COVID-19 Data

DSC 680 Applied Data Science

Bellevue University

Spring 2021

Conrad Ibañez

Abstract

The Covid-19 pandemic turned people's lives upside down beginning in early 2020. Since the pandemic began, many people around the world have been infected or have died due to the virus, and new waves of infections are still occurring even today, over a year later. Every person has been impacted by the disease, and we have had to make many adjustments in our daily lives. Some of us may have been infected by the virus or know someone who was infected and even died from it. At some point, society was at a standstill with many establishments being shut down. Many people have lost jobs, and many activities, such as work, school, and meetings, have taken place from home. Because of the effects of this pandemic, a large volume of data has been recorded and is available for analysis.

Entities such as federal and state governments, as well as health organizations have created websites to convey information and daily status for the disease. This project analyzes Covid-19 data around infections, deaths, and vaccinations in the United States. Data will be presented in different ways to show relationships among those features and to offer views that are different to what is currently available.

Background

The Covid-19 pandemic has had a broad range of impacts to society. Numerous statistical data is readily available across many domains and presented on many websites of numerous organizations. There is much information being captured around infections, deaths, and vaccinations to help guide public health policy and to influence certain behaviors among people, such as wearing masks and social distancing.

The Centers for Disease Control and Prevention (CDC) has established a tracker for the impacts of the disease on its website with abundant information for national, state, and local

levels in the U.S. Additionally, the information can be downloaded in various formats. There are also links to other resources and websites with even more information about the disease.

Due to the vast amounts of information, it is sometimes difficult to browse the details for Covid-19 impacts. Furthermore, the relationships between infections, deaths, and vaccinations are not apparent at times because each attribute is displayed independently of each other. This project will analyze the Covid-19 data and offer different visualizations that shows trends and relationships among these attributes.

Problem Statement

The Covid-19 pandemic has changed our lives since it began in early 2020. Metrics around infections, deaths, and vaccinations are being recorded and monitored closely. The data has been used to guide public health policy and behavior, such as wearing masks and social distancing.

This project analyzes Covid-19 for the United States and for the state of Georgia where I live. Information will be presented to show how metrics have changed over time and the current trends. Additionally, different visualizations will show how vaccinations have impacted the number of infections and deaths. This project will address various questions including the following:

1. How many total infections have there been in the U.S. and how as this increased over time?

2. How many total deaths have there been in the U.S. and how as this increased over time?

3. When did the highest number of infections occur?

4. When did the highest number of deaths occur?

5. Did different waves of infections and deaths occur?

6. What do the previous metrics look like for the state of Georgia?

7. How has the roll out of the vaccinations impacted the number of infections and deaths?

8. How many people have gotten the vaccines in the U.S. and in Georgia?

9. Are we safer now that vaccines are more readily available?

10. Can we live our lives normally and not wear masks or social distance now that people are getting vaccinated?

Data Understanding

The datasets for this project can be found on the Centers for Disease Control and Prevention (CDC) website and healthdata.gov. The datasets appear complete and clean.

- Cases & Deaths
  - https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36
- Covid-19 Testing Data
  - https://healthdata.gov/dataset/COVID-19-Diagnostic-Laboratory-Testing-PCR-Testing/j8mb-icvb
- Vaccinations
  - https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc

Methods

Tableau was used to generate various time series line and bar graphs for the Covid-19 data. Data was filtered by the fifty states to show the U.S. data and by the state of Georgia to show its state level metrics.
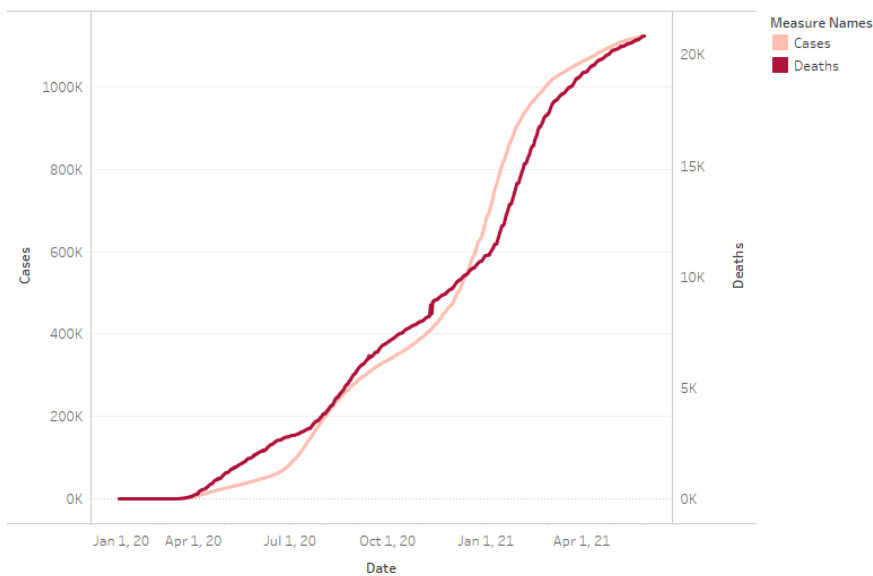
U.S. Total Cases and Deaths
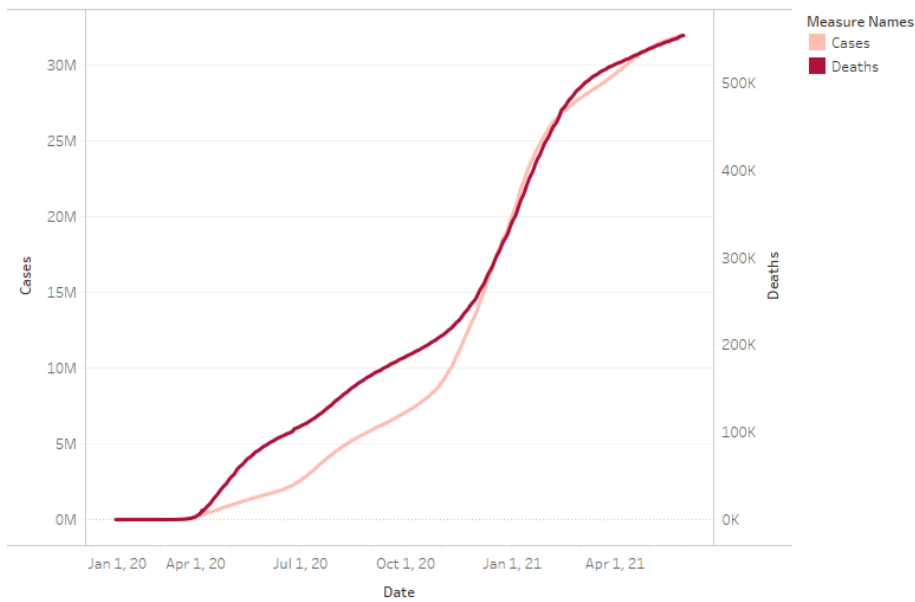
The trends of Cases and Deaths for Submission Date Day. Color shows details about Cases and Deaths.
The data is filtered on State, which keeps 50 of 60 members.

This graph of the running total of U.S cases and deaths shows a very close relationship between the two attributes even though the measures are on different scales. As the number of cases increased, we see similar pattern of increase with deaths.
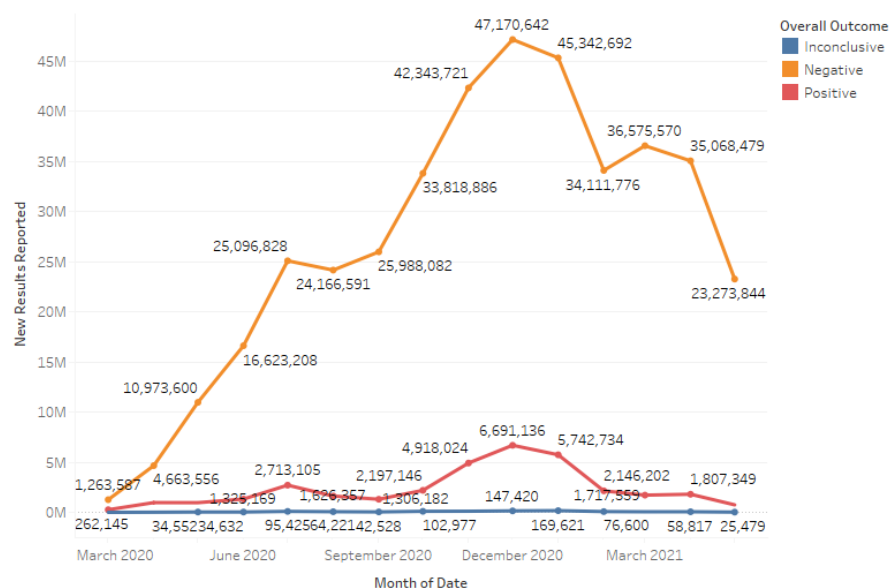
GA Total Cases and Deaths

The trends of Cases and Deaths for Submission Date Day. Color shows details about Cases and Deaths.
The data is filtered on State, which keeps GA.

There is the same positive relationship between the total cases and deaths for the state of Georgia.
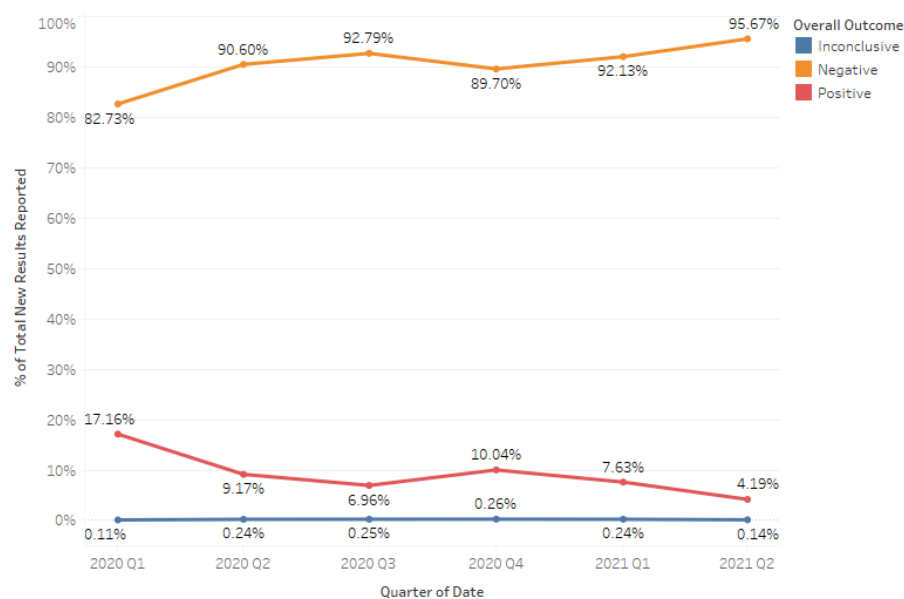
U.S. Covid-19 Tests Administered

Overall Outcome
- Inconclusive
- Negative
- Positive

New Results Reported

45M
40M
35M
30M
25M
20M
15M
10M
5M
0M

47,170,642
45,342,692
42,343,721
36,575,570
35,068,479
33,818,886
34,111,776
25,096,828
25,988,082
24,166,591
23,273,844
16,623,208
10,973,600
6,691,136
5,742,734
4,918,024
2,713,105
2,197,146
1,263,587
4,663,556
1,325,169
1,626,357
1,306,182
2,146,202
1,807,349
147,420
1,717,559
262,145
34,552
34,632
95,425
64,221
42,528
102,977
169,621
76,600
58,817
25,479

March 2020   June 2020   September 2020   December 2020   March 2021

Month of Date

The trend of sum of New Results Reported for Date Month. Color shows details about Overall Outcome. The data is filtered on State, which keeps 50 of 56 members.

There were claims that Covid infections greatly increased because there were more tests being done. We see that from March to September 2020 as more tests were done, the number of positive cases did not increase tremendously. In some periods, the ratio actually decreased.

U.S. Covid-19 Tests Administered

Overall Outcome
- Inconclusive
- Negative
- Positive

% of Total New Results Reported

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

95.67%
92.79%
90.60%
92.13%
89.70%
82.73%

17.16%
10.04%
9.17%
7.63%
6.96%
4.19%
0.26%
0.11%   0.24%   0.25%   0.24%   0.14%

2020 Q1   2020 Q2   2020 Q3   2020 Q4   2021 Q1   2021 Q2
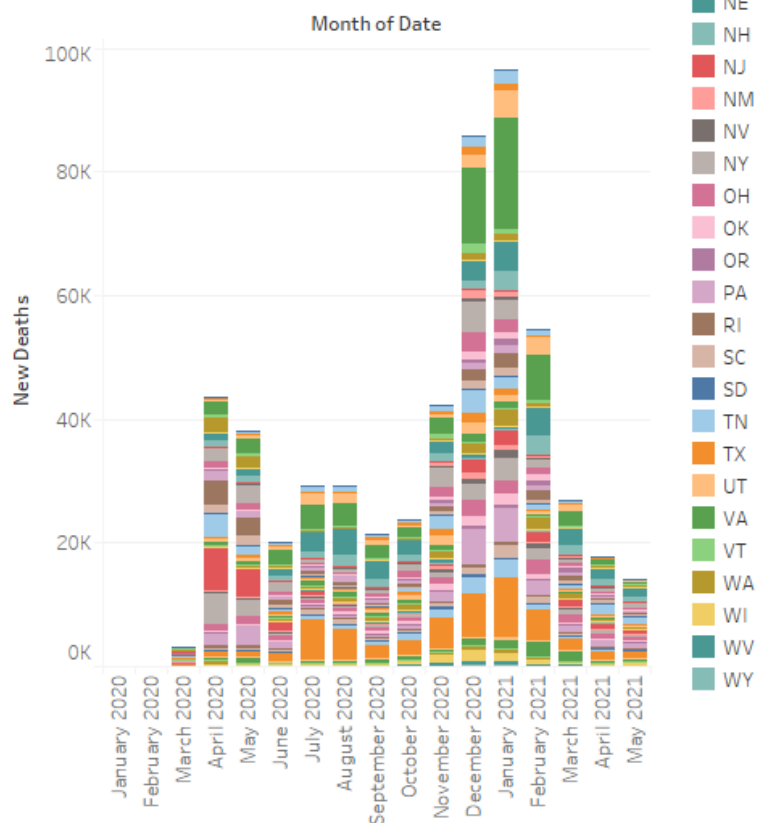
Quarter of Date

The trend of % of Total New Results Reported for Date Quarter. Color shows details about Overall Outcome. The data is filtered on State, which keeps 50 of 56 members.

U.S. New Cases Across States
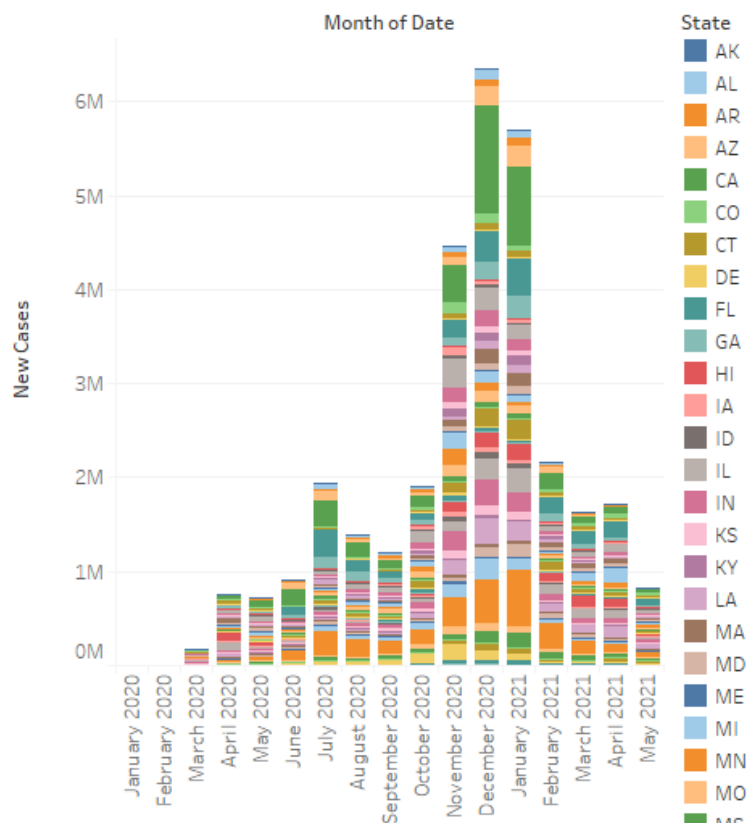
Sum of New Case for each Date Month. Color shows details about State. The view is filtered on State, which keeps 50 of 60 members.

A high number of new cases were reported in November 2020, December 2020, and January 2021, peaking to over 6 million for the month of December. California and Texas were the hardest hit of all the states during those months.

U.S. New Deaths Across States

There is a strong relationship between the number of infections and the number of deaths as the number of deaths follows a similar pattern to the number infections. There were over 80K deaths reported in December 2020 and January 2021, and California and Texas again appear to be the hardest hit. There is a similar pattern in regards to the number of deaths due to Covid.
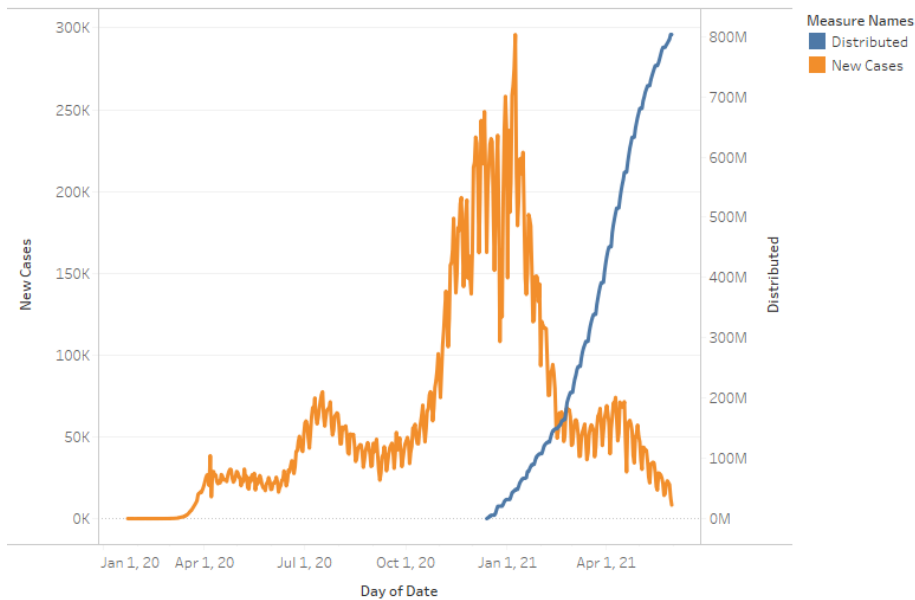
Sum of New Death for each Date Month. Color shows details about State. The view is filtered on State, which keeps 50 of 60 members.

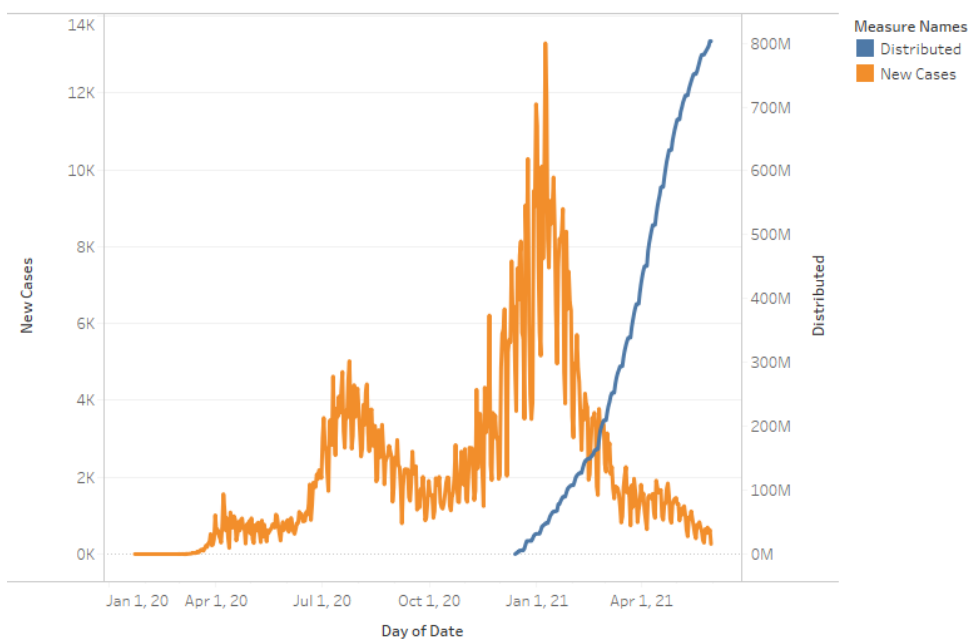## U.S. New Cases and Distributed Vaccinations

The trends of New Cases and Distributed for Date Day. Color shows details about New Cases and Distributed. The data is filtered on State, which keeps 50 of 60 members.

There is a drastic reduction of new cases shortly after the rollout of the vaccination efforts started at the end of 2020 and early into 2021. Towards the end of 2020, there were 100K new cases reported per day. After the vaccination rollout, the number has dropped to under 50K today.

GA New Cases and Distributed Vaccinations

The trends of New Cases and Distributed for Date Day. Color shows details about New Cases and Distributed. The data is filtered on State, which keeps GA.

For Georgia at the state level, we see a similar impact of the vaccination efforts to the number of new cases per day. At its peak the number of new cases reached over 13K per day, but today, the level is below 1K

Results and Conclusion

Covid-19 is a terrible disease and has had a dramatic impact in the U.S. Over 30 million cases have been reported since the pandemic began. Over 500 thousand people have died in the U.S. There is a direct relationship between number of infections and number of deaths. As infections increase, the number of deaths also increases. We saw the highest number of infections and deaths occurring in December 2020 and January 2021. States, such as California and Texas, were hardest hit during this time.

We saw the number of tests conducted increase dramatically over time as the pandemic started to spread. However, the ratio of positive tests versus the number of tests completed did not increase drastically as more tests were done. We saw the number of tests completed peak around December 2020 as the impact of the virus in the U.S. appears to have peaked at that time.

The data supports the good news coming from the effectiveness of the vaccinations created by Pfizer and Moderna. As distributions of these vaccines ramped up earlier this year, the number of new cases both in the U.S. and the state of Georgia began to drop significantly. However, new cases are still being reported as well as new variants of the disease. There are still some countries today where the pandemic is still at its peak. Although we may be able to relax on the social distancing and masking rules due to the success of the vaccines, we should continue to be vigilant and follow the guidelines of the CDC.

Acknowledgements

Thank you to the CDC and other government entities for recording and making this data available. Most of all, thank you to all of our heroes, the front-line workers and first responders, who have helped save many lives during this pandemic.

References

CDC Case Task Force. (2021). United States COVID-19 Cases and Deaths by State over Time.

United_States_COVID-19_Cases_and_Deaths_by_State_over_Time.csv. Retrieved from

https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-

o/9mfq-cb36

Onyeaka, H., Anumudu, C. K., Al-Sharify, Z. T., Egele-Godswill, E., & Mbaegbu, P. (2021). COVID-19

pandemic: A review of the global lockdown and its far-reaching effects. Science Progress.

https://doi.org/10.1177/00368504211019854

Immunization Information Systems (IIS). (2021). COVID-19 Vaccinations in the United States,Jurisdiction.

COVID-19_Vaccinations_in_the_United_States_Jurisdiction.csv. Retrieved from

https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-

b7fc

U.S. Department of Health & Human Services. (2021). COVID-19 Diagnostic Laboratory Testing (PCR

Testing) Time Series. COVID-

19_Diagnostic_Laboratory_Testing__PCR_Testing__Time_Series.csv. Retrieved from

https://healthdata.gov/dataset/COVID-19-Diagnostic-Laboratory-Testing-PCR-Testing/j8mb-icvb