DATA 472 Project Report

Hien Nguyen, 300199540

Due: 8 June 2021

1. Introduction and Motivation

Since COVID-19 was discovered in December 2019, over a year ago, it is still going strong as one of the biggest global pandemics, making thousands of people fall sick or die every day, causing various cities to lock down, stopping millions of travels around the World, and causing financial distress to many. Thus, it is in everyone's best interest to create vaccines in response to this pandemic, and as of April 2021, 13 vaccines are now authorized for public use. However, rolling out the vaccines to everyone is a lengthy process as much as it is a chance to keep track of its distribution and see how it impacts the spread of the virus.

The COVID-19 application was built to provide a quick snapshot of the current situation. How the pandemic evolves is measured mainly by the number of cases, number of tests, and number of deaths. The rolling out of vaccinations will also be visualized. In particular, the app will showcase how some potential factors influence COVID-19 cases and death rates and how vaccine hesitancy and the affordability of the vaccines can be the major hindrances in the fight with the current COVID-19 pandemic.

The wide variability in the number of deaths/cases among different countries has fueled the interest in what could cause such variations. Pointing out risk factors for severe illness and death could identify vulnerable groups and shed light on choosing the proper practices to minimize the transmission. The potential factors influencing COVID-19 cases and death rates are grouped into four main categories: population characteristics, environmental/geographic factors, healthcare policy, and virus-related factors (Hammad et al., 2020). However, this app will mainly focus on population characteristics like age, handwashing-facility, population density, and health care policy like testing rate due to the limitation of accessing published raw data on the internet.

In many countries, vaccine hesitancy and misinformation pose a substantial obstacle in achieving community immunity coverage (Sallam, 2021). Therefore, this app also aims to visualize any link between the vaccine acceptance rate and the number of people fully vaccinated per hundred. Furthermore, another well-known challenge in ensuring global access to COVID-19 vaccines is affordability (Wouters, 2021). For this reason, we will also look at if there is a correlation between GDP per capita and total vaccinations per hundred.

The World is a big place, and there is a large variability in terms of the above stats among different countries and regions. Therefore, representing an overall world map as well as breaking it down to specific locations is necessary. In addition, the application will provide multiple widgets and filters to help users navigate through desired dates, continents, or countries.

2. Data Overview

Data was taken from 3 sources, PubMed Central® (PMC) (a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7920465/, Our World in Data: https://github.com/owid/covid-19-data/tree/master/public/data/ and Worldometers: https://www.worldometers.info/coronavirus/.

PMC provides the latest results of the COVID-19 vaccine acceptance rates in different countries. This research focused on different groups, including the general population, nurses, healthcare workers, university students/staff, parents, and guardians. However, only a few countries were recorded, and not all of them have the data for all the listed groups. With the general population, this would still provide valuable insight into the general perception of vaccine acceptance. My app focuses mainly on whether vaccine acceptance links with the number of people fully vaccinated, not the change in people's perception about Covid-19 vaccines. Therefore, only the latest results are retained. One thing worth noting is that the date of the survey varies from country to country, between April 2020 to October 2020, and out of approximately 200 countries and regions around the World, only 29 countries are available in this study.

The second source is Our World in Data (OWID). OWID collects information from various reliable sources with clean, processed data and ready to be used. For this reason, this one plays a significant role and provides almost all the data represented in this report. Another good thing about it is that it contains statistics for every day since the beginning of the outbreak for almost every country, allowing us to explore how the pandemic evolves on different scales and parameters. OWID also provides a link to enable downloading the data straight to R. As mentioned before, OWID's data includes the date variable, thus "as.Date" was used to convert data into date format in R. Also, with the purpose of including a map of vaccinations around the World, the combination of OWID data and polygons data was essential. Though I could use the country codes as the reference to merge data from two sources to one source, I decided to use the country names as the reference because vaccinations map was not the only map I would like to draw. Representing the number of COVID-19 cases in a big picture is also necessary, and the number of cases that I took from Worldometers does not have the country codes; thus it looked like using the country names as the reference was a reasonable approach. The shape file, which included countries' polygons, was provided by Bjorn Sandvik, thematicmapping.org, and included more countries than other data sources, so it was better to use it as a base file, then merge it and others. In order to do that, they both needed the country names to match, so after actioning the joint, I had to figure out which country's name was miss-matched by looking for NA using "is.na" function and then changed the names accordingly.

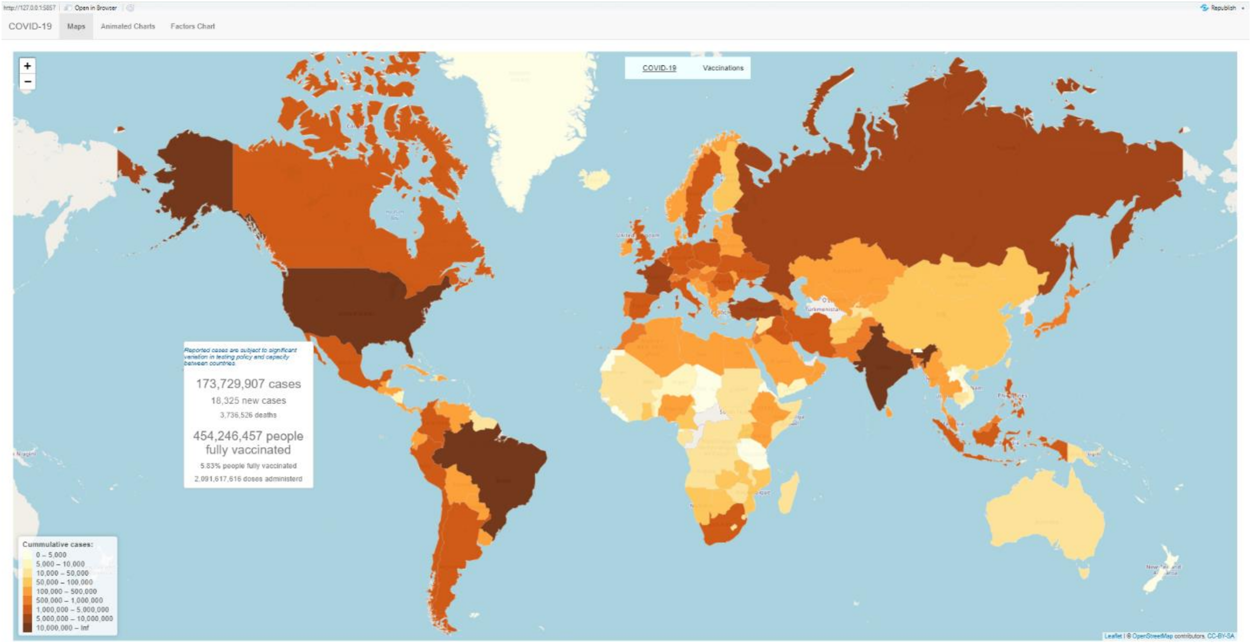
On the other hand, data recorded daily makes visualizing the data without date feature a little harder because many static stats are duplicated, and so getting only the total number is not intuitive. Fortunately, in terms of total cases, total tests, or total deaths, the number should continuously increase and not decrease over time. This makes thing easier to filter out the other values by using "which.max" function. Other parameters that were used in this report, such as "aged_65_older”, “handwashing_facilities”, "population_density" and "gdp_per_capita", are recorded the same for every day; therefore, it does not matter which function we use as long as we can pick one number out of the duplicate values. Due to some graphs used data that was not based on date, a new data frame was needed to store the latest or the maximum value for each parameter correspondent to each country. OWID provided data on not only country level but also continent and World Wide levels; therefore, sub-setting those were compulsory when representing data on the country level.

Though OWID provides lots of relevant information for this report, it lacks one parameter that I'm interested in: the active cases that Worldometers has on their website updated daily. Luckily, the data is stored as a table, so all I needed to do was read that table into R using package "rvest" and "tidyverse". Worldometers also has the total number of cases, the total number of deaths, and the total number of tests available, which was handy for my COVID-19 world map. When drawing this map, I had the same problem as drawing the vaccines map, where some country names on Worldometers table different from the polygon table, so the same fix was applied for this one, like with OWID data. Table scouted from the web often has an incorrect data type. In this case, I had numbers stored as string due to the comma separator, hence "gsub" and "as.numeric "were used to fix this problem.

3. App Overview

This application aims to give insights on the current global pandemic COVID-19 with a fresher perspective. We could get a quick update on the number of cases, tests, deaths, and vaccinations just after a quick search, however, information is usually not gathered in one place. Moreover, the impact of various factors on cases/death rates and vaccinations has not yet been widely visualized and discussed. As such, this application should serve as a dashboard to represent those. I always believe that less is more, and the best user experience is the less click the better. Therefore, everything should represent itself. While the interactive application is a good approach to increase user experience, exploiting it might backfire on the original purpose.

As mentioned above, the main purpose of this application is to give a quick update on the pandemic and explore how various factors affect the cases/death rates as well as the rolling out of vaccinations globally. With this in mind, there will be two Worldwide maps, being one represents the overall COVID-19 pandemic, the other one represents the vaccinations. Users will have a widget to freely switch between 2 maps, with the set of metrics changing according to which option is chosen. Additionally, color gradients will represent the scale of the cumulative cases or the number of people fully vaccinated. These scales will vary country by country. Users can also place their mouse cursor over some countries to see more details such as "active cases ", "cumulative cases ", "total cases per 1M population" or "doses administered", "number of people fully vaccinated ", and "percentage of people fully vaccinated". The map will allow users to zoom in and zoom out to any specific location freely. If anyone is curious about the whole World's total number, there will be an absolute panel with some of the World's metrics on it. This panel can be moved around to not get in the user's way while navigating around the map.

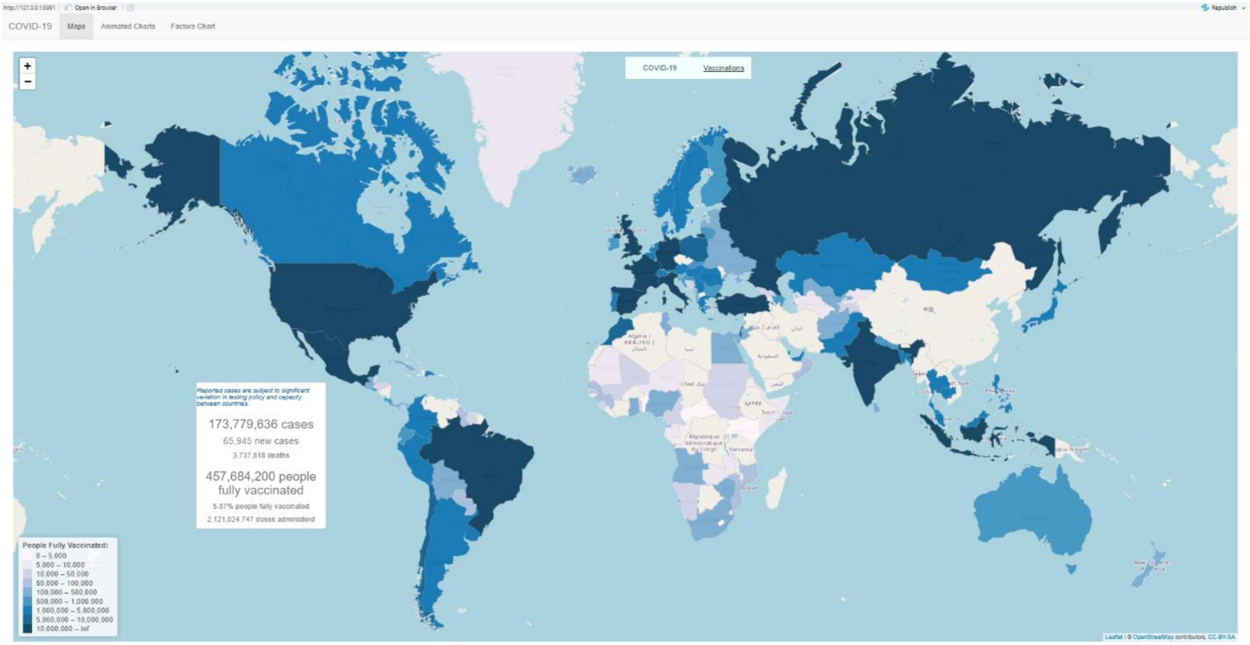


Reported cases are subject to significant variation in testing policy and capacity between countries.

173,729,907 cases
18,325 new cases
3,736,526 deaths

454,246,457 people fully vaccinated
5.83% people fully vaccinated
2,091,617,616 doses administerd

The color scales are based on cumulative cases for the COVID-19 map and the number of people fully vaccinated for the Vaccinations map.



Reported cases are subject to significant variation in testing policy and capacity between countries.

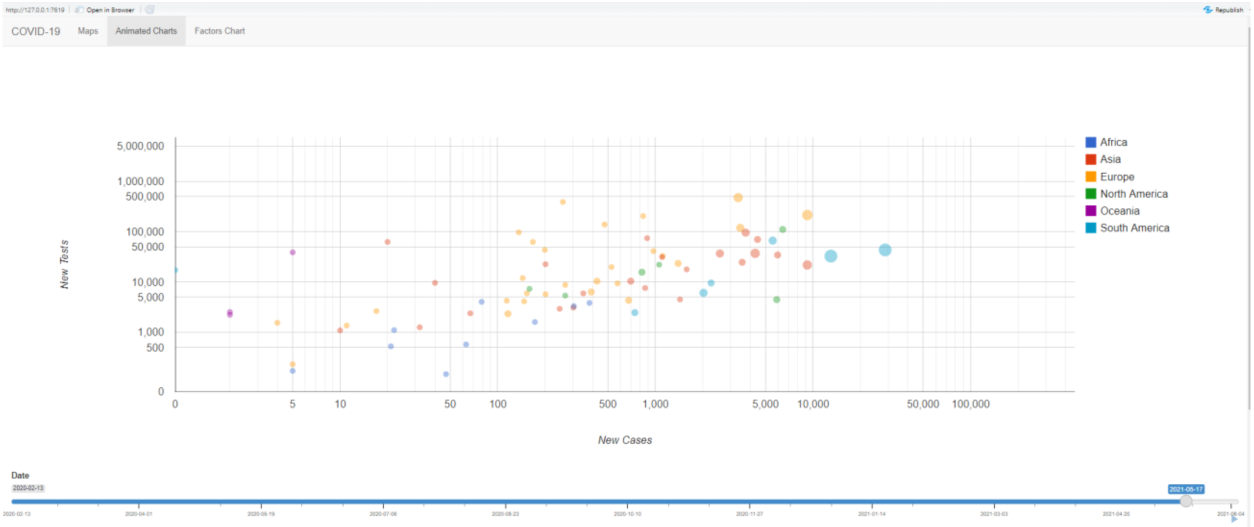
173,779,636 cases
65,945 new cases
3,737,818 deaths

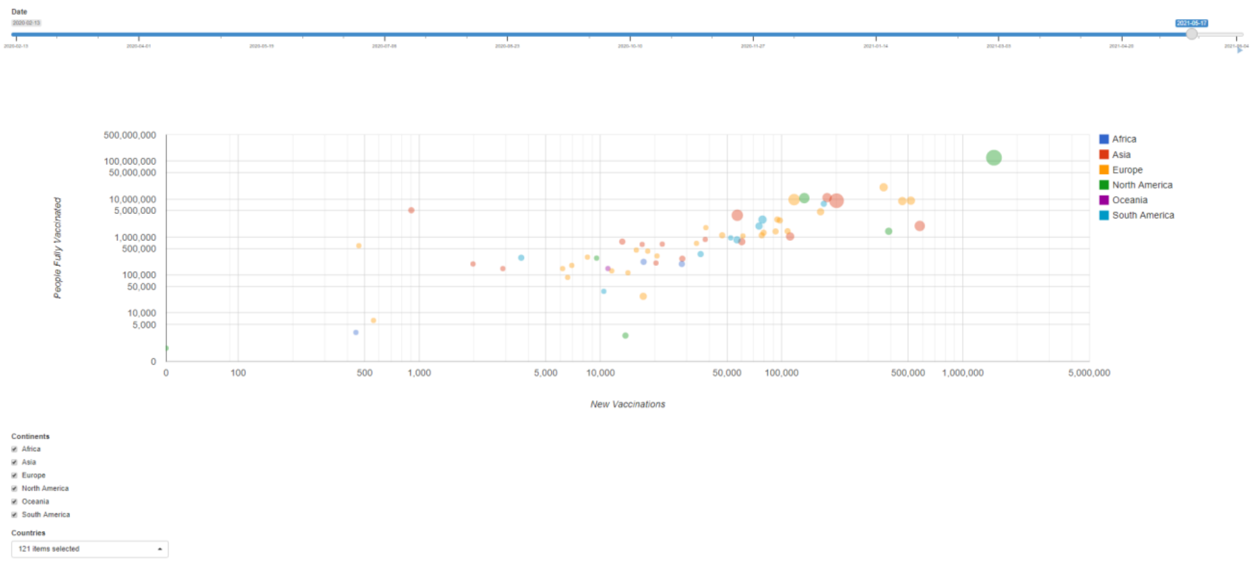
457,684,200 people fully vaccinated
5.87% people fully vaccinated
2,121,024,747 doses administerd

Color scales could also be based on the total cases per 1M population or the percentage of people fully vaccinated. Countries with a larger population often have more cases than the rest. However, with a wide variety of countries worldwide, these two metrics will lessen the big gap in the color gradients and hence lessen the efficiency of a World map visualization. The map will look rather monotone, thus it gives the impression that the situation is similar around the world.

When we have data recorded over time, it is always good to look at how things change as time passes, especially when the spread is fast, the outbreak could be enormous, and the change from one day to the next is very unpredictable. As a result, second tab showcases two animated bubble charts, where users can either click the “play” button, and they will play themselves. Alternatively, users can drag the slide to the desired date, and the chart will change accordingly. For the first chart, x represents number of new cases, y represents new tests, where the bubble size represents the new deaths. The number of new vaccinations, number of people fully vaccinated, and population are the x-axis, y-axis, and bubble size respectively for the second animated chart. There is a considerable variation between some countries, thus, both x and y are log-scale. The bubbles are categorized into six different colors corresponding to six different continents. Like with the maps, if users move the cursor to the bubbles on the charts, the country name and its metrics will pop up to show more details.

Furthermore, one chart with all six continents may be rather busy at some points. Users then could choose to filter out undesired continents/countries—the continents/countries filter and time slider widget control both COVID-19 and Vaccinations animated charts to avoid duplications. Regarding the tab level filter, the time slider holds the highest level, following Continents and then Countries has the least control power among the three.





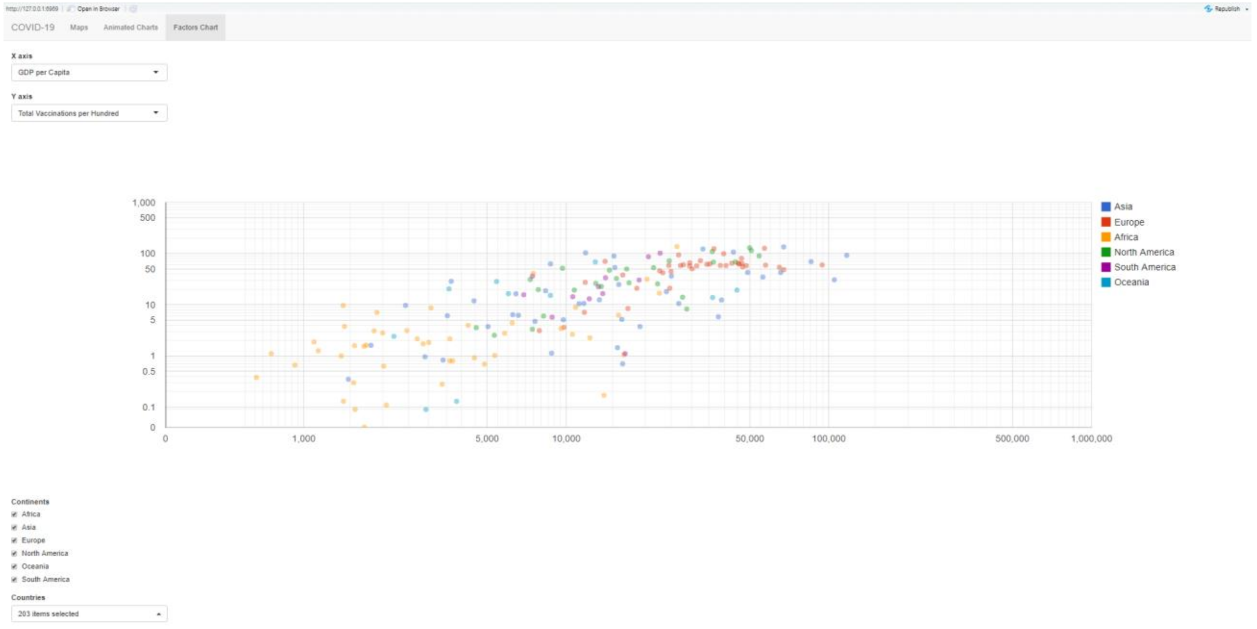
Continents

☑ Africa
☑ Asia
☑ Europe
☑ North America
☑ Oceania
☑ South America

Countries

121 items selected

The third tab, called "Factors Chart", is reserved to visualize the factors that affect the transmission/ death rates of COVID-19, the number of people fully vaccinated, and the total vaccinations allocated to each country. Again, a bubble chart is utilized, however unlike the animated chart where the x variable and the y variable are fixed, for this chart, users can choose those variables from two drop-down lists. There are ten options; to gain more insights on those factors, the most intuitive selection for x would be "Share of 65 Years and Older", "Hand-washing Facilities", "Population Density", "Total Tests per Thousand", "COVID-19 Vaccination Acceptance Rate" and "GDP per Capita". As for y, they could be "Total Deaths per Million", "Total Cases per Million", "Percentage People Fully Vaccinated" and "Total Vaccinations per Hundred". Nonetheless, users could choose any other combinations to gain other insight.



X axis

GDP per Capita

Y axis

Total Vaccinations per Hundred

Continents

☑ Africa
☑ Asia
☑ Europe
☑ North America
☑ Oceania
☑ South America

Countries

203 items selected

4. Insights

Searching for the data, processing, and transforming them consumed half of my time. During the process, I learned web scraping, transforming data using self-written functions as well as other available methods such as "sapply" to help with data processing. The other half-time was all about learning how to incorporate R Shiny with my vision, which is very challenging. Though drawing a World map required much research to understand how it works, R Shiny was much more complicated. Fortunately, there was also much inspiration on the internet that helps with ideas developing, and at the same time, it also pushed me further to learn new things such as CSS and "googleCharts" since the app would look so much better with their help.

In terms of researching for the data, there were so many available sources on COVID-19, it was not straightforward to find one that had information on the pandemic itself and on social or health statistics around the World. In addition, it would be much more complex and time-consuming to collect data from different sources and combine them myself.

Map drawing also posed a new challenge, and I needed to extensively research and read on how “leaflet” function and “addPolygons” function work, it was confusing at first since they are inseparable, and there were so many elements to play around with. When it came to visualizing the World map, one of the most popular examples was to use country population as value. The population came with the shape file provided by Bjorn Sandvik, thematicmapping.org, but I needed different metrics to represent COVID, and at the time, I did not know just using the "merge "function could help with that, and it took me a while to figure it out.

However, the Shiny aspect of building this application was the most difficult part. The package's syntax looked simple, but it made things complicated to change or customize according to my need. Therefore, I had to learn a bit of CSS and tried to learn from available examples on R Shiny gallery. Generating the maps in R is only half of the task, the goal is to visualize them in R Shiny under one tab. At first, I thought of using the select box widget but later realized the radio buttons widget with some CSS touch would make the switch for the two maps looked nicer. Before knowing of CSS, I struggled a lot with getting the switch button to show up on the app and did not know how to call out the right map corresponding to the user's input until I stumbled on the function “get”.

Regarding the animated chart, R had a package called “gganimate” but to make things work in R Shiny, it was better to use “googleCharts”. Nevertheless, with lots of arguments to play with, “googleCharts” was hard to understand at first.

There are many other chart types that I have not tried. Additionally, although I am a big fan of minimalism and minimum click, there is still room to add more widgets and filters. For example, the animated chart can only show one day at a time, and it does not give the users option to choose a longer time frame. Log-scale switch could also be a potential widget. In the case of the factors chart, it would be interesting to also look at the data and find out if there are any interactions between those factors on the number of cases/deaths and plot them. Various factors influence the spread of COVID-19, which was not mentioned in this app, it would be nice if they are also included. I would also like to publish this app on R Shiny server and get it to update daily at a specific time automatically.

5. Conclusion

This project was a fun one to work on. COVID-19 is a hot topic around the World. Building an application on this dataset is relevant now more than ever. Through this project, I could see how this pandemic affects us on a bigger scale, at the same time, it does not fail to give insights on a smaller scale. GoogleCharts is a good find, and it incorporates very well with R Shiny. However, the app still needs some improvements, especially in terms of log-scale switch or interaction factors visualization for a more specific user experience.

6. References

Hammad A., O., et al. (2020, July 17). *Factors Influencing Global Variations in COVID-19 Cases and Fatalities; A Review*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7551068/

Sallam M., (2021, February 16). *COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7920465/

Wouters J, O., et al. (2021, February 12). *Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment*.

https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00306-8/fulltext