

## COVID-19 Deaths in U.S.

Michael David  
CS82 Data Science  
Pr. Seno  
5/31/22

### *ABSTRACT*

*This data-driven analysis seeks to explore the rate of COVID-19-related deaths in the U.S. It looks at data collected from February 15, 2020, to March 15, 2022. Spanning a total of 25 months, it looks at the cumulative increments of 100k deaths; going from 0 cases on February 15, 2022, to more than 1 million by the end of the analysis timeline on March 15, 2022. Visualizing the data, increases and decreases in the rate of change can be noted and further explored and connected to societal events and trends. The analysis seeks to associate significant national events with the changes in the death rate.*

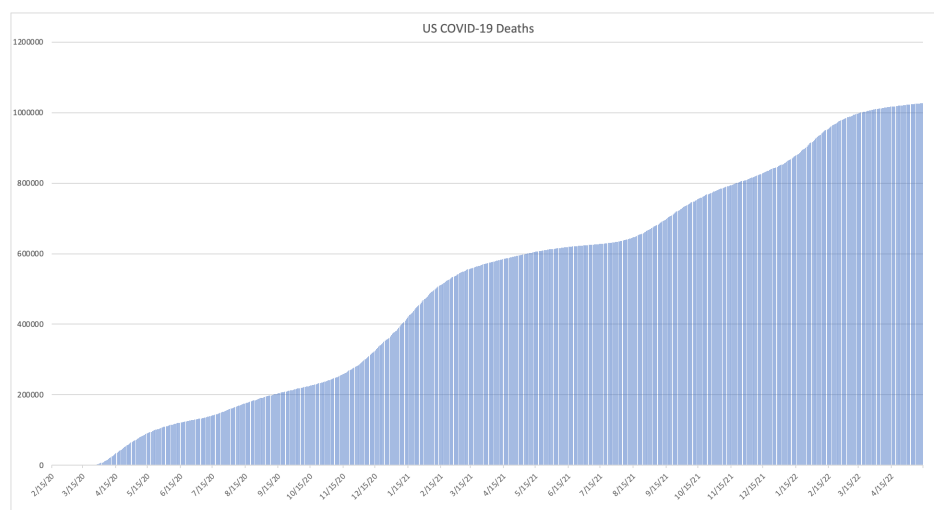
### Introduction

COVID-19 is the name for the infectious disease caused by SARS-COV-2; a virus from the coronavirus family. Since the first reported cluster of cases in Wuhan China in December 2020, the virus has spread to all continents and surpassed hundreds of millions of cases and millions of deaths. The ramifications of COVID-19 cannot be understated; its impact as one of the most significant global events is solidified in all those who have experienced it. The far-reaching implications of COVID-19 have disrupted the lives of billions and changed countless industries. Apart from the sectors that COVID-19 has impacted, the death toll of the pandemic has passed grim milestones repeatedly since its initial spread in 2020. Despite responses from national governments across the world, outbreaks have proved inevitable. Emergency policies and social

distancing measures have aided in slowing down transmissions, but often these responses from government officials were late. Across the globe, adapting to the new way of life caused by COVID-19 created conflict. Many were opposed to lockdown measures. Patterns of social movement caused by the seasons continued to play a role in the transmission of COVID-19 along with developments of the virus (genetic variants).

## Data

The data obtained from Kaggle: *Covid-19 Global Dataset* by Joseph Assaker, contains COVID-related data from February 15, 2020, to May 14, 2022, for 225 countries. It includes features for daily new cases, cumulative total cases, daily new deaths, cumulative total deaths, and active cases. For the purposes of this analysis, only data pertaining to cumulative total deaths in the U.S. was explored. Readability and format led me to choose this dataset over others. Furthermore, the data involving the U.S. was split into various data files that each contain data for deaths in increments of 100,000 deaths; with 10 files dividing 1,000,000 deaths equally. The first data file spans from February 15, 2020, to May 15, 2022. The last begins on January 15, 2022, and ends on March 15, 2022. A final data file was created for the 25 months that the original dataset spans. This data file named “monthly” has a single entry for each month.



## Methodology

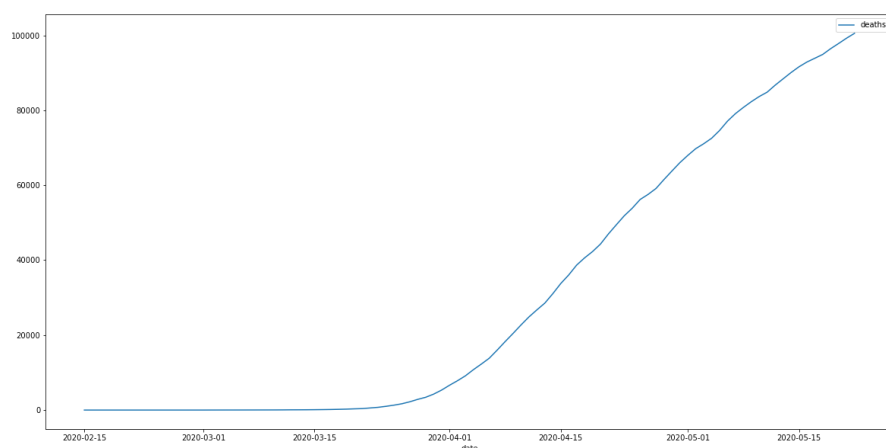
Using excel and google sheets (for user interface and file-sharing ease) for preprocessing, 10 unique files were created to analyze the timeline with a higher degree of accuracy. These files would eventually be uploaded to Jupyter Notebook for visualization. The “monthly” file was used as a quick overview of how the data behaves and to determine which timeframes need further inspection for possible change in the rate of change of deaths.

Visualization is made possible by using data science relevant libraries in Jupyter Notebook such as Seaborn and matplotlib. Graphing line charts, changes to the rates of change of deaths are more easily visualized. Such visualization of the changes can be further facilitated because each data file now represents only a fraction of the original data frame. While an overview of the data is useful, the purpose of this analysis was to associate certain time frames with increases in the death count.

## Analysis

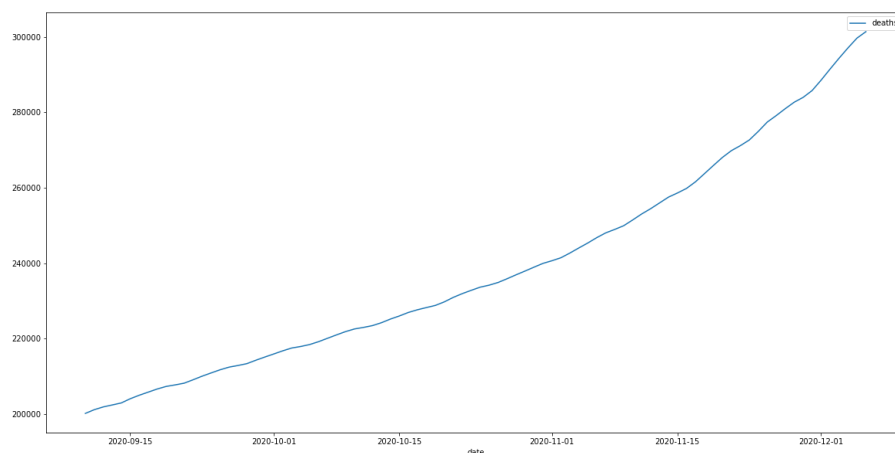
By first looking at the summarized file “monthly” and then graphing a line chart, three of the graphs demonstrate an increase in the rate of change in deaths.

Firstly, the initial outbreak in the U.S. following the arrival of the index patient in Washington can be obviously noted in the graph below. The graph contains data for the first 100,000 cases in the U.S. spanning from 2/15/2020 to 5/15/2020. The graph follows something similar to an exponential curve as COVID rapidly spread in the U.S. upon first contact with a susceptible population.

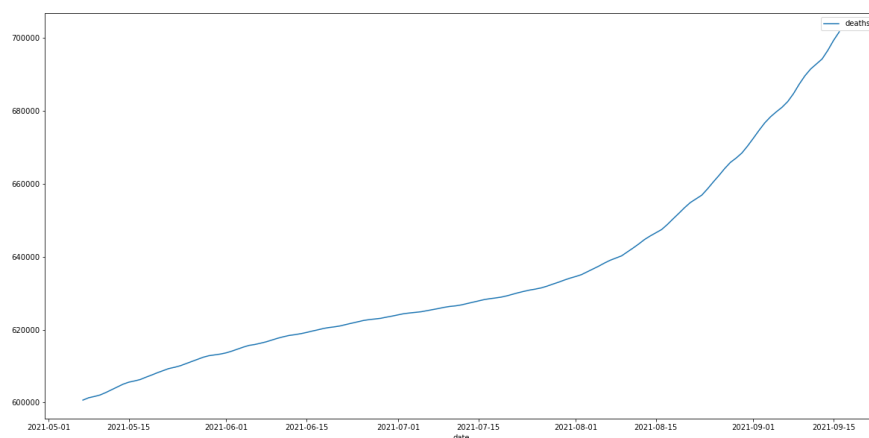


Secondly, during the winter of 2020, another increase in deaths can be seen in the graph below. The increase in the rate of change in deaths can be connected to the festivities in

American culture and the associated gatherings and travel. Disregard on behalf of some and a hesitancy to adapt to the social distancing guidelines led to the “second wave” of the pandemic in the U.S.



Thirdly, the arrival of the Delta variant to the U.S. and its highly contagious nature created yet another spike in cases which was followed by an increase in the number of deaths. In the summer of 2021, the Delta variant would eventually become the dominant strain and create a third wave of deaths.



## Conclusion

Upon visualization and further research surrounding COVID trends, meaningful insights can be deduced from the data. The ability to transform a csv file into graphs and then create meaning out of the graphs has serious implications and uses. From the analysis, historical events were discovered through visualization using various Python libraries.