

# **Written Report on the Impact of COVID-19 Across Demographics**

## **Lydia Wang**

### Introduction:

The COVID-19 pandemic has had a profound impact on global health, exposing vulnerabilities in public health systems and exacerbating existing health conditions. Our study focuses on the United States, aiming to uncover how underlying health conditions have influenced COVID-19 related deaths across various demographics. This dataset shows health conditions and contributing causes in conjunction with deaths involving COVID-19 by age group, date, and state. Data from 2022 and 2023 are provisional; estimates for 2020 and 2021 are based on final data. The analysis seeks to answer pivotal questions regarding the age groups most affected and the top 5 pre-existing health conditions(out of 23 in total) that posed the greatest risks during the pandemic.

### Methods:

#### Data Acquisition

The dataset employed in this analysis was procured from Data.gov, specifically from the dataset titled "Conditions Contributing to Deaths Involving Coronavirus Disease 2019 (COVID-19) by Age Group," which is disseminated by the Centers for Disease Control and Prevention (CDC). It is a comprehensive collection encompassing 621,000 entries across 14 distinct columns, providing granular details on the fatalities associated with COVID-19, along with demographic specifics and comorbid conditions.

#### Data Cleaning and Wrangling

The preliminary step in our analysis involved a meticulous inspection of the dataset to ascertain the integrity of the data. Approximately 5.1% of the records contained missing values, notably within the 'Year', 'Month', and 'Flag' columns. These entries were excised from the dataset to ensure the robustness of our analysis. The 'data.table' package facilitated the elimination of any duplicate records, thereby purifying the dataset. Concurrently, the 'dplyr' package was instrumental in sculpting the dataset further—by selecting pivotal columns, discarding rows marred by missing values, and performing data grouping and summarization tasks pertinent to our research question.

#### Tools for Data Exploration

The 'ggplot2' package was the primary tool for data visualization, enabling the creation of illustrative and publication-grade figures. Its comprehensive suite of functions allowed for intricate customization and refinement of the visual output. Furthermore, the 'knitr' package was utilized to present the data frames in a structured and aesthetically pleasing tabular format, enhancing the readability and interpretability of the tabulated data.

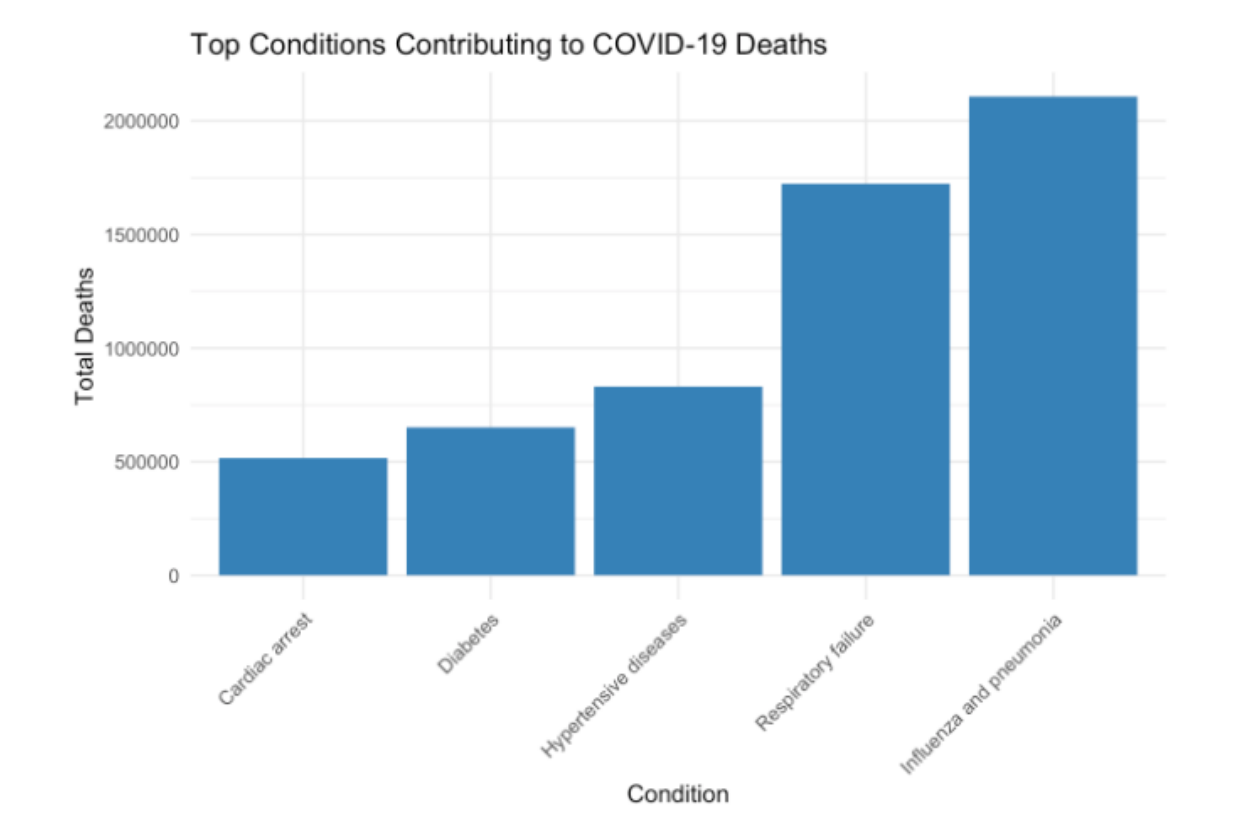
### Results:

Two interactive maps could be found on the homepage of the website: the first map shows the total number of deaths per state, while the second map shows the most prevalent conditions and their counts per state.

Our analysis revealed that certain pre-existing conditions notably increased the risk of death from COVID-19. These conditions varied in impact across different age groups.

We identified the top five underlying health conditions that contributed significantly to COVID-19 deaths:

1. Influenza and Pneumonia
2. Respiratory failure
3. Hypertensive diseases
4. Diabetes
5. Cardiac arrest



Each condition's impact was further analyzed across age demographics to understand the age groups most at risk.

Influenza and Pneumonia: Most prevalent in the 75-84 age group.

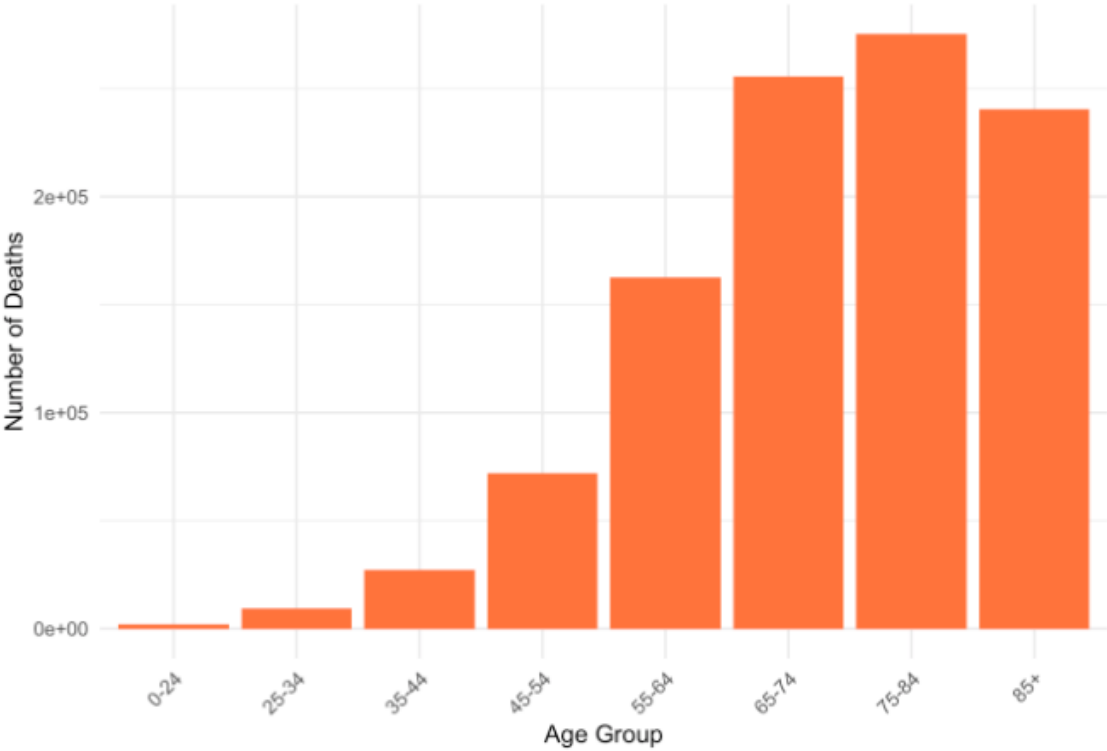
Diabetes: Had the highest impact on the 65-74 age group.

Cardiac Arrest: Showed a significant effect on the 75-84 age group.

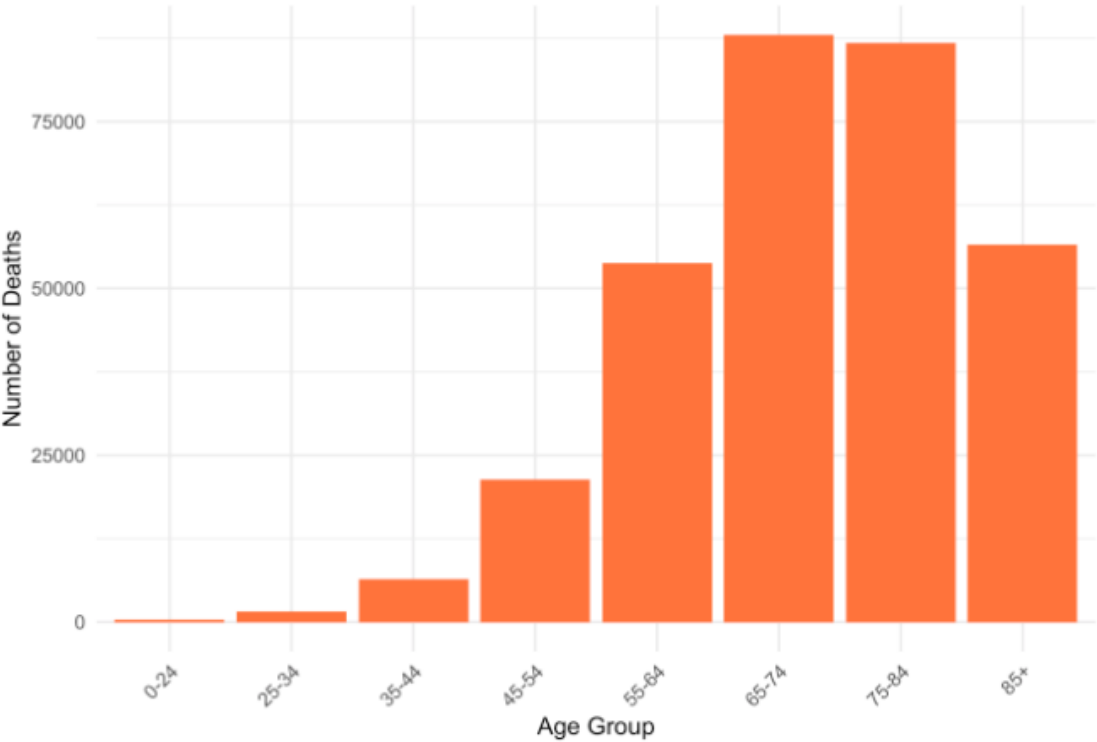
Hypertensive Diseases: Were most impactful for individuals aged 85 and above.

Respiratory Failure: Had a pronounced impact on the 75-84 age demographic.

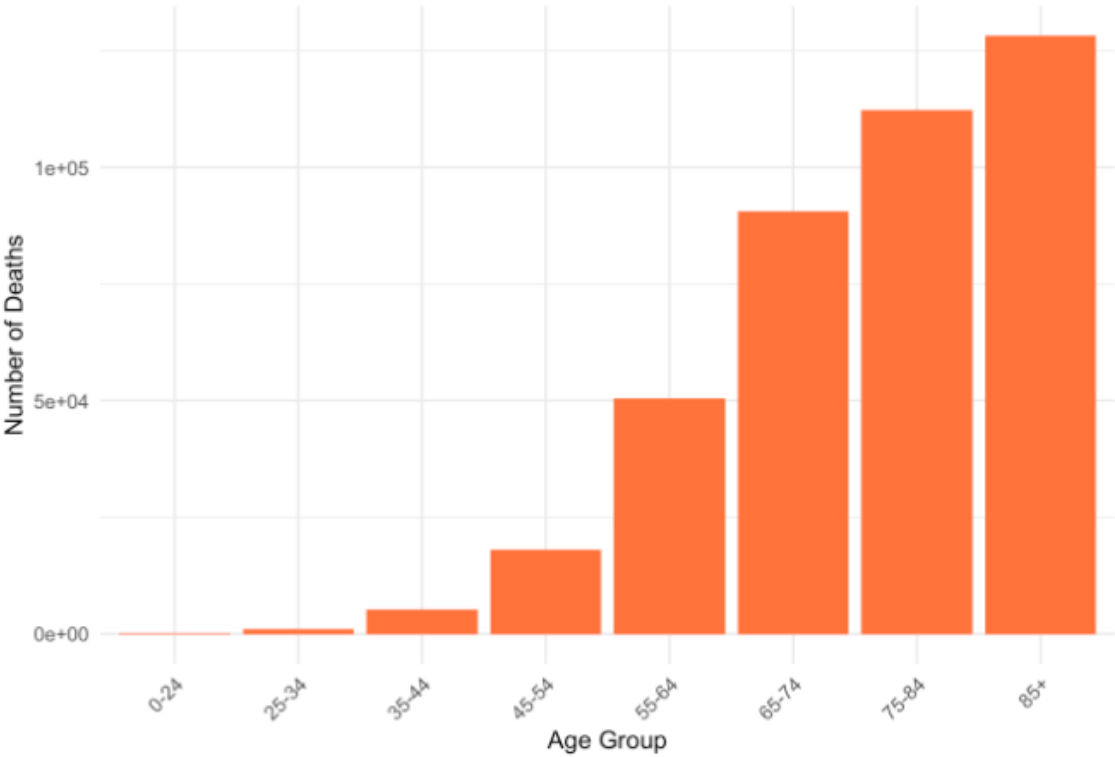
Impact of Influenza and Pneumonia on COVID-19 Deaths by Age Group

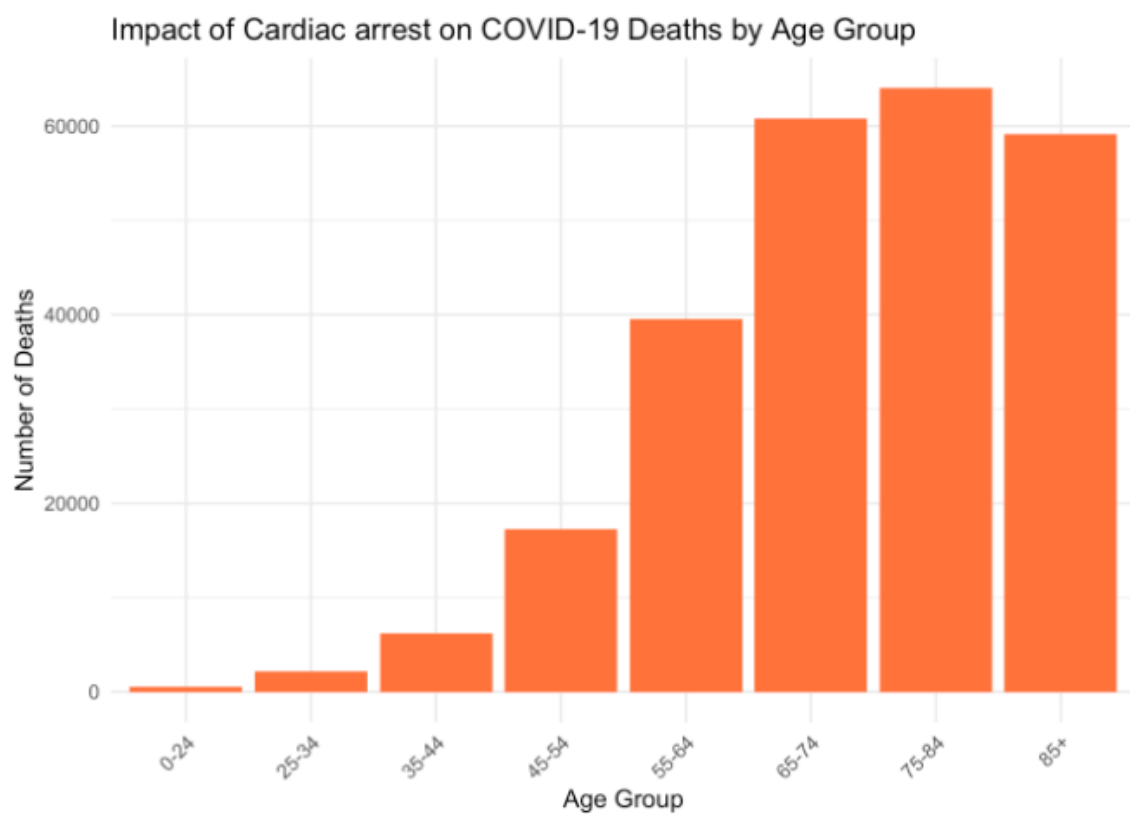


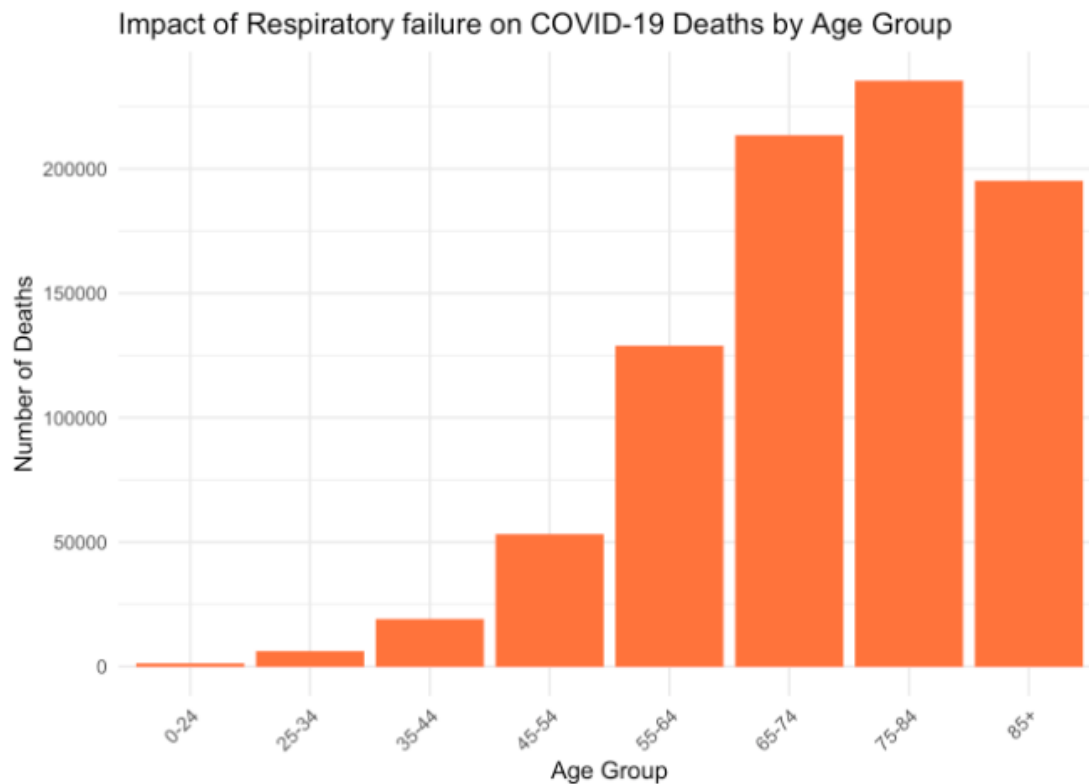
Impact of Diabetes on COVID-19 Deaths by Age Group



Impact of Hypertensive diseases on COVID-19 Deaths by Age Group







### Conclusion:

This comprehensive report delineates the multifaceted impact of COVID-19 on the demographic tapestry of the United States, with a keen focus on how underlying health conditions have influenced mortality rates. The meticulous analysis, grounded in a robust dataset, has unearthed pivotal insights that underscore the differential impact of the pandemic across age groups, underpinned by pre-existing health conditions.

Notably, the investigation has illuminated the heightened vulnerability of the elderly population—particularly those contending with influenza and pneumonia, respiratory failure, hypertensive diseases, diabetes, and cardiac arrest. These conditions, prevalent in the 65-84 age bracket and above, have been significant contributors to the mortality rates, emphasizing a clear age-related susceptibility to the virus's lethality.

The findings of this report accentuate the critical need for targeted public health interventions. Such strategies must be tailored to protect the most vulnerable demographics, especially the elderly and individuals with specific health vulnerabilities, who have been disproportionately affected by the pandemic.