# 260_Final_Report

## Abstract

Quarto enables you to weave together content and executable code into a finished document.

## Introduction

The COVID-19 pandemic has profoundly influenced global health, economies, and societal structures. From January 2020 to December 2024, the United States experienced multiple waves of COVID-19 infections and deaths. These waves provide a unique opportunity to analyze the evolution of COVID-19's impact over time. By examining mortality data across states and time periods, this study aims to investigate patterns and trends to better understand the trajectory of this pandemic.

Recent studies have investigated and highlighted the importance of analyzing mortality trends to better understand the pandemic's impact. For instance, Chan et al. (2021) quantified the pandemic's effects on excess mortality and life expectancy (Chan et al., 2021). Moreover, Woolf et al. (2021) analyzed excess deaths in the United States and showed increased mortality from non-COVID-19 causes, such as heart disease and Alzheimer's disease during COVID-19 waves, demonstrating the pandemic's indirect effects on public health outcomes (Woolf et al., 2021). While these studies have provided valuable insights, there are limitations that our study intend to address. Most prior analyses focus on aggregated data or specific time periods and pay little attention to state-level variations and temporal shifts of the pandemic's impact. Our study tries to offer a much comprehensive and longitudinal analysis that encompasses the entire pandemic period in both national and state level.

In this study, the analysis focuses on several key aspects of the pandemic. First, we divide the timeline into distinct waves based on data visualizations of infection and mortality trends. This section allows for understanding the virus's evolution and impact over time. Second, we compute death rates for each state during these periods, identifying states that performed better or worse in terms of mortality outcomes. These findings could potentially identify the effectiveness of various public health strategies and problems within state-level healthcare systems. Third, we explore whether COVID-19 became more or less virulent across different periods by investigating changes in mortality rates and excess deaths over time. Furthermore, we extend the scope to estimate weekly excess mortality for each state and evaluate whether COVID-19 deaths explain these excesses.

By investigating those key aspects of the pandemic, this study seeks to provide a comprehensive understanding of the dynamics of COVID-19 mortality in the United States, potentially contributing to future pandemic analysis and serving as resources for public health officials to study healthcare systems and improve response strategies and allocation of resources across the nation.

## Results

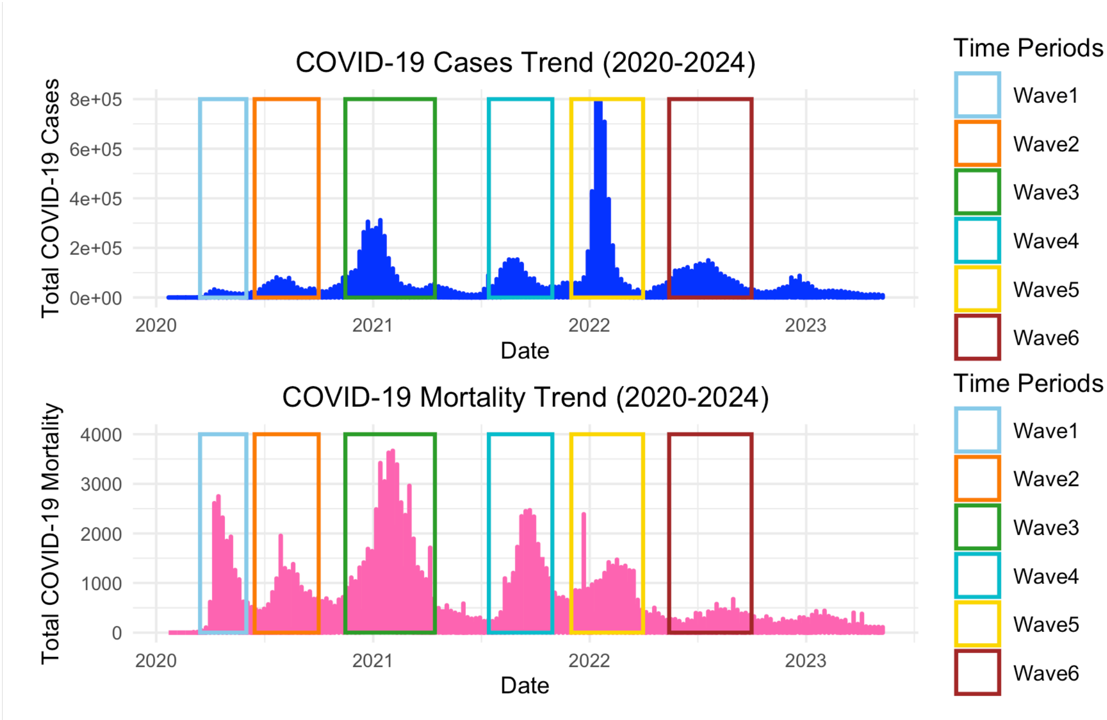
Q1

```r
library(magick)
```

```
Linking to ImageMagick 6.9.12.93
Enabled features: cairo, fontconfig, freetype, heic, lcms, pango, raw, rsvg, webp
Disabled features: fftw, ghostscript, x11
```

```r
img <- image_read("Graph1.png")
print(img)
```

```
  format width height colorspace matte filesize density
1    PNG  1826   1132       sRGB  TRUE   268718   57x57
```




The COVID-19 case and mortality trends from 2020 to 2024 reveal six distinct waves, each varying in intensity and duration. Case trends show an initial gradual increase during Wave 1, followed by a sharp rise in Wave 2, peaking significantly in Wave 3, and then tapering into moderate peaks during Waves 4, 5, and 6. Mortality trends closely align with these waves, with the highest mortality observed in Wave 3, a noticeable decline in Wave 4, and relatively lower but steady levels in Waves 5 and 6. The provided charts highlight the temporal dynamics of the pandemic, illustrating the fluctuation of cases and mortality over time.

## References


Woolf, S. H., et al. (2021). "Excess Deaths from COVID-19 and Other Causes in the US, March 1, 2020, to January 2, 2021." JAMA. Retrieved from jamanetwork.com. https://jamanetwork.com/journals/jama/fullarticle/2778361

Chan, E. Y. S., et al. (2021). "Impact of COVID-19 on Excess Mortality, Life Expectancy, and Years of Life Lost in the United States." PLOS ONE. Retrieved from journals.plos.org. https://pubmed.ncbi.nlm.nih.gov/34469474/

## Methods

Population Data: Source: NST-EST2024-POP.xlsx, collected by the U.S. Census Bureau. Website: https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html. Time Frame: 2020 to 2024. Provides annual state-level population estimates, used for calculating per capita COVID-19 case and death rates.

Regional Data: Source: JSON file hosted on GitHub. URL: https://github.com/datasciencelabs/2024/raw/refs/ Defines regional groupings of U.S. states and territories (e.g., "NY & NJ & PR & VI") for aggregated analysis.

COVID-19 Cases and Death Data: Source: Centers for Disease Control and Prevention (CDC) API. Endpoint: https://data.cdc.gov/resource/pwn4-m3yp.json. Time Frame: January 22, 2020, to May 10, 2023. Provides daily records of new cases and deaths at the state level.

for Q1 This analysis processes and analyzes population and COVID-19 data (2020–2024) using a combination of data cleaning, transformation, and visualization techniques. The R programming language and libraries such as readxl, dplyr, tidyr, ggplot2, janitor, and lubridate were employed. Population data from an Excel file was cleaned by standardizing geographic names and reshaping it into a long format for ease of analysis. It was merged with regional data extracted from a JSON file using jsonlite, mapping states to predefined U.S. regions. COVID-19 case and death data were retrieved via an API using httr2, cleaned to handle anomalies like negative values, and normalized by population to compute metrics per 100,000 people. Data visualization, using ggplot2, involved smoothed time-series plots for identifying pandemic waves and faceted plots for regional analysis. A rolling average approach was applied to smooth fluctuations, highlighting significant trends in cases and deaths. Each region's unique patterns were visualized, emphasizing geographic variability. While assumptions include consistent reporting and static yearly populations, this approach integrates advanced data handling and visualization techniques to provide clear insights into pandemic trends.

To clean and analyze the data, we utilized various R packages, including readxl for reading Excel files, dplyr and tidyr for data cleaning and manipulation, jsonlite for processing JSON data, httr2 for API requests, lubridate for date management, and ggplot2 for visualization. The analysis started with population data from an Excel file, where unnecessary rows were filtered, geographic names cleaned, and the data reshaped into a long format using pivot_longer. State abbreviations were added for consistency, and regions were incorporated using external JSON data. For COVID-19 cases and deaths, we retrieved data via an API, selected relevant columns, converted data types, and handled anomalies such as negative values by replacing them with NA. These datasets were joined by state to compute cases and deaths per 100,000 population for normalization. Finally, we visualized trends using time-series plots, faceted by regions, and smoothed data to identify significant waves from 2020 to 2024. This approach ensured comprehensive data integration, normalization, and insightful visual analysis across different demographic dimensions.