COVID-19 Vaccination Efficacy

Kayla Thompson

DSC 680 - Fall 2021

https://github.com/kaymaethomp1

Abstract

In this project I plan to take a closer look at the COVID-19 pandemic and efforts that are on-going to curb the impact of this virus. Through my research and analysis, I plan to compare the rates of vaccinations (first shot, fully vaccinated, booster shots) with the rates of infection and the death rate in several of the world's most populated countries. A correlation map will also be used to look at potential correlations in disease impact and mitigation efforts. The benefits of doing this analysis include the potential to clear up any misinformation circulating as well as provide a framework for data informed decision making with regards to COVID-19. I plan to use Python to generate visualization and statistical values to either support my hypothesis that higher vaccination rates lead to lower COVID-19 infection and deaths or support the null hypothesis that vaccination does not impact these values.

Problem Statement

Almost two years into the COVID-19 pandemic we know that COVID-19 is highly contagious, and we know the effect that this disease has had on the world. After losing 3.8 million people to COVID-19 globally, this virus has caused one of the most consequential global health crises since 1918 [1]. This type of destruction has left the globe desperate for a fix and that fix is believed to be found in the three COVID-19 vaccines. The vaccine is believed to prevent the transmission and severe illness associated with COVID-19. The three vaccines currently

available are Pfizer-BioNTech, Moderna and Johnson and Johnson. Pfizer-BioNTech and Moderna are both MRNA vaccines. These vaccines deliver instructions to our cells to make the S protein found on the surface of the virus. Once our cells make the protein, our body detects this protein and creates antibodies to destroy it. Thus, when COVID-19 infection occurs, our body already has a blueprint to fight the virus. The Johnson and Johnson vaccine is a vector vaccine. This means that the S protein genetic material from COVID-19 is placed in a modified version of a different virus (vector) and that materials is delivered to our cells upon injection. Then, similarly to the MRNA vaccine, our cells make the protein, that protein is detected by our immune system and then our immune system creates antibodies for that protein [2]. All of this is a good thing but has been clouded by bouts of misinformation and questions about vaccine efficacy. In this study, I am to answer the question – do the vaccines help to prevent COVID-19 infection and death?

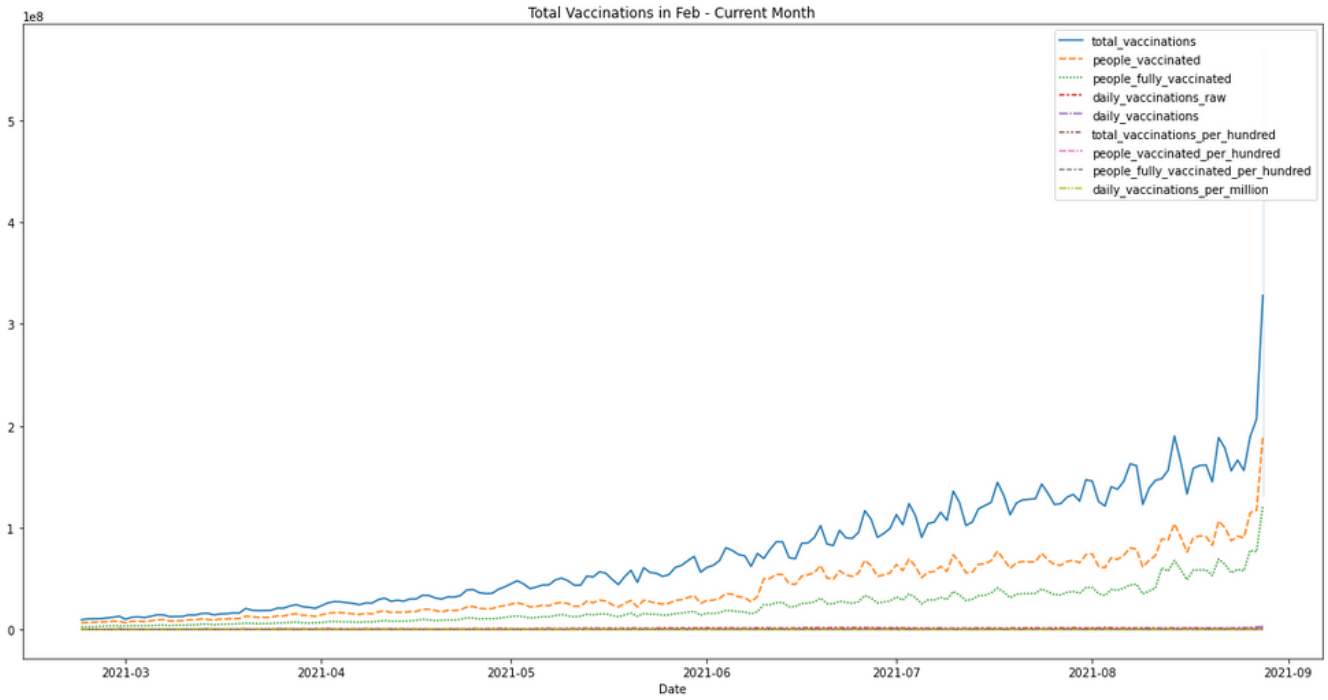
Methods

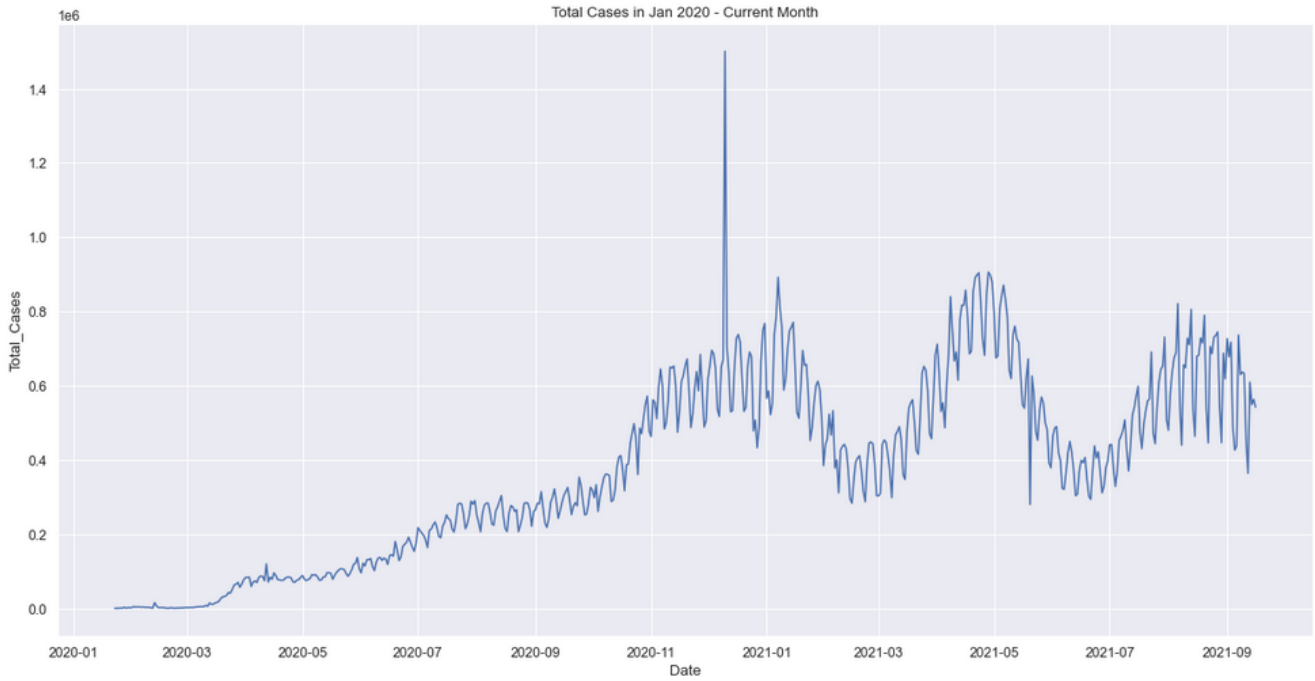
To begin the analysis, I uploaded three data sets into pandas data frames. These data frames represent the number of positive COVID-19 cases globally by country, the number of deaths globally by country and the number of vaccinations administered globally by country. All this data is also tagged to a date, so we can look at trends over time using this information. Additionally, I loaded in global meta data that can be used for visualizations if necessary. I began my exploratory data analysis with some data clean up. The most important of which was to convert the dates in each data frame to date time format and to make the date the index in the data frame. This ensured that all dates are consistent and allows for the flexibility needed to combine multiple data frames together. Once the data was clean, I began creating

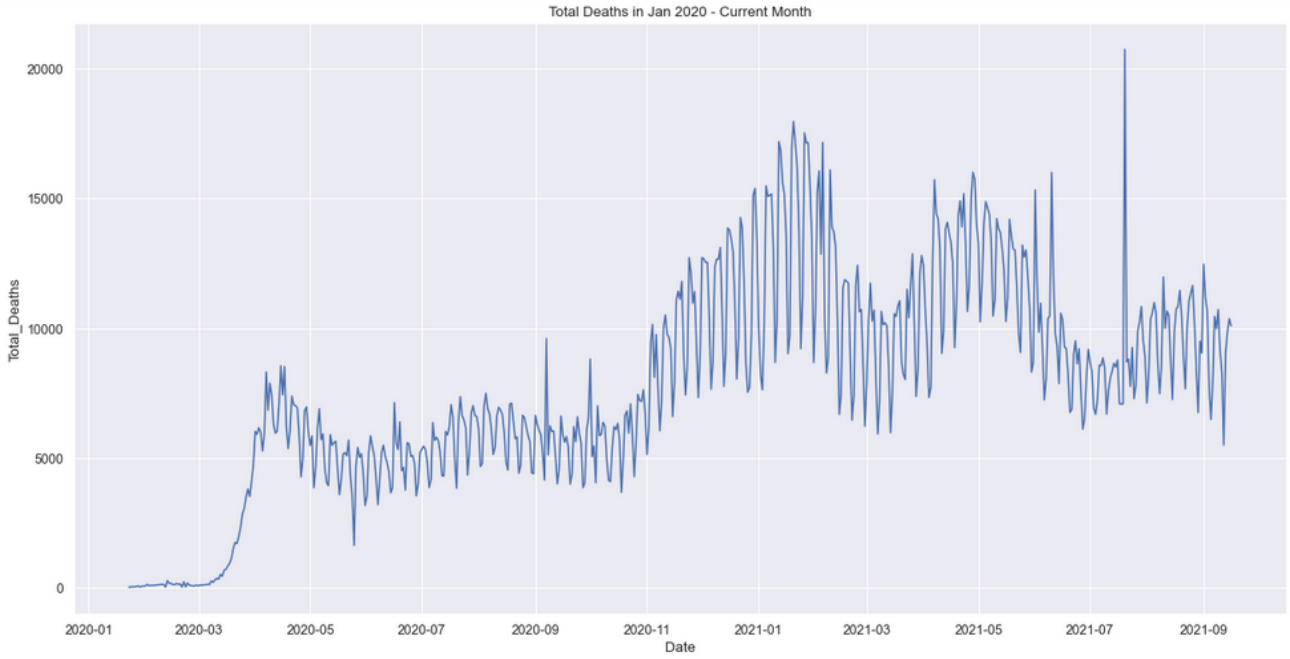
visualizations to begin the comparison of infection, death, and vaccination. Before breaking down any of the data by country, I looked at the global positive case total, global death count and the total of fully vaccinated people globally. To create these visuals, I used MatPlotLib and Seaborn. After creating these initial visuals, my next step was to combine the most important information from these data frames into one data frame for further analysis. In order to merge the three data sets, I used the reduce function from functools. I joined all three datasets on the date column. This created a data frame with five hundred and seventy one columns. In order to keep this analysis efficient and feasible, I trimmed the data frame down to only include total cases, total deaths, total vaccinations and the count for fully vaccinated people. I then visualized all of these data points over time in using a line graph. This line graph will need to be improved, currently it is on a log scale which makes interpreting it difficult. Additionally, the total cases and total deaths do not seem to be plotting appropriately. My last visualization is a correlation matrix that I created using seaborn. Once the visualization was generated, it was apparent that I need to further segment the data to solve for my problem statement. Before my next submission I will need to accomplish several things to solve my problem statement and to answer my submitted questions. First, I need to determine a better way to plot the values in my merged data set. This visualization is not currently suitable for interpretation. My correlation matrix is currently very skewed as initial vaccination totals did not pick up for several months once the pandemic started. The number of null values in the total vaccination column has skewed my correlations between total cases, deaths, and vaccinations. I believe looking at a smaller, more representative time frame will help as well as breaking down the

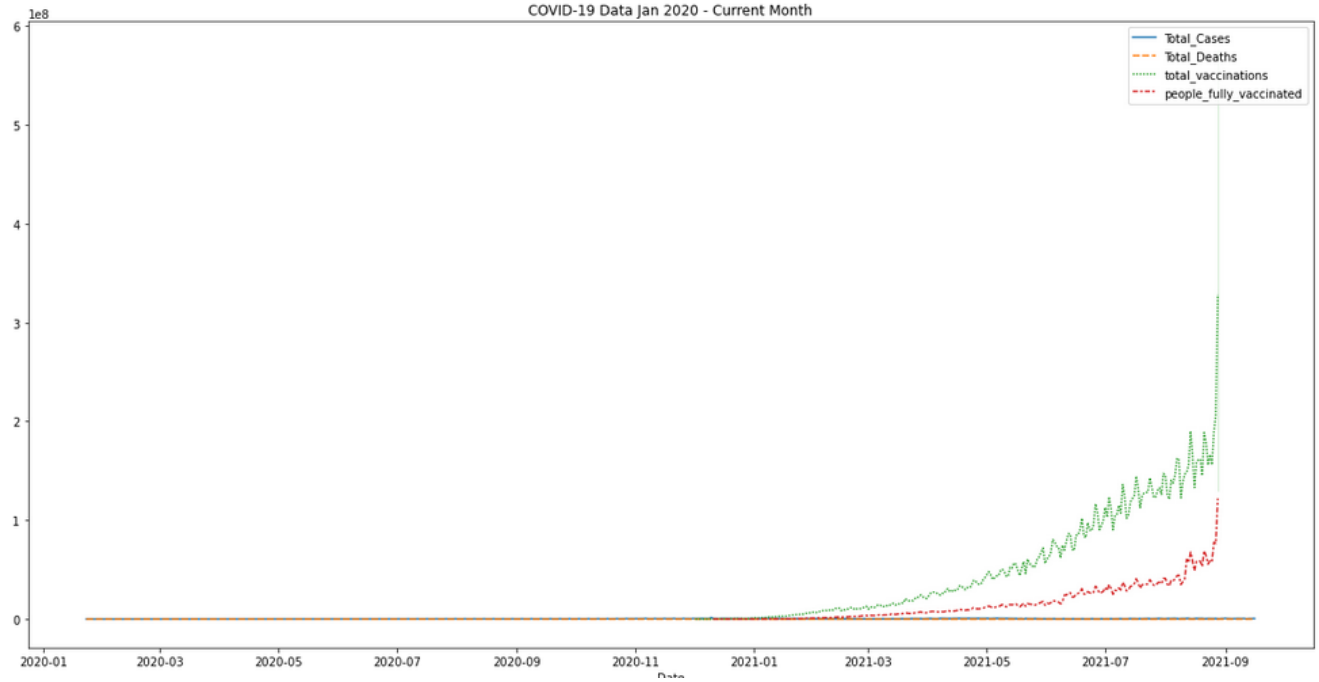
data by countries with high case counts, high death counts and high vaccination counts. This will be completed in future work before final submission.

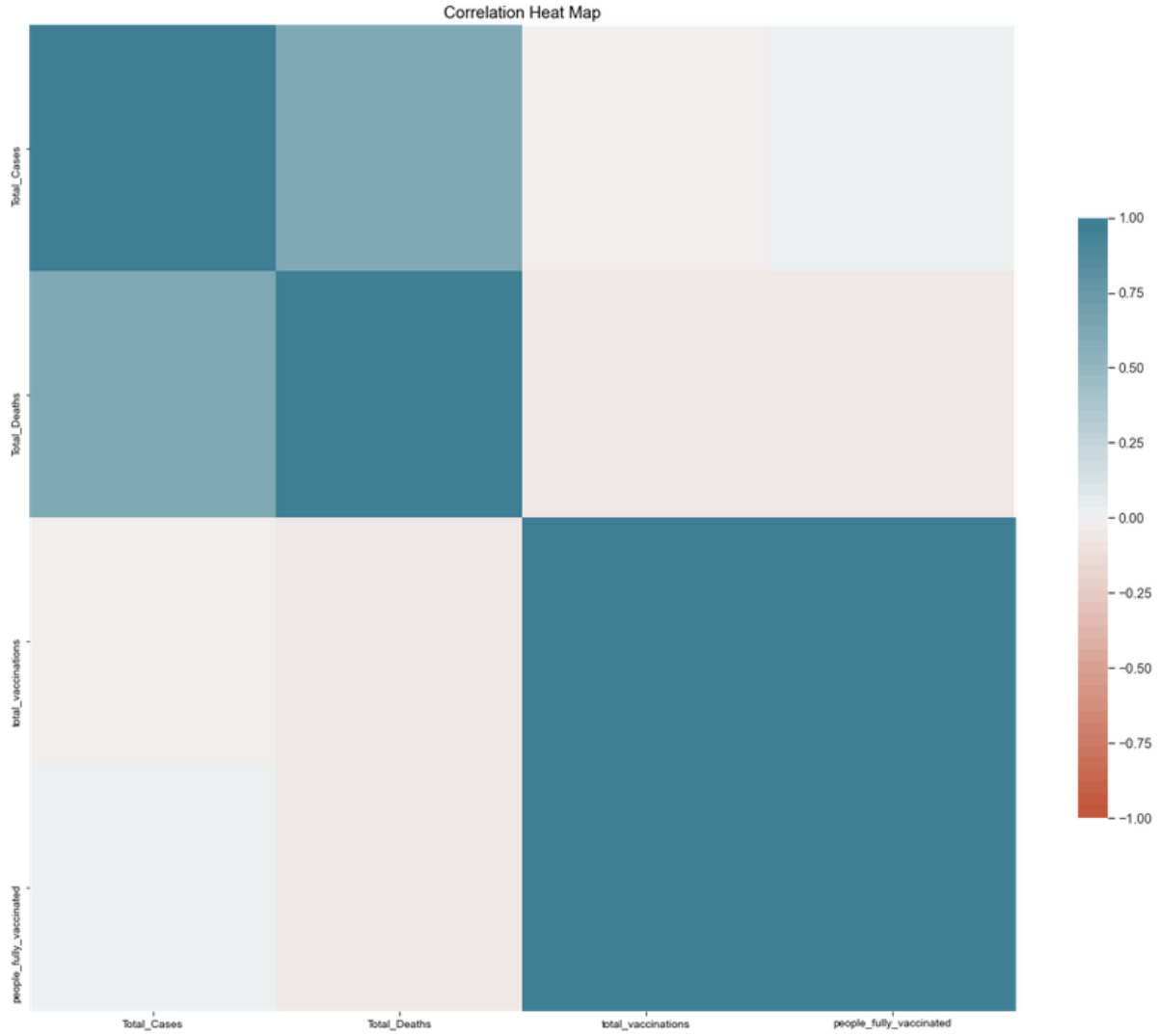
Appendix






Total Deaths in Jan 2020 - Current Month
Total_Deaths
Date
0
5000
10000
15000
20000
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09


COVID-19 Data Jan 2020 - Current Month
1e8
Total_Cases
Total_Deaths
total_vaccinations
people_fully_vaccinated
0
1
2
3
4
5
6
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09



Conclusion

Pending.

References

[1] Cascella, M. (2021, July 30). *Features, evaluation, and treatment of Coronavirus (COVID-19)*. StatPearls [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK554776/.

[2] Mayo Foundation for Medical Education and Research. (2021, August 27). *How do different types of COVID-19 vaccines work?* Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/different-types-of-covid-19-vaccines/art-20506465.

COVID-19 Vaccination Data: https://www.kaggle.com/priteshraj10/covid-vaccination-all-countries-data
COVID-19 Infection Data: https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university

Additional References

Centers for Disease Control and Prevention. (n.d.). *Key things to know about covid-19 vaccines*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/keythingstoknow.html?s_cid=10493%3Acovid+19+vaccine%3Asem.ga%3Ap%3ARG%3AGM%3Agen%3APTN%3AFY21.

Centers for Disease Control and Prevention. (n.d.). *Understanding mrna covid-19 vaccines*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/mrna.html.

Commissioner, O. of the. (n.d.). *FDA approves FIRST Covid-19 Vaccine*. U.S. Food and Drug Administration. https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine.

Coronavirus disease (covid-19): What is it, symptoms, causes & prevention. Cleveland Clinic. (n.d.). https://my.clevelandclinic.org/health/diseases/21214-coronavirus-covid-19.

Elamroussi, A. (2021, September 4). *Already vaccinated against COVID-19? Experts say you're protected, even without a booster shot*. CNN. https://www.cnn.com/2021/09/04/health/us-coronavirus-saturday/index.html.

Mayo Foundation for Medical Education and Research. (2021, August 24). *Coronavirus disease 2019 (COVID-19)*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963.

Sauer, L. M. (n.d.). *What is Coronavirus?* Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus.

World Health Organization. (n.d.). *Coronavirus disease (COVID-19)*. World Health Organization.

https://www.who.int/emergencies/diseases/novel-coronavirus-2019