# Exploratory Data Analysis – Data Analysis of Deaths during the pandemic

## Mohammed Shabuddin Ramizuddin

### Week 12 Project

## Introduction:

The Coronavirus disease (COVID-19) is caused by a newly discovered virus. There are more than a 36.1Million deaths due to the virus within 11 months of virus first found. The COVID-19 stands for ‘COrona Virus and Disease started in the year 2019’. The Virus has spread across the globe, the World Health Organization has declared it as a pandemic. Although, there are a big loss due to COVID-19, I was trying to find if there are some other reasons for the death. Also, I am planning to find the friends who are having to suffer the big loss.

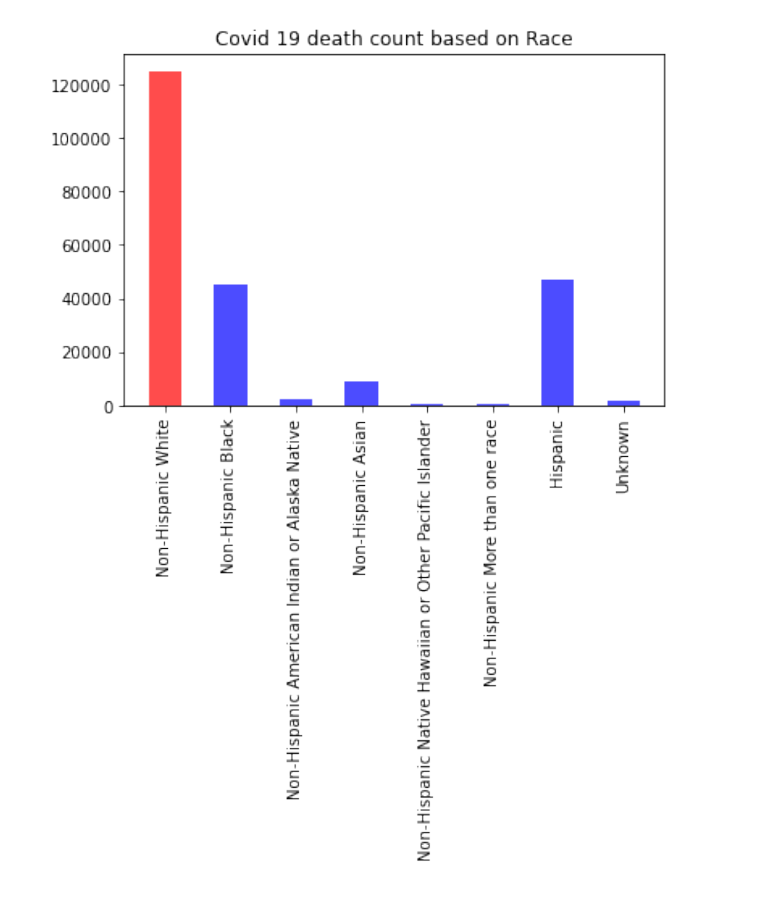
## Statistical/Hypothetical Question

In this project, I am planning to research the following questions.

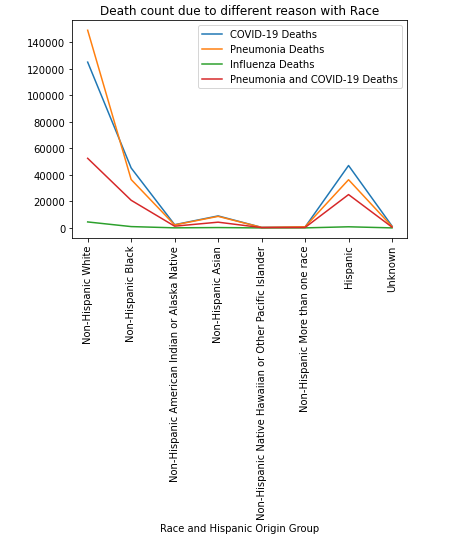
* Total number of Deaths due to COVID-19 based on Race.
* Total number of Deaths due to different reasons by Race.
* Top 10 States where the deaths due to COVID-19 is high.
* Total number of deaths with age
* COVID-19 Death vs Pneumonia Death by Race.

## Outcome of EDA:

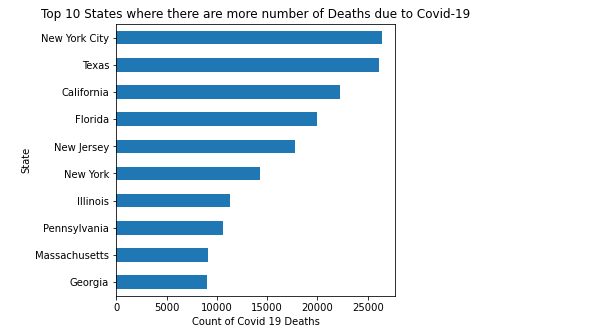
* Total number of Deaths due to COVID-19 based on Race. As per the dataset, It looks like Non-Hispanic White people have more number of deaths during this pandemic.



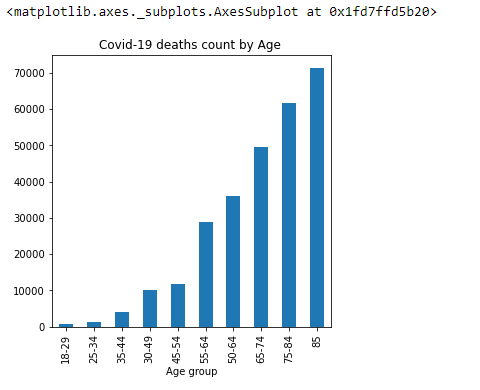
* Deaths due to different reasons is outlined below. As per the dataset, the number of deaths due to Pneumonia is more than any other death reason for Non- Hispanic White Race. Also, It is going hand in hand with other races.



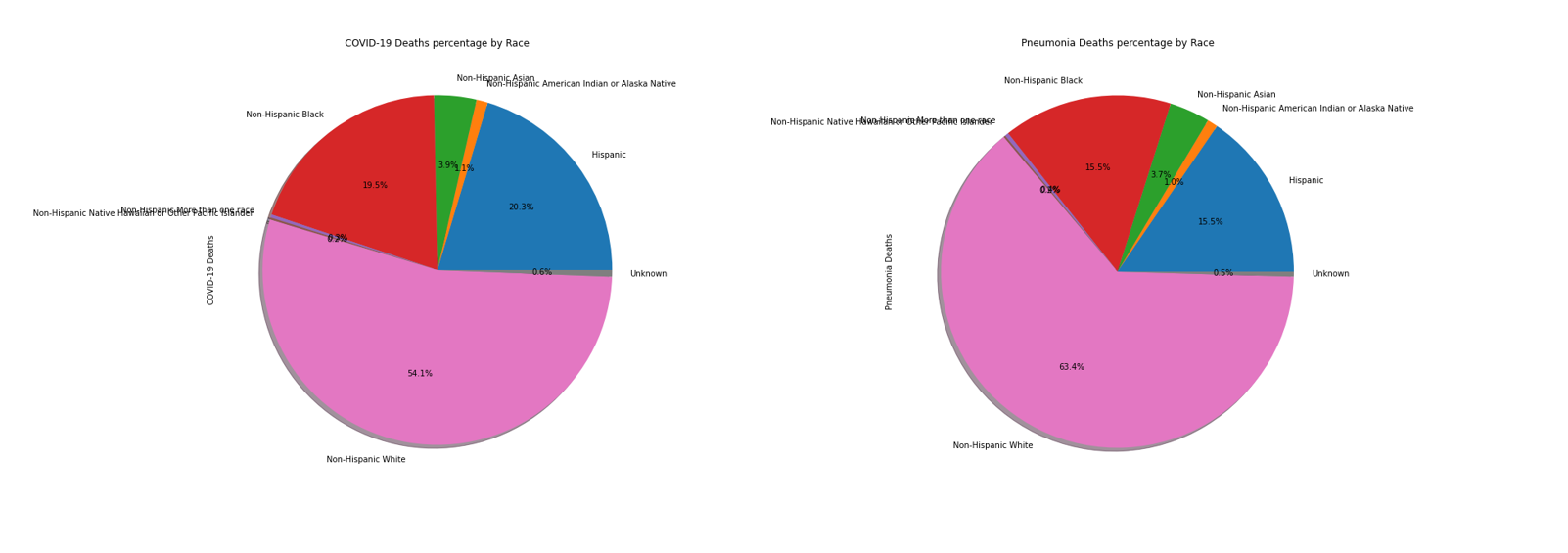
* Top 10 States where deaths due to COVID-19 is higher.



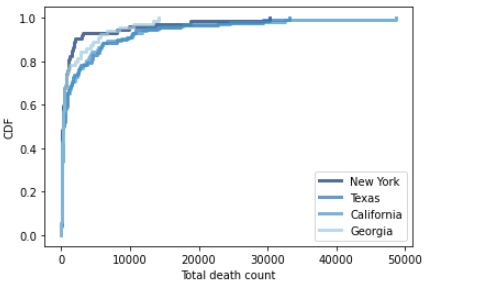
* Number of deaths based on the age range.



* COVID-19 Death vs Pneumonia Death by Race



* CDF for 4 of the cities



## Reference:

<https://healthdata.gov/dataset/deaths-involving-coronavirus-disease-2019-covid-19-race-and-hispanic-origin-group-and-age>

Allen B. Downey (2015). Think Stats, Second Edition.