# **Death by State**

## Introduction

Our project focuses on the death rate by state in the United States. We used data from the CDC to further break down the mortality rate from diseases such as COVID 19, heart disease, Alzheimer’s, Cancer, Influenza and more. This project is important because it highlights the states that are most affected by these illnesses and the seasons of the year that receive the largest number of deaths. We break the data down and comb through each year to fully analyze the data and compare the mortality rate of different diseases throughout the years 2020-2023.

Map

Description automatically generated

Chart, treemap chart

Description automatically generated

Questions we tried to answer:

* Which state has reported the highest number of deaths in the last 3 years.
* Which are the top 10 states with highest number of deaths?
* What are the main causes of death in US?
* In what season do most deaths occur by natural causes and what is the correlation between the seasons where the most deaths occur by all causes and natural causes?
* In what season is the death rate highest due to Cancer & Septicemia?
* In what season is the death rate highest due to Influenza & Respiratory Diseases?
* Does COVID-19 have any impact on the death totals and rates for the other mortality measures?
* In what season is the death rate highest due to COVID 19 and heart disease?

Where we are getting data from

CDC - National Center for Health Statistics. Weekly Provisional Counts of Deaths by State and Select Causes, 2020-2022.

Libraries /modules used

Used different libraries like Pandas, Matplotlib, seaborn and we are using choropleth (from plotly.express) for plotting the maps.

Data cleaning and preparation

* Imported data files and read the data using pd**.**read\_csv(death\_data\_path)
* Obtained a list of columns for death\_data and population\_data then reduced the number of columns to only show 2020-2023
* Filtered the data to only show the diseases were wanted to further analyze
* Changed the data type from object to datetime and added month, year and month-year columns.
* Extracted month by the month names.
* Created a season list by month.
* Renamed the columns so it’s easier to read & understand.
* Combined Cerebrovascular diseases and heart diseases into one column named “all heart diseases”.
* Combined columns for chronic lower respiratory diseases and other diseases in the respiratory system into one column named “respiratory diseases”.
* Reordered columns for visibility and so data is able to be understood easily.
* Identified incomplete rows and filled the NAs with 0 and replaced the NaN values with 0
* Verified all row counts so they are equal to 8,164
* Converted measured to an integer data type
* Organized the data so it shows the deaths by year and month
* Created a population by year and state data frames

## **Analysis**

The questions we answered:

1. What are the main causes of death in the United States? - Through our data we learn that COVID 19, heart disease and cancer are the main causes of death in the United States.
2. What are the top 10 states with the highest number of deaths?

Chart, bar chart

Description automatically generated

Map

Description automatically generated

The states with the highest number of deaths in the United States are California, Texas, Florida, Pennsylvania, Ohio, Illinois, New York, North Carolina, Michigan and Georgia. Through this chart you can clearly see the number drop as it goes from California to Georgia.

1. In what season do most deaths occur by natural causes and what is the correlation between the seasons where the most deaths occur by all causes and natural causes?

A picture containing text, clipart, vector graphics

Description automatically generated

As you can see in the bar chart, the seasons where the most deaths occurred due to all causes as well as natural causes in 2020-2023 is the winter. This could be due to a number of reasons such as improper shelter during harsh weather and illnesses such as pneumonia and the flu. In 2022 we see the winter is clearly a much larger death rate compared to the other months. COVID 19 could also play a role in this because many of the symptoms caused by COVID mimic the flu and the common cold.

1. In what season is the death rate highest due to Cancer & Septicemia?

A picture containing text, clipart, vector graphics, window

Description automatically generated

These charts show that winter was the season for the highest death rate due to Septicemia. It was very interesting that the death rates stayed pretty consistent throughout the four seasons when looking at deaths due to cancer except for a small uptick in the winter season of 2022. This tells us that there is no correlation between the seasons and cancer death rates. There is no season of the year where people are more susceptible to cancer.

1. In what season is the death rate highest due to Influenza & Respiratory Diseases?

A picture containing text, clipart, vector graphics, window

Description automatically generated

This was interesting because you can see the effect COVID 19 had on the mortality rates. There is a clear uptick in spring 2020 of respiratory and influenza and pneumonia related deaths. Spring 2020 is also when COVID 19 will start to truly affect the country and further affect so many people’s lives.

1. Does COVID-19 have any impact on the death totals and rates for the other mortality measures?

Chart, histogram

Description automatically generated

This graph shows the total number of deaths along with the months and year where people were affected. You can clearly see how COVID 19 has such a high number compared to the other diseases. Especially during 2020-2023 COVID 19 was at its peak. The coronavirus has affected and killed more people than any other illness in 2020.

1. In what season is the death rate highest due to COVID 19 and heart disease?

Chart, bar chart

Description automatically generated

This is a very important question because it is comparing two of the top leading causes of death in the United States in 2020-2023. It further shows how the huge jump in the mortality rate in winter due to COVID 19 compared to any other season. Heart disease numbers stay consistent during the seasons with only a small increase in winter 2022.

## **Conclusion**

Our project is an opportunity to look at the leading causes of death in the United States. The data we selected shows the number of deaths by states and their causes. This dataset and the visuals we provided are important because they show us the prevalence of certain diseases from the years 2020-2023. In the last couple of years health in America has been a crucial topic of conversation because of COVID-19. The coronavirus has made many Americans focus more on the importance of health and the causes of many diseases. In 2021 the highest number of deaths occurred in the United States, this is because the fatality rates of covid were rising while there were still several people suffering from other illnesses such as cancer, diabetes, respiratory diseases, influenza and more. While covid emerged and affected over 100 million Americans, the leading cause of death before covid was heart disease. Our project is essential because knowing the leading causes of death in America as well as the numbers of deaths will cause more people to engage in healthy habits and take all the right preventative measures so that they can increase their chances of not getting sick. Throughout our project we see visuals that show different aspects of the number of deaths by state. California has the top number of deaths followed by Texas and Florida. This could be because of the fact the population of states like California, Texas and Florida are larger and have more people that can be susceptible to disease. Understanding the prevalence of these diseases from 2020-2023 shows us what diseases most people are suffering with in different states. Our project highlighted how essential proper shelter is during winter months. We showed the death rate due to influenza, pneumonia and COVID 19 during the winter season. Those numbers were significantly larger than the death rate during the spring and summer. The point of our project is to spread awareness on health disparities across the country and show how crucial it is that there are the right healthcare resources for people all year round. Changes in our healthcare system that will provide proper education on the causes of death and implement preventative measures will increase the quality of life in Americans worldwide.