

Data Analytics and Visualization in Healthcare: Mid Term Project

By Lakshmi Sabari Priya Jaini

Introduction and Background

For this project, the topic that I have chosen is “*COVID-19 Deaths by Sex and Age Over Time and their Comparison with Pneumonia and Influenza Deaths.*” My primary motivation for choosing this topic is that it is a topic of paramount importance. As the COVID-19 pandemic impacted global populations, understanding how the disease affected different demographic groups is critical for public health planning and intervention strategies. By delving into the data, we can uncover valuable insights into which age and sex groups are most vulnerable to severe outcomes, guiding targeted vaccination campaigns and healthcare resource allocation. Moreover, comparing COVID-19 deaths to deaths from other respiratory illnesses like pneumonia and influenza can provide a unique perspective on the pandemic's gravity, shedding light on whether COVID-19 poses a distinct threat to specific populations. Such comparative analyses can inform evidence-based decision-making by health authorities and policymakers, potentially influencing vaccination prioritization and healthcare infrastructure investments. I personally got inspired by NYU Professor, Dr. Anasse Bari, who did a lot of research related to the pandemic during its early phases.

Dataset

- **Source:** I obtained the dataset from the Centers for Disease Control and Prevention (CDC) official website.
- **Data Creation:** The data was collected to track and report COVID-19 deaths by various demographics and locations, and compare them with deaths caused by some other diseases. The primary purpose was to inform public health policy, medical research, and emergency response efforts.
- **Data Collection:** The data was collected by National Center for Health Statistics and published by CDC.
- **Funding:** The project was funded by the U.S. federal government and supported by public health agencies.
- **Timeline:** The dataset covers a timeline from 2020 to 2023, capturing the evolving nature of the pandemic.
- **Size:** The dataset comprises of 138K rows and 16 columns.
- **Locations:** It covers all U.S. states and territories, offering a comprehensive view of the pandemic's impact.

Data Story

Here are insights for each of the visualizations along with explanations of best practices used in them:

COVID-19 Deaths by State Map

- Insight: The map highlights the significant regional disparities in COVID-19 deaths. It's clear that some states have been more severely affected by the pandemic than others.
- Best Practice Explanation: Geographic Representation - The use of a map is effective in conveying geographical patterns. Color-coding provides a quick visual reference for variations in COVID-19 deaths across states.

Deaths by Age Group and Cause Bar Graph

- Insight: The bar chart demonstrates that deaths due to all causes are most concentrated in the older age groups, particularly the "75+ years" category.
- Best Practice Explanation:
 - Stacked Bars - Stacking the bars allows for easy comparison of COVID-19, pneumonia, and influenza deaths within each age group. This visually represents the composition of deaths by cause.
 - Color-Coding - Different colors for each cause of death aid in differentiating data series.

Deaths by Year and Gender Table

- Insight: It can be seen from the table that apart from influenza, deaths of males are higher in number than deaths of females. This indicates that females are more susceptible to falling prey to influenza than males.
- Best Practice Explanation:
 - Structured Layout - The structured layout, with clear headers and gridlines, enhances readability and makes it user-friendly.
 - Having the Total Deaths column either at the beginning or at the end helps in users as it makes the largest numbers to be at the beginning or at the end.

Time Series of COVID-19 Deaths Line chart

- Insight: We can clearly see that there were by-and-large 2 waves of increased deaths, indicating 2 COVID waves, one at the end of 2020 and the other at the beginning of 2022.
- Best Practice Explanation:
 - Time Series Representation - A line chart is ideal for displaying time-dependent data, enabling users to identify trends and patterns over different quarters.
 - Clear Axes - Properly labeled and formatted axes make it easy for users to interpret the timeframe and data scale.

These insights and best practices help ensure that each visualization effectively conveys meaningful information while maintaining clarity and visual appeal.

References

CDC. (2020). Provisional COVID-19 Deaths by Sex and Age.
<https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-by-Sex-and-Age/9bhg-hcku>.