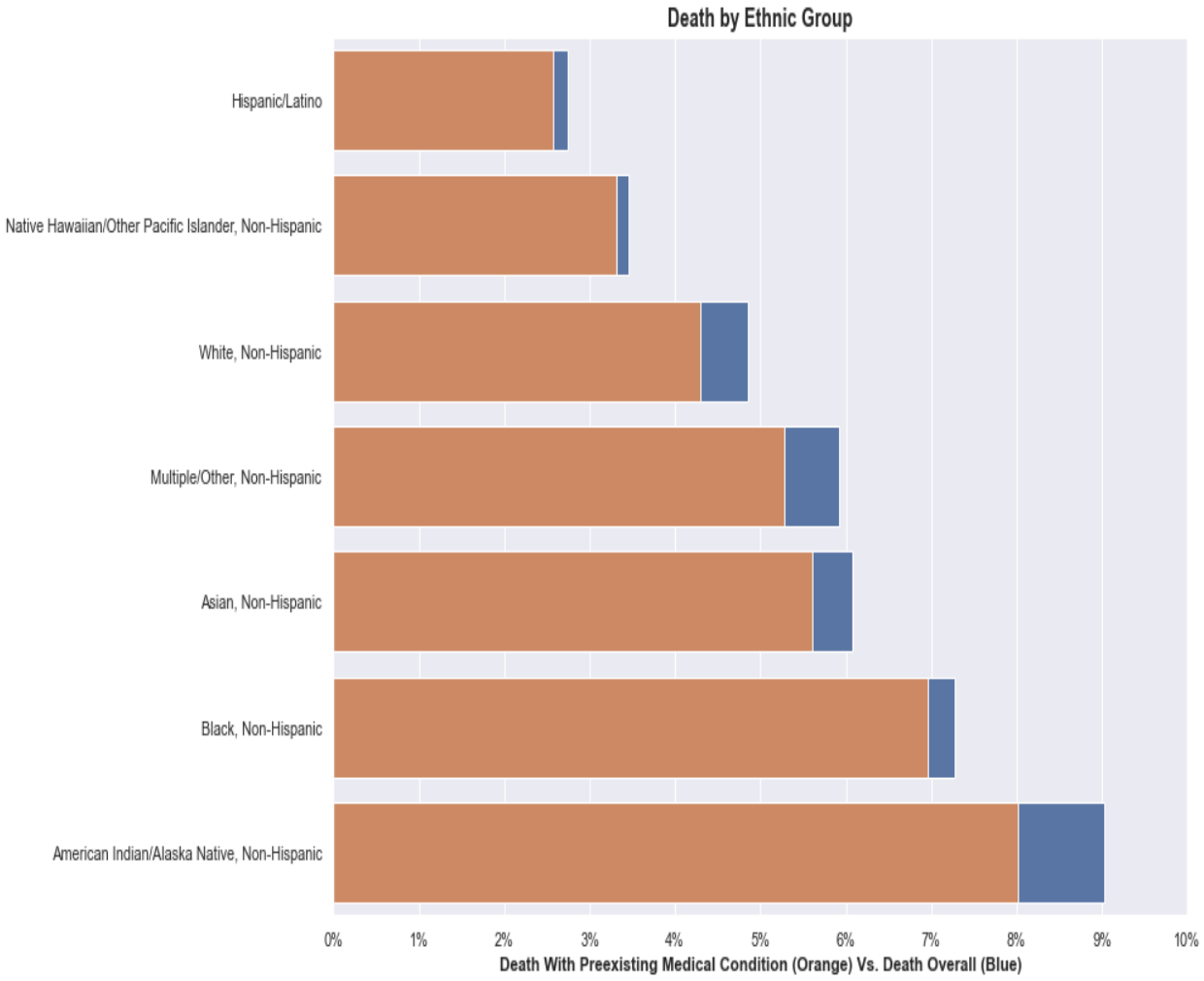
The dataset is modified further to include just deaths only. From 294,496 rows, it now contains only 13,694 rows. This shows that the death rate, overall is 4.65%. This dataset also contains if the person died already had a preexisting condition before contracting the coronavirus. The count for people dying having a preexisting medical condition(s) is 12,718 or 92.9%. This leaves only 976 people dying from coronavirus without having a recorded preexisting condition. This doesn't take into account the person's age, nutrition, lifestyle, or other factors that could possibly be a detriment on an individual's health.

The next visual illustrates the death rate by age when an individual was tested positive for COVID-19. It contains people that died if they already had a preexisting condition. Again, it shows both a logarithmic and linear scale.





The final visual illustrates death by ethnicity with preexisting conditions.



Conclusion

While the distributions across race highlight what ethnicity is more vulnerable to death in regards to a person dying when they contracted coronavirus with or without a preexisting condition, a person's age is a clear indicator on what outcome that person will face statistically. In addition, most deaths from people contracting COVID-19 are from individuals that already had a preexisting condition, almost 93%.

The survival rate, regardless of race or preexisting medical condition is above 90%. The same percentage holds true if a person is below the age of 70. So, if a person is below the age of 70 and does not have a documented preexisting condition, the survival rate is at least 98% (over 99% if below 60).

In conclusion, coronavirus, by itself does not cause death, but can be a catalyst for people that are already diagnosed with a medical issue or for elderly people that may or may not have medical problems.