

Covid-19 Vaccinations - Final Project

Anitha Ganapathy, Keerthana Reddy Varakala, Saymon Mameza

Dr. Ahn

D590: Data Visualization

Luddy School of Informatics, Computing and Engineering, Indiana University

22 July 2022

CONTENTS

Abstract	3
Introduction	4
Background	4
Motivation	4
Objective	5
Existing Visualizations	6
Process	11
Failed Experiment	14
Results, Insights, and Methods	15
Discussion, Future Work, and Conclusion	32
References	32

Abstract

The first case of COVID-19 virus was detected in Wuhan City, Hubei Province of China. Covid-19 was declared as a pandemic in March 2020 and since then, it's been more than 2 years. The quick widespread spread of the disease has affected the way of life in every aspect known to us. Many preventive measures have been undertaken by the World Health Organization (WHO) and Center of Disease Control (CDC) to prevent and slow down transmission by providing details to the public about the disease itself and how it spreads. Much vaccine research was undertaken, and vaccines were developed within 12 months from the identification of the first case of the novel coronavirus. Many companies were in the race to create the vaccine for the virus and aid in bringing the epidemic to an end. The top companies being Johnson & Johnson, AstraZeneca, Moderna, BioNTech, Pfizer and Novavax to name a few. The vaccines created must satisfy the diversity in population, global access and provide immunity against the variants. Further research and trials are still ongoing to create a vaccine that can provide immunity against various symptoms of the novel coronavirus.

Keywords: COVID-19, SARS-CoV-2, types, vaccine, pandemic, coronavirus

Introduction

Background

Coronavirus disease 2019 (COVID-19) is a virus that was discovered in December 2019 in Wuhan China (CDC, 2021). COVID-19 originates from SARS-CoV-2 which are all a part of the coronavirus family. All viruses that are a part of the coronavirus family are considered a repository virus which spreads rapidly through little droplets that come from the mouth or nose (CDC, 2021). It can affect humans and other animal species.

The root word coronavirus which is corona gets its name because of the appearance of the spike protein that sticks out. The spike protein that is a part of the virus sticks to the human cell and affects it. As time goes on, the virus can mutate which causes the spiked protein to change. The process of mutation is called genomic surveillance. This can make finding a vaccine difficult, and make current vaccines useless in certain circumstances.

The virus may gradually evolve into a seasonal low-level epidemic. Even if the virus can be completely eliminated from the population, the transmission mechanism from the host to the person is still unclear due to the population's general susceptibility. There is a risk of recurrence or periodic epidemics. With the risk of occurrences, the development of vaccines are crucial. Currently, there are over five COVID-19 vaccines worldwide. The effectiveness rate ranges from 80% to 95% during a certain period.

Motivation

Covid-19 is a virus that affects the respiratory system of an individual. In July 2022, over a half a billion people have been infected and over six million deaths have been reported. Of the six million

reported deaths, the top three countries account for one third total deaths and ten percent of infections. This is astonishing considering that the virus has only been around for two years. When the pandemic first started in March 2020, almost all countries were in strict lockdown. Some countries such as China, Italy, and parts of the USA made staying indoors mandatory because of the spikes in cases. This ultimately led to fear and uncertainty in people. Other repercussions of the pandemic was devouring the mental health of individuals that were isolated indoors. According to the Kaiser Family Foundation (KFF), 41% of adults experienced symptoms of anxiety and/or depressive disorder.

In December 2020, the Pfizer vaccine was approved under an emergency procedure to fight against the virus. It is estimated that more than 12.2 billion doses have been given and over 4.5 billion people have been vaccinated. For every 100 people in the world, 156 shots have been administered. Most of the vaccinations have been done in countries such as China, US, and India.

Objective

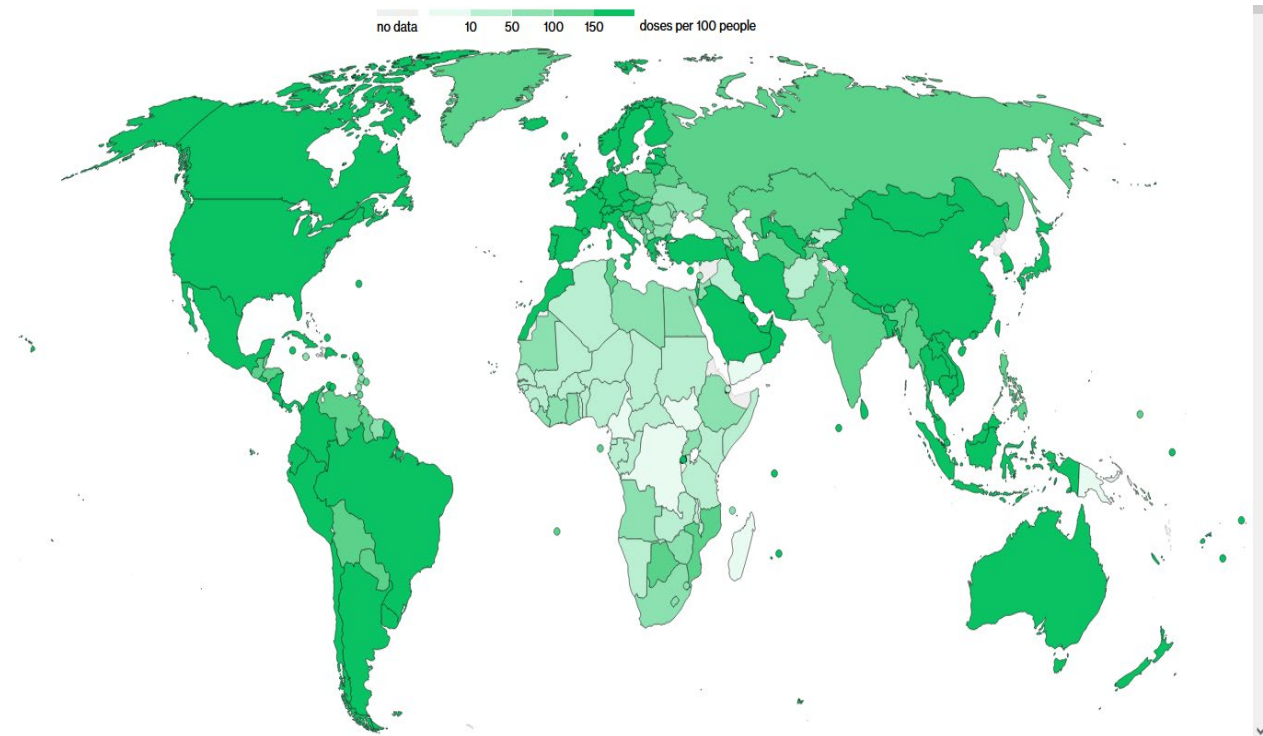
Overall, COVID-19 has put the world into shock and fear. It has killed and affected millions of people around the world. The objective of this paper is to analyze the current and past Covid-19 infections and deaths. We believe that analyzing the data can help us explore how the virus spreads and the countries that are most vulnerable. When knowing that information, it could lead to preventative actions that can be in place to prevent another future pandemic.

Another outcome we hope to achieve is to analyze what countries were hit the hardest and what measures did they put in place to slow the spread of the virus. This would allow us to put together a better action plan for future outbreaks. Also, connecting the link between spreads

of the virus and other diseases would be a big help to researchers. The overall objective is to ensure an outbreak like this never happens so the world would not be in a static shock.

Existing Visualizations

Figure 1

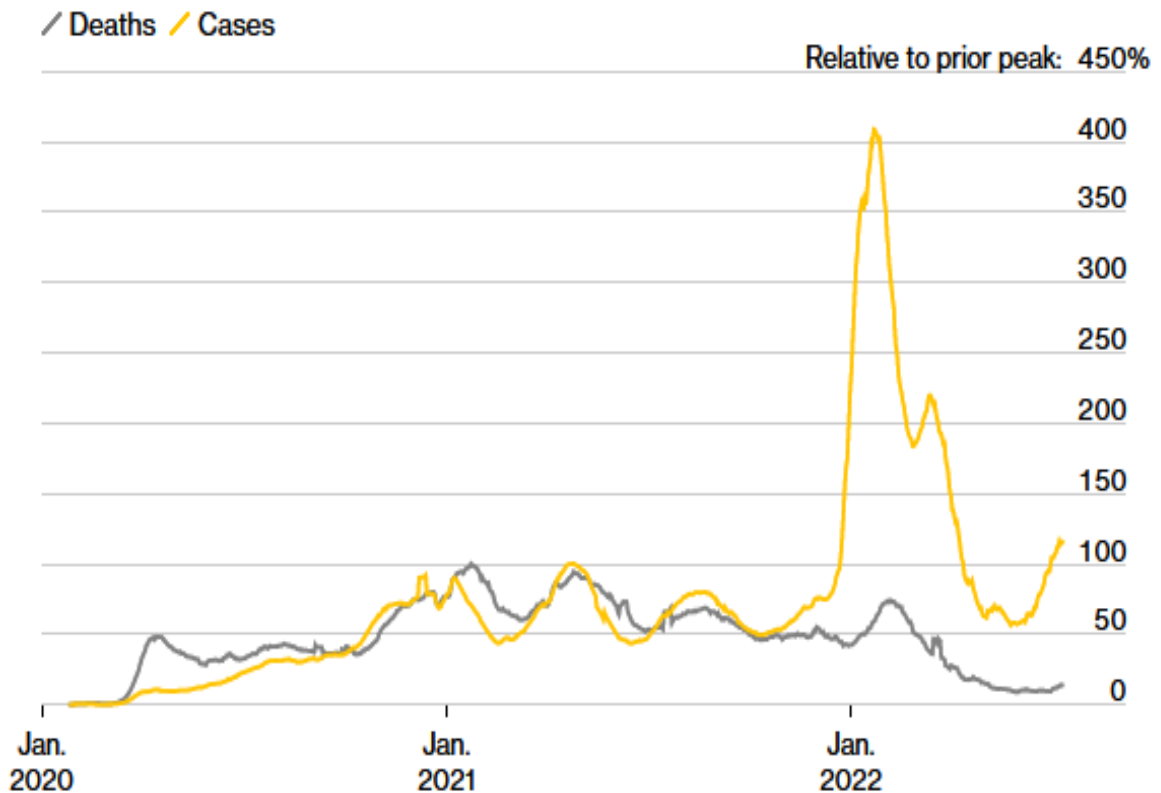


(Bloomberg, 2022)

Shown in figure 1 is a world map of Covid-19 vaccinations doses per 100 people. The visualization was built and presented by Bloomberg. Overall, the map gives the viewer a good understanding of vaccinations throughout the world. However, there is some issue with the visualization that could affect the viewer point of view. As you can see, they only use the color green as the color. This can affect the view of certain areas making them look greener or lighter in certain areas. Also, when this visualization is printed, it can look distorted because of the

color. If the author used different colors, it would make the visualization easier to read without any issue.

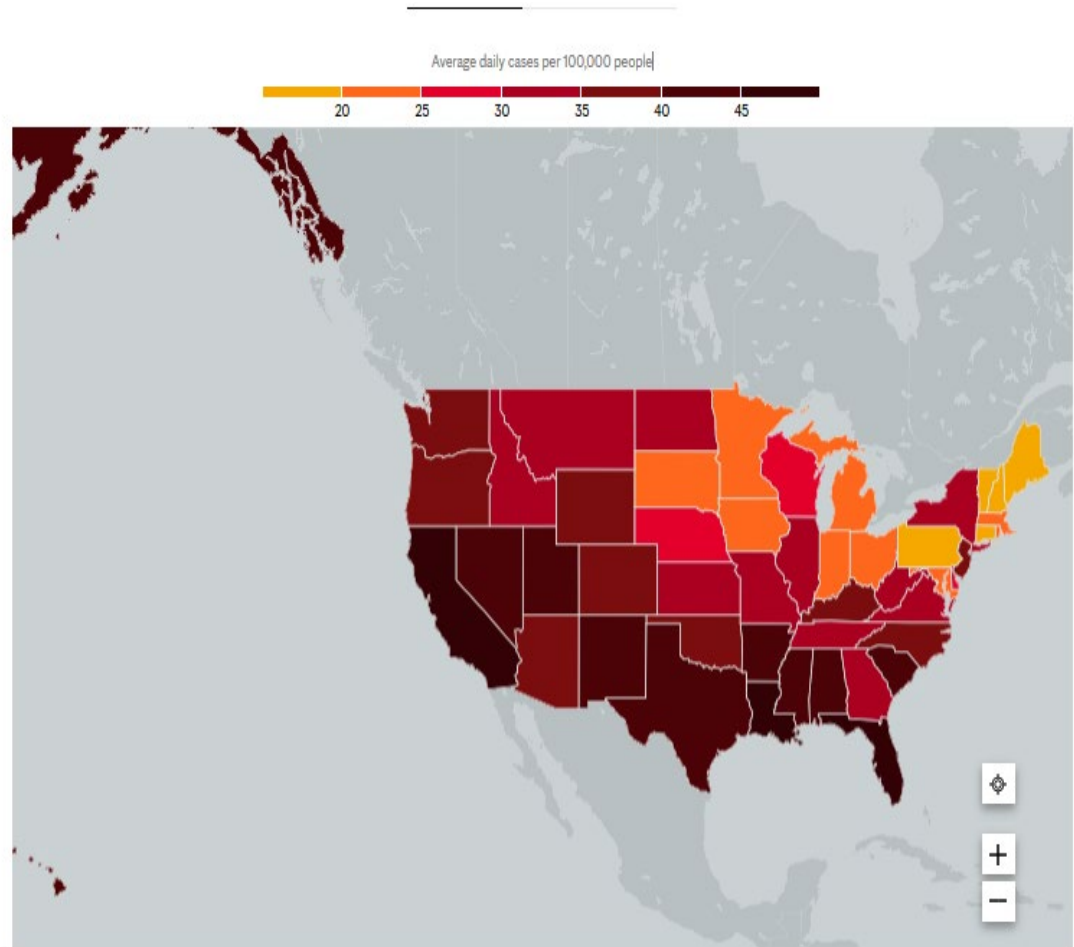
Figure 2



(Bloomberg, 2022)

Shown in figure 2 is a time series chart of deaths and cases from January 2020 through July 2022. Overall, the chart gives the viewer a good understanding of Covid-19 cases and deaths. A viewer could look at this chart and see that February 2022 had the most cases throughout the existence of Covid-19. There are some changes that could be made to make the chart look even better. Labeling the x and y axis would help the viewer understand what they are looking at. Also, adding a title to this chart would make it easier to understand as well. We also recommend changing the death line chart color from gray to blue. This combination of blue and yellow in a chart helps people with colorblind issues read the chart.

Figure 3



(Mayoclinic, 2022)

Shown in figure 3 are daily cases in the United States. Overall, this is a great visualization. The color scheme used is easy to differentiate. Any viewer can look at this chart and understand what is going on with Covid-19 in the United States. Since different areas have different populations, changing the legend to a ratio would be better to understand what areas are being more affected based on the state's population size. Therefore, a viewer who is doing research could then analyze the ratio and compare it to their policy in fighting Covid-19.

Figure 4

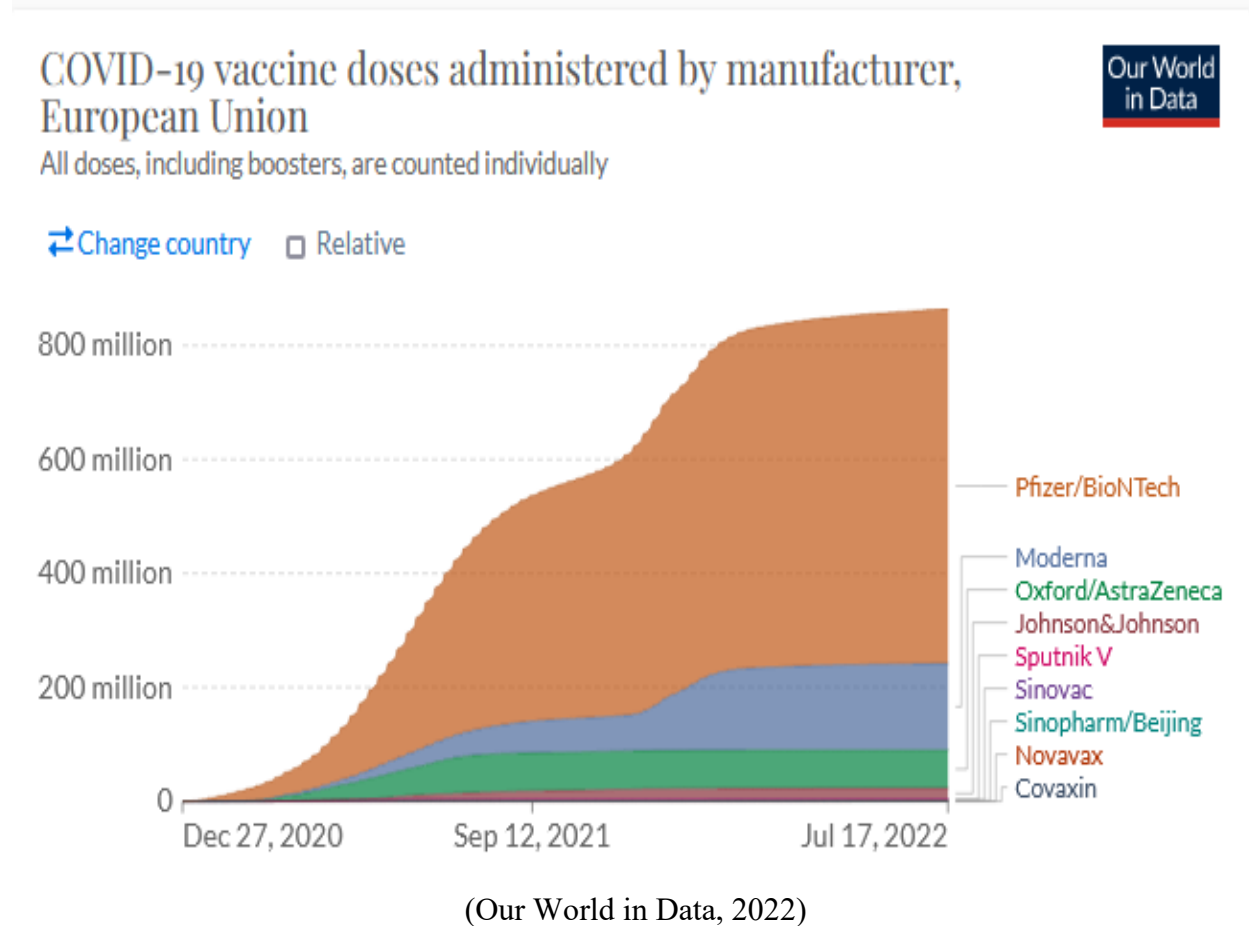
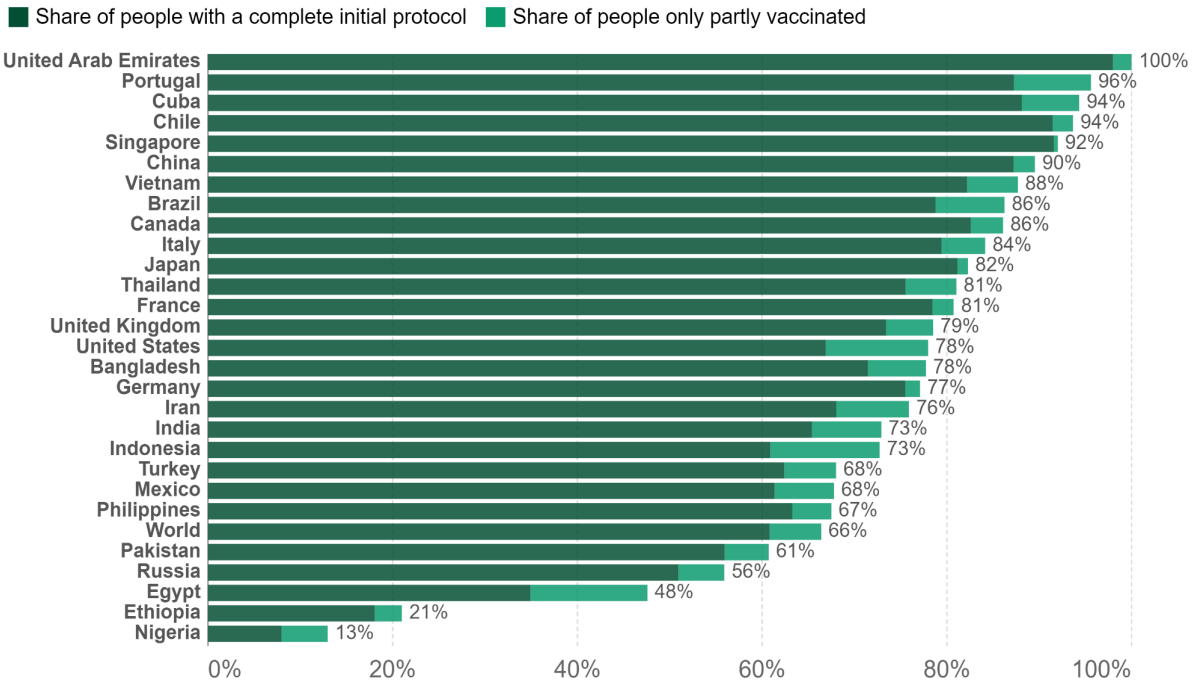


Figure 4 shows doses administered by the manufacturer. The chart itself is easy to read and very powerful. By looking at the chart, you can see what manufacturer is the most popular throughout the world. The color scheme used is really good which allows the viewer to decipher each brand. Since it is hard to see the data for the bottom five, we recommend grouping those into “other” bucket. From there, create a different chart for those manufacturers. This way, you can still see the trend of doses by each manufacturer.

Figure 5

Share of people vaccinated against COVID-19, Jun 26, 2022



Source: Official data collated by Our World in Data

CC BY

Note: Alternative definitions of a full vaccination, e.g. having been infected with SARS-CoV-2 and having 1 dose of a 2-dose protocol, are ignored to maximize comparability between countries.

(Our World in Data, 2022)

Figure 5 breaks out the percentage of people vaccinated against Covid-19 by each country. The visualizations are powerful because it truly shows what countries are doing their job by vaccinating their people. Not all countries have the same population. Therefore, looking at it in this view helps the viewer understand what countries are fully vaccinated. The only thing they should change is the colors. The color green they used is hard for the viewer to analyze. By changing it to blue and yellow, it would be better for the viewer who are color blind.

Process

Analysis of data

For our project we used different datasets from the same source, the CDC. Some of the datasets contain missing values. We mainly used Vaccinations.csv, us_state_vaccinations.csv datasets.

The Vaccinations dataset contains 16 columns and 114892 rows of data. Location column is a string datatype and contains the names of the countries across the world. Date column is of date datatype and contains the date with which the rest of the columns have values corresponding to these dates. Rest of the columns are numeric. Of them some of the columns were total_vaccinations which is the count of the total number of doses administered. Covid-19 vaccine requires multiple doses. So, in this column even if the same person received multiple doses it is added into this column. For example if a person receives two doses and a booster dose, then the total_vaccinations count is increased by 3, one per each dose. The daily_vaccinations contain the values of the new doses that were administered that day. The daily_vaccinations_per_million contains the values of the daily_vaccinations per million people in the total population of that country. The people_fully_vaccinated column contains the total number of people who got all the doses recommended by the initial vaccination protocol. The total_boosters column has the total number of booster doses administered. The daily_people_vaccinated_per_hundred contains the daily people vaccinated per hundred people in the total population of that country.

The Us_State_vaccinations dataset contains 16 columns and 35364 rows of data. Location column is a string datatype and contains the names of the states across the United States. Date column is of date datatype and contains the date with which the rest of the columns have values

corresponding to these dates. Rest of the columns are numeric. Of the columns, the `people_fully_vaccinated` column contains the total number of people who got all the doses recommended by the initial vaccination protocol. The `total_vaccinations_per_hundred` column contains the total_vaccinations per hundred people in the total_population of the state. The `total_distributed` column has the cumulative counts of Covid-19 vaccine doses recorded as shipped in CDC's Vaccine Tracking system.

The `people_vaccinated` column has the values of the total number of people who received at least one vaccine dose. For example if a person receives the first dose of the recommended two-dose vaccine, the count increases by one, and if the same person receives any other dose or booster, the count does not vary. The `daily_vaccinations` column has the count of the new doses that are administered on that particular date. The `total_vaccinations` which is the count of the total number of doses administered in the state. Covid-19 vaccine requires multiple doses. So, in this column even if the same person received multiple doses it is added into this column. For example if a person receives two doses and a booster dose, then the `total_vaccinations` count is increased by 3, one per each dose. The `total_boosters_per_hundred` has the total per hundred people in the total population of the state.

The `Vaccinations_by_manufacturer` dataset contains 4 columns and 41653 rows of data. Location column is a string datatype and contains the names of the countries across the world. Date column is of date datatype and contains the date with which the rest of the columns have values corresponding to these dates. The vaccine column has the name of the manufacturer of the Covid-19 Vaccine. The `total_vaccinations` has the cumulative count of that particular vaccine given on the date in the country.

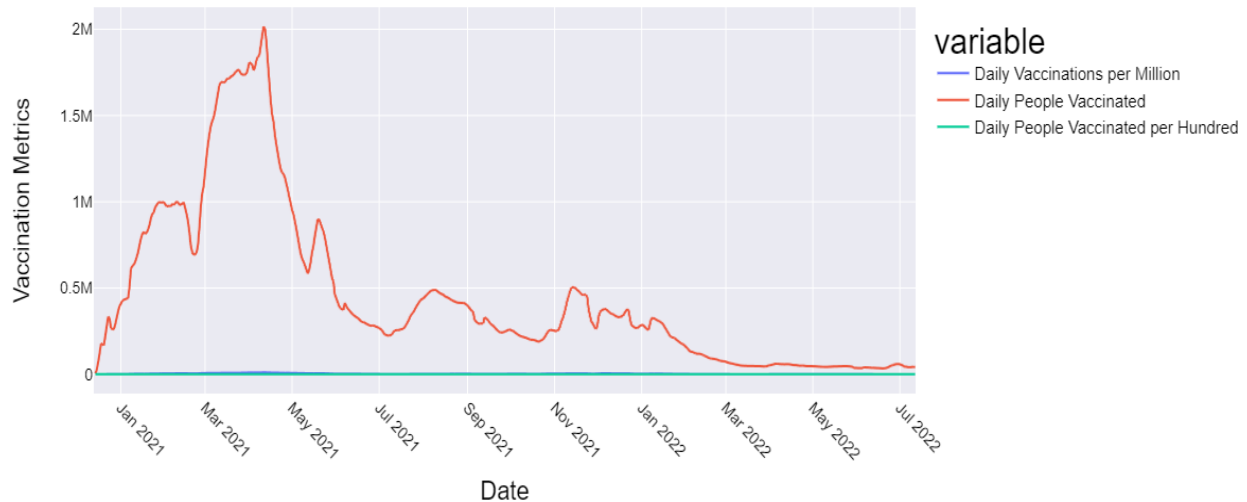
The Vaccinations_by_manufacturer dataset contains 4 columns and 41653 rows of data. Location column is a string datatype and contains the names of the countries across the world. Date column is of date datatype and contains the date with which the rest of the columns have values corresponding to these dates. The vaccine column has the name of the manufacturer of the Covid-19 Vaccine. The total_vaccinations has the cumulative count of that particular vaccine given on the date in the country.

The Vaccinations_by_age dataset contains 6 columns and 44781 rows of data. Location column is a string datatype and contains the names of the countries across the world. Date column is of date datatype and contains the date with which the rest of the columns have values corresponding to these dates. The age_group column has different age groups. The people_vaccinated_per_hundred column has the count of people per hundred in that age group who received the vaccine. The people_fully_vaccinated_per_hundred has the count of people per hundred in that age group who received all the recommended doses of the Covid-19 vaccine. The people_with_booster_per_hundred has the count of people per hundred in that age group who received all the booster doses of the Covid-19 vaccine.

While visualizing these datasets, we eliminated the missing values. Most of these missing values are due to the missing entries of data on a particular date. Some locations have no data entries on a particular date.

Failed Experiment

USA Daily Vaccinations Metrics



USA Daily Vaccination Metrics

In the above visualization we have the daily comparison metrics per million for the country USA. The metrics plotted include daily vaccination per million, daily people vaccinated and daily people vaccinated per hundred. The metrics include the non aggregated version of the actual numbers. For this plot we have used the `vaccinations.csv` dataset. The dataset is filtered on the location specific to the USA.

We can depict from the plot that the daily vaccination administered to people was at a peak during March and April of 2021. It has been a few months since the approval of covid vaccines.

Now we categorize it as a failed experiment because when we look at the data and the scales that they represent, each is at a different scale and hence we would not be able to cross compare the plots for any actual numbers.

Results, Insights, and Methods

We began the visualization by directly fetching the data from the repository mentioned in the website [Ourworldindata](https://ourworldindata.org) and the repository can be found [here](#).

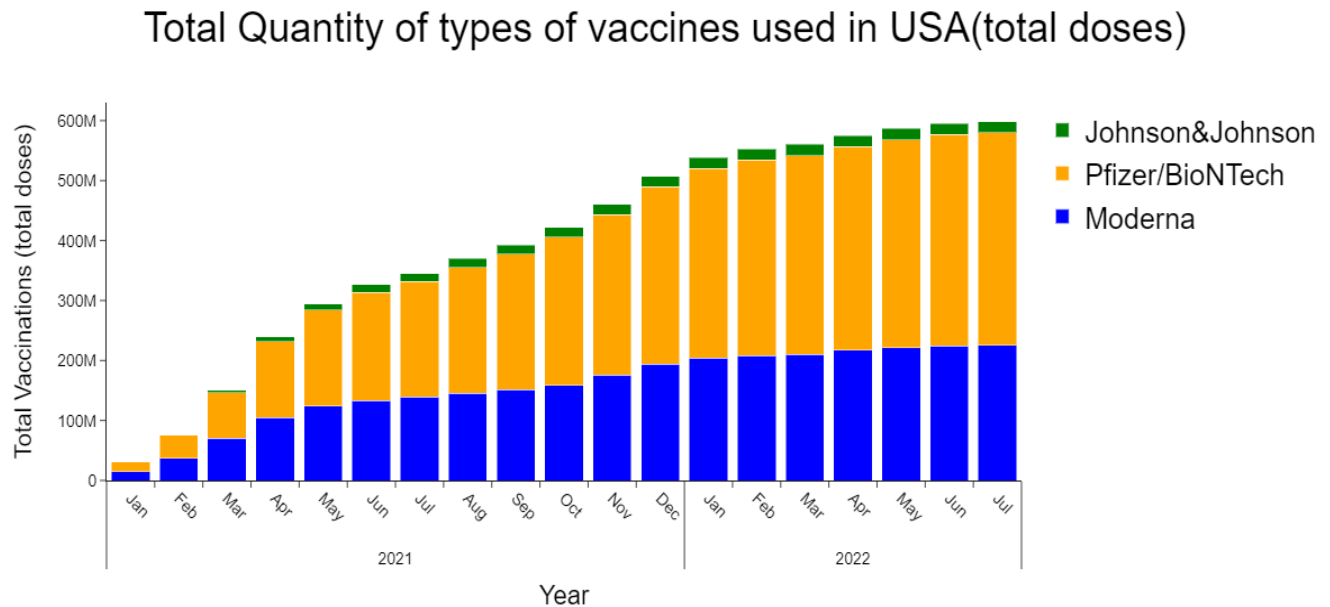


Fig 1. Total Quantity of types of vaccines used in USA(total doses)

In the above visualization we try to plot the total number of doses used in the USA for each of the vaccine types. The data is fetched from the file `vaccinations-by-manufacturer.csv`. We tried to filter the data specific to the location/country 'USA'. We have tried to plot this distribution using the plotly stacked bar graph which helps us to visualize the data in a clear and precise manner. The numbers on the y-axis here indicate the total vaccinations used for the first and second dose including the booster shots.

The date representation on the x-axis helps us to know that the covid vaccination was recommended and was administered to the people living in the USA from the beginning of

January 2021. We can see from the legend that the different vaccines used in the USA are Johnson & Johnson, Pfizer from BioNTech and Moderna. The Moderna vaccine was only distributed and was administered to the resident only from May 2021. We can clearly see that Pfizer is the most distributed vaccine which is closely followed by Moderna.

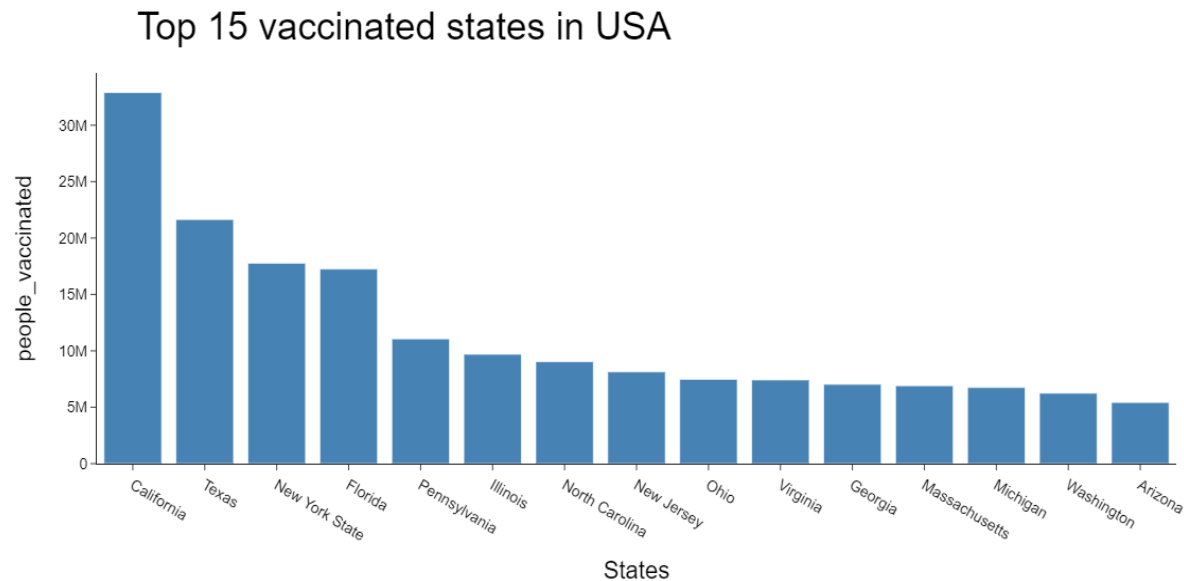


Fig 2. Top 15 vaccinated states in USA

In the above visualization we try to plot the top 15 states with the most number of people vaccinated. For this visualization we use the `us_state_vaccinations.csv`. We have fetched the last entry in the file and based on the date and the total number represented in the `people_vaccinated` columns, we have tried to plot a bar graph using plotly.

We can see from the visualization that the top 5 most vaccinated states are also the top 5 most populated states and this should be a surprise to us. For reference I have listed the [2020 census data](#) which lists the states in the USA according to the rank by population.

Rank	State	2020 Census	Percent of Total
------	-------	-------------	------------------

1	California	39,538,223	11.91%
2	Texas	29,145,505	8.74%
3	Florida	21,538,187	6.47%
4	New York	20,201,249	5.86%
5	Pennsylvania	13,002,700	3.86%

Out of the 39.5 millions of people living in California, we have the total number of people vaccinated to be 32.9 Millions. For the state of Texas we have close to 21.6 Million people vaccinated out of 29.1 Millions population count. The data can be interpreted in a similar fashion for all the other states as well.

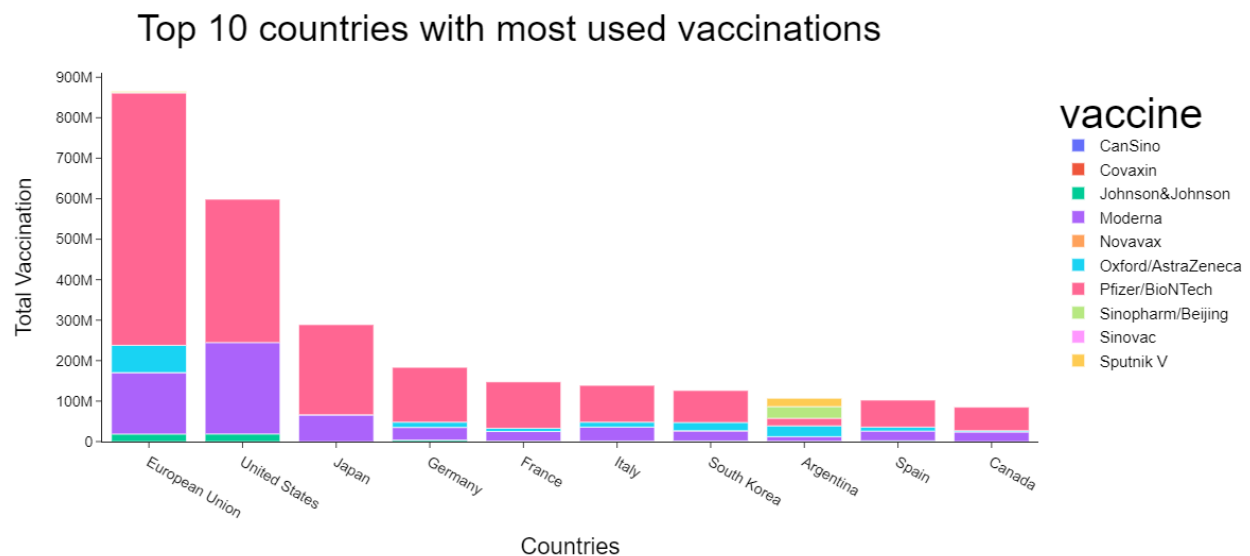


Fig 3. Top 10 countries with most used vaccinations

In the above visualization, we have plotted the top 10 countries with top distribution of different types of vaccines used. We have used vaccinations-by-manufacturer.csv dataset for this purpose. We grouped the dataset by location and vaccine and then went ahead and created a

pivot table and created an aggregated sum of the columns. We then sorted the dataset on total_vaccinations in descending order so as to be able to plot the top n countries if need be.

The visualization renders a stacked bar plot showcasing the top vaccination used in a particular location. The top 3 are the European Union, USA and Japan. For Argentina, we can see the top 5 vaccines used in that country.

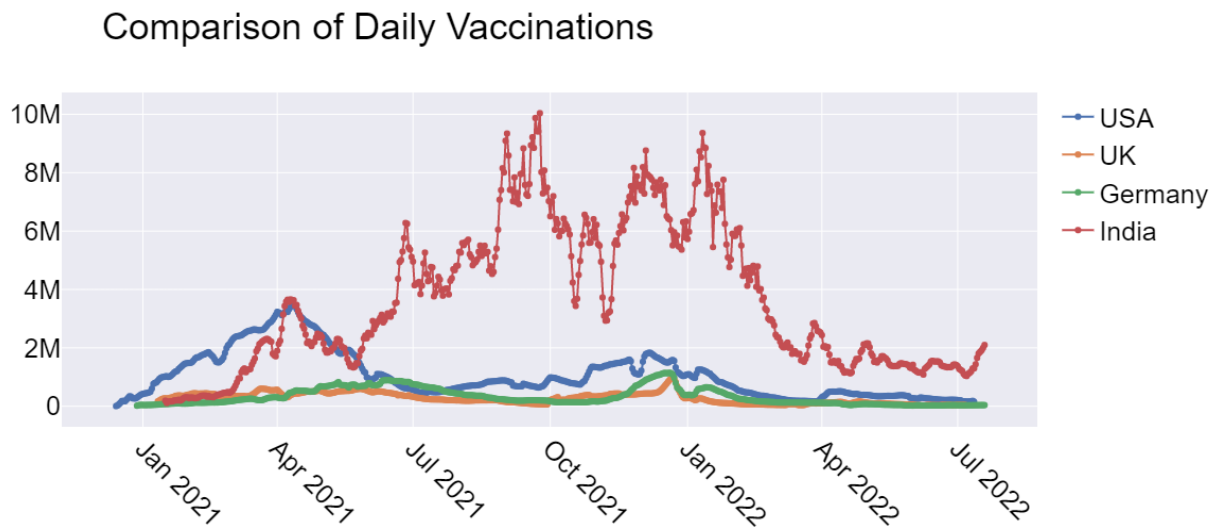


Fig 4. Comparison of Daily Vaccinations

In the above plot we have plotted the comparisons of daily vaccination for the 4 countries of our interest. We have chosen the USA, UK, Germany and India and filtered the data for each location into an individual dataframe. We have used the vaccinations.csv dataset from the repository for our visualization.

We can visually see a great variation in daily vaccination counts for India. The numbers are definitely more significant from June 2021 to March 2022 and the counts have tapered off after that. For the USA, we can see that there was a significant amount of people vaccinated

between Feb 2021 and June 2021.

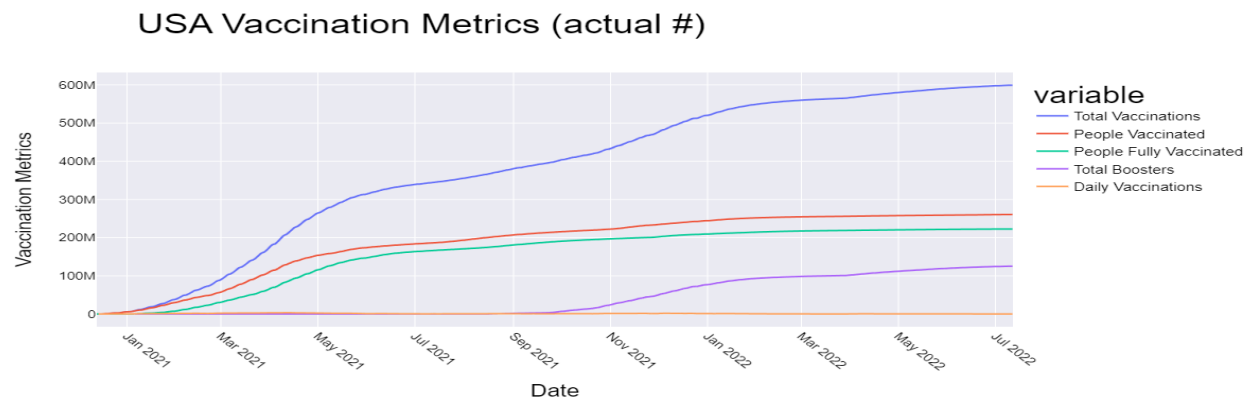


Fig 5. USA Vaccination Metrics (actual #)

In the above visualization we have the daily comparison metrics for the country USA. The metrics plotted include total vaccination, people vaccinated, people fully vaccinated, total boosters and daily comparisons. The metrics include the non aggregated version of the actual numbers. For this plot we have used the `vaccinations.csv` dataset. The dataset is filtered on the location specific to the USA.

We can visually see that the date range marks the beginning of the covid vaccination from January 2021 onwards to current date. The total vaccinations include all the covid vaccination recommended doses including the first, second shots including the booster as well.

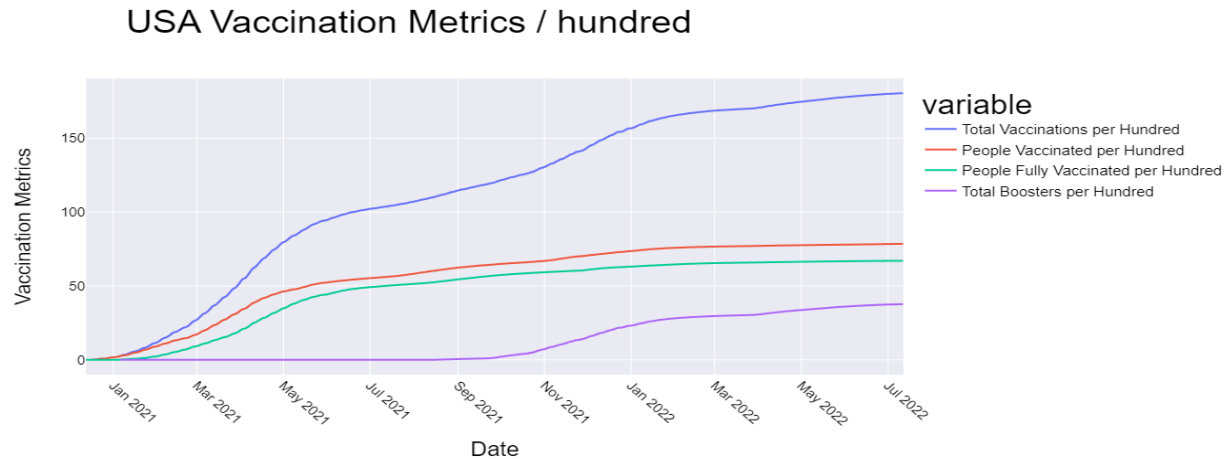


Fig 6. USA Daily Vaccination Metrics / Hundred

In the above visualization we have the daily comparison metrics per hundred people for the country USA. The metrics plotted include total vaccination per hundred , people vaccinated per hundred , people fully vaccinated per hundred and total boosters per hundred. The metrics include the aggregated version of the actual numbers. The aggregation is per hundred people living in the USA. For this plot we have used the vaccinations.csv dataset. The dataset is filtered on the location specific to the USA.

The above plot clearly depicts the difference between the people who are vaccinated and fully vaccinated, which means many people have not completed the full recommended doses for the covid vaccine. The booster shots were recommended and we can see an increase from November 2021. The booster shot still is around 37 per hundred people living in the USA. But the people vaccinated per hundred is at 78.5 so we can say that we have a long way to go to get all the residents to get boosted on the covid vaccines.

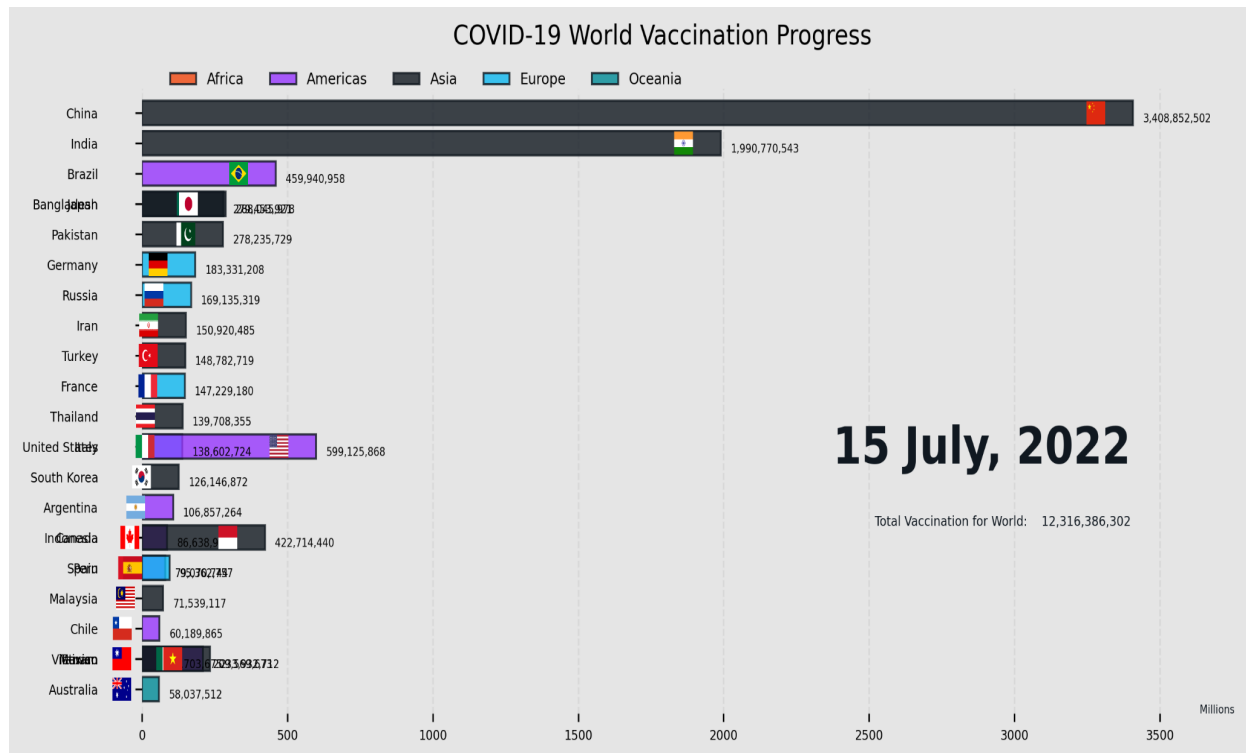


Fig 7. Covid-19 World Vaccination Progress as of July 15th, 2022

The above chart represents the Covid-19 World Vaccination progress which we have visually depicted as a race between the nations to showcase the total number of vaccines administered to people per country. This is the status as of 15th of July , 2022. This is the time stamp almost at the end of the video that we have captured to visually represent the race. Since China is the most populated country we can say that China is the most vaccinated country followed by India.

The chart has a tick on the x-axis representing every 500 millions of vaccines, and the different countries are depicted on the y-axis. We have color coded the counties into 5 different continents and color coded them according to the continent category they fall into. The different continent categories are Africas, Americas, Asia, Europe and Ocenia. The video for this race spans to 1 minute 57 seconds beginning from 10 December 2010 to 19 July 2022 as of the day of submission of this project report.

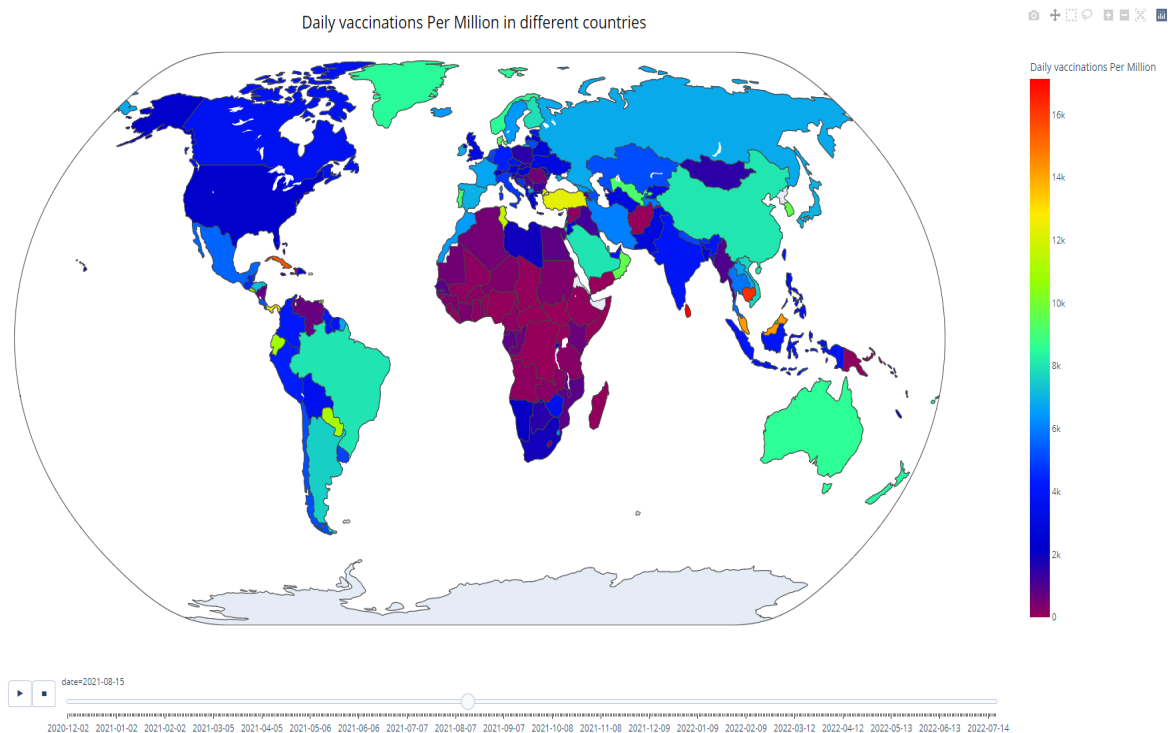


Fig 8. Map Plot of Daily Vaccinations Per million in different countries.

For the following four visualizations we used data from Vaccinations.csv dataset. The above map visualization shows Daily Vaccinations per million in different countries. The data entry for the column `daily_vaccinations_per_million` is the count of the Covid-19 vaccinations that are being administered on a daily basis per million of the total population of countries across the world. The Covid-19 vaccine is a multi-dose recommended vaccine, so every time a dose is administered the count of the `daily_vaccinations` increases in turn affecting the values in this column. In this visualization we can see that there is an animation frame at the bottom. This animation frame is the date which has the corresponding values of `daily_vaccinations_per_million`.

The method used above is the `choropleth` method of `plotly.express`. This method allows us to add an animation frame. For the above visualization the animation frame we chose is `date`. The

locations on the map are identified by the `iso_codes` of the countries. The purpose of these codes is to define internationally recognised codes of letters that can be used to refer to the countries and their subdivisions. It is better to use these codes as there is no need to go through the whole string of the name instead we can just parse these shortened codes to identify the countries. The color parameter helps us to give the column values we want to map on the visualization. This color is also used to scale. As the animation frame moves the scale also differs according to the values varying in the column. The projection parameter helps us to determine what type of visualization we want to see. The projection we used in the above visualization is natural earth projection. We can also use different types like mercator, albers usa when we are visualizing the United States data.

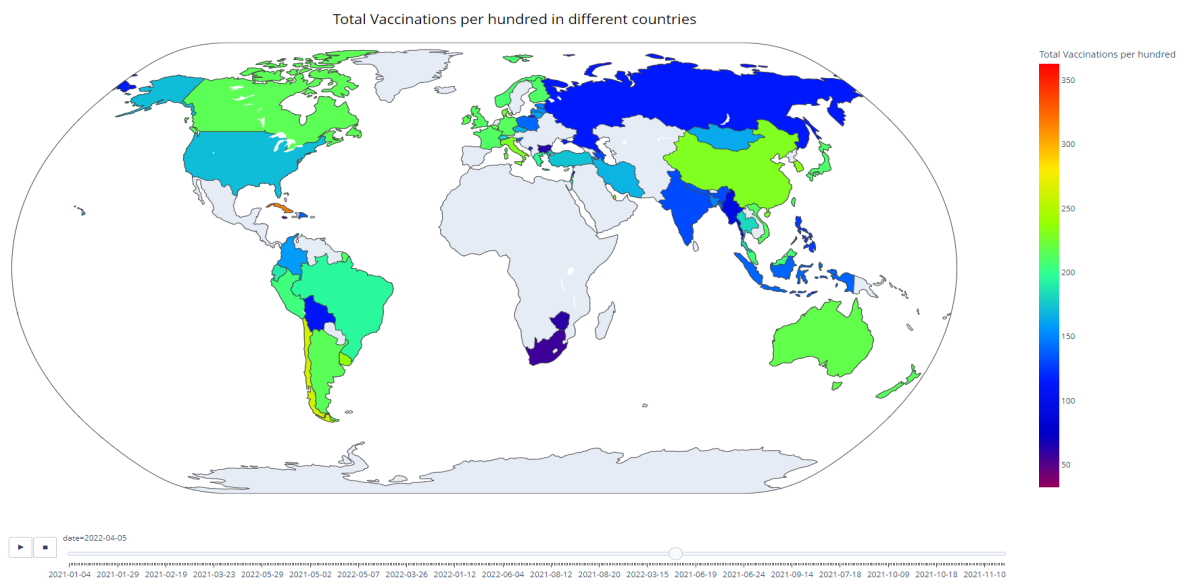


Fig 9. Map Plot of Total Vaccinations Per Hundred in different countries.

The `total_vaccinations_per_hundred` column has the values of each dose of vaccine that is administered till that date per hundred of the population. As mentioned earlier the Covid-19 vaccine is a multi-dose. So, everytime a person gets a dose the count in the `total_vaccinations`

column is increased, thus in turn affecting the values in this column. The same animation frame is used as in the above visualization.

The method used here is the choropleth method of `plotly.express`. The animation frame and the color scale are `date` and `total_vaccinations_per_hundred` respectively. The scale in this changes according to the values in the `total_vaccinations_per_hundred` column. The `iso_codes` are used to identify the countries in the map as these codes are easier to identify the countries. The projection used is the same as above visualization, natural earth.

As we look at the visualization, we can see that there are countries that have no color. This is because in the data, the information on a particular date has entries only for some countries. When we try to clean the data that has no entries of all the locations, we are left with very little data which doesn't represent the whole dataset. So, instead we handled the missing data and grouped the values according to the location and the `total_vaccinations_per_hundred` values. This is because the values in the `total_vaccinations_per_hundred` is the cumulative values of the column for the location at its last occurrence on a particular date. Based on the above assumptions, we visualized the above data.

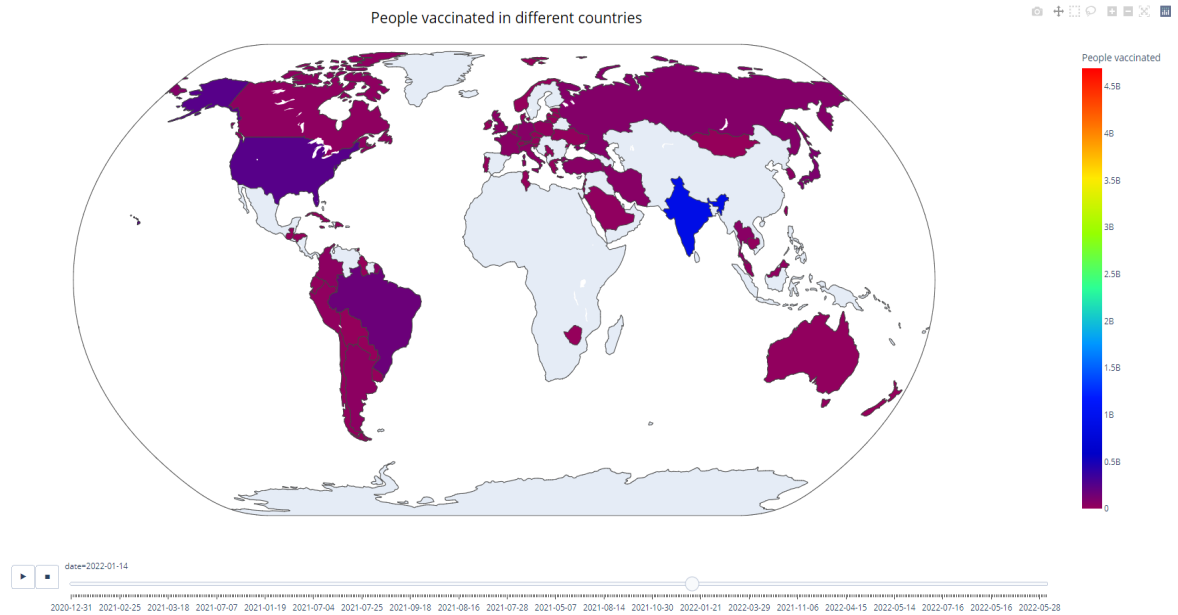


Fig 10. Map Plot of People Vaccinated in different countries.

The above visualization is the natural earth projection map visualization of People Vaccinated in different countries. The `people_vaccinated` column data is used for visualization and is the data of the people that are administered with the Covid-19 vaccine. This column acts differently compared to the above two columns of data. In this, when a person is vaccinated for the first time with a dose of Covid-19 vaccine the count is increased. If the same person gets the next dose or a booster dose the count does not vary. This column is to know how many people are vaccinated even if it's the basic dose. The animation frame here is the date. The play button on the frame enables us to see the flow of the data into different dates. And there is a stop button, to stop and inspect the data more closely. We can also drag the pointer on the frame to see data on a particular date.

The method used here is the `choropleth` method of `plotly.express`. The animation frame and the color scale are `date` and `people_vaccinated` respectively. The scale in this changes according to the values in the `people_vaccinated` column. The `iso_codes` are used to identify the countries in

the map as these codes are easier to identify the countries. The `color_continuous_scale` allows us to use different colormaps to see the visualizations. We used the rainbow in our visualizations. We are also using the `write_html` method which is used to present the whole visualization in a new html page which can help in seeing the data clearly.

As we look at the visualization, we can see that there are countries that have no color. This is because in the data, the information on a particular date has entries only for some countries. We handled the missing data and then visualized the data.

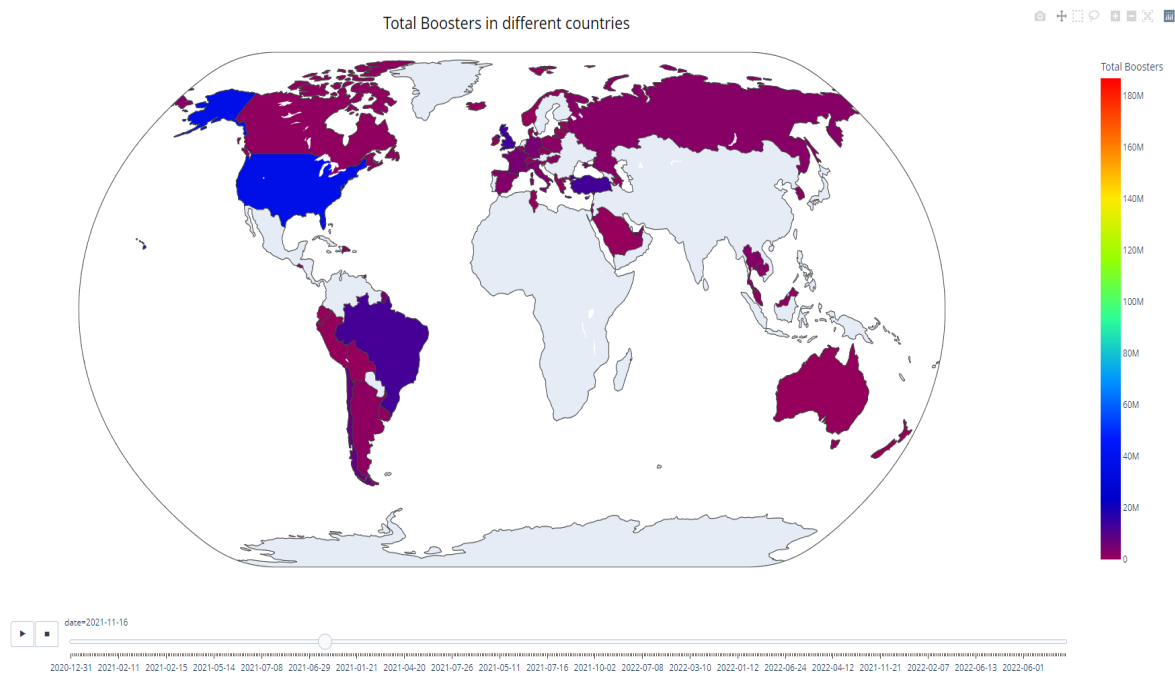


Fig 11. Map Plot of Total Boosters in different countries.

In the above visualization, we can see a world map with the data of Total Boosters in different countries. The `total_boosters` column has the total number of COVID-19 vaccination booster doses administered, excluding the information about the initial prescribed doses. The initial vaccination protocol prescribed only two doses of Covid-19 vaccine. The booster doses are

the vaccination doses administered beyond the number prescribed by the initial vaccination protocol. All the visualizations above are interactive, we can zoom into a particular state of the country and inspect the data closely.

The method used here is the `choropleth` method of `plotly.express`. The animation frame and the color scale are `date` and `total_boosters` respectively. The scale in this changes according to the values in the `total_boosters` column. The `color_continuous_scale` allows us to use different colormaps to see the visualizations. This method also provides a parameter, `hover_name`. This parameter helps us interact with the map. We can pass a single value or an array of values that are the column names. When we hover the mouse on a particular country with the values that are being represented on the map or the values we pass in the array.

As we look at the visualization, we can see that there are countries that have no color. This is because in the data, the information on a particular date has entries only for some countries. We handled the missing data and then visualized the data.

Vaccinations per hundred of people with Age greater than 70

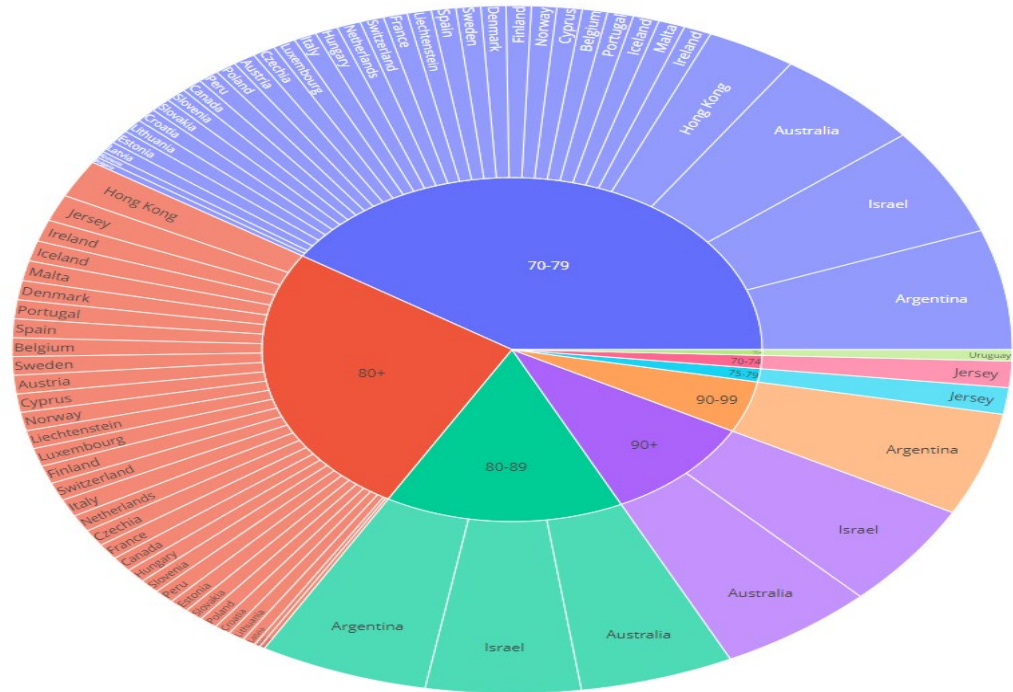


Fig 12. Sunburst plot of Vaccinations per hundred of age group greater than 70 in different countries.

For the following three visualizations, we used the data from the `Vaccinations_by_age.csv` dataset. We filtered the data to represent the data in three different parts. This is because when the data is used as a whole, as the data is heavy the chart looks more crumpled and is hard to read or get a picture of what is being represented.

The following visualization is called the sunburst. This visualization is similar to pie charts. This can be seen in different layers. This chart has the center layer as the age. We can see there are age groups that are greater than 70. As we can see that the age group 70-79 have the most vaccination per hundred.

This visualization uses the `plotly.express` sunburst method. This representation is for the Vaccination per hundred people with Age greater than 70. On top of this age group we can see the

countries. The categories here have areas allocated based on the people Vaccination per hundred people scale. This method is given with the values of `people_vaccinated_per_hundred`. This helps in dividing the area of the map based on these values.

Vaccinations per hundred of people with Age greater than 50 and less than 70

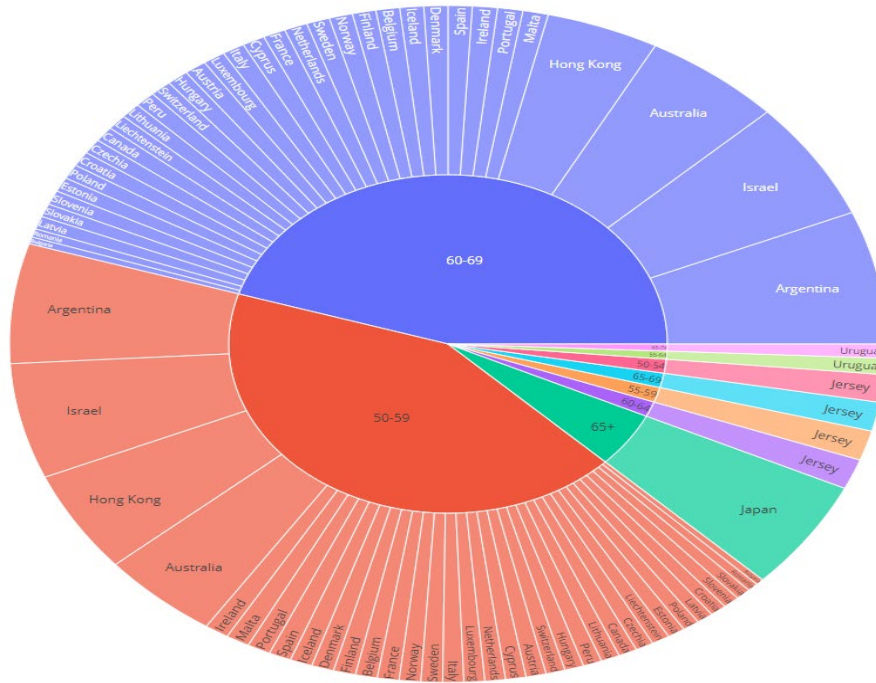


Fig 13. Sunburst plot of Vaccinations per hundred of age group between 50 and 70 in different countries.

We can see there are age groups that are between 50 and 70. This chart also has the center layer as the age. The visualization shows the layers of data age group and locations. The data represented the order of people who got vaccinated per hundred in a country of a particular age group.

The above visualization uses the sunburst method of `plotly.express`. This method has a parameter called `path`. This path parameter is to determine the flow of the layers in the chart. An array with the names of the columns that we want to see on the map can be passed. The very first in the array is the center part of the chart followed by the others according to the array entries.

Vaccinations per hundred of people with Age less than 50

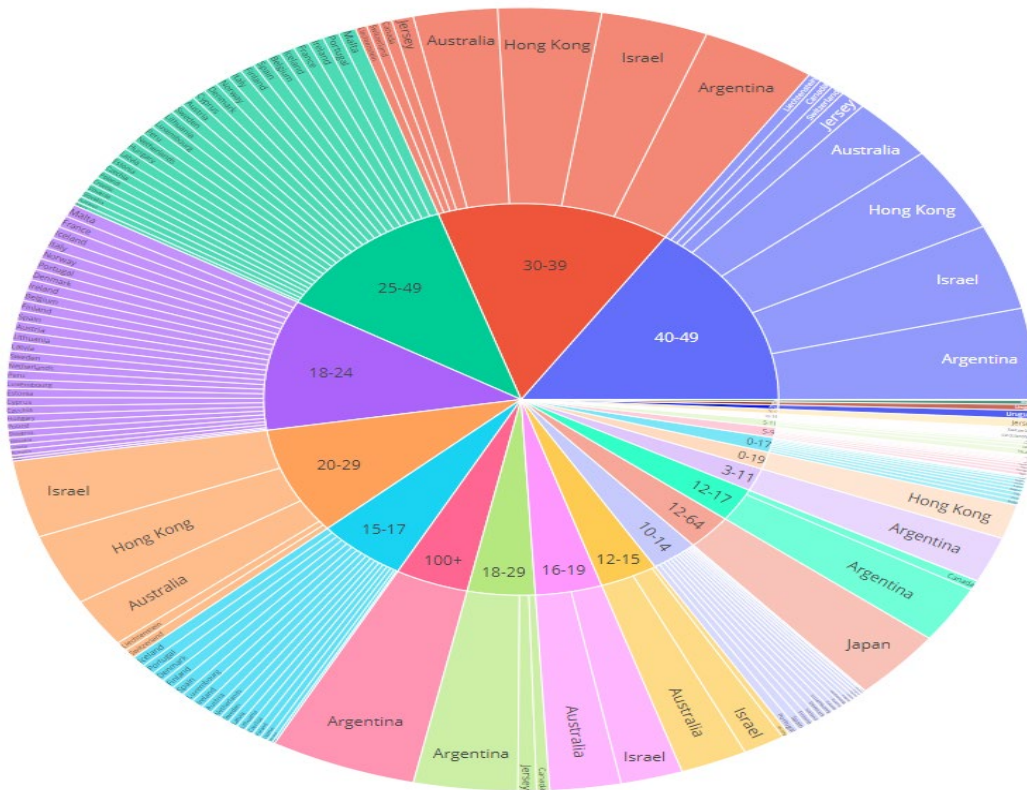


Fig 14. Sunburst plot of Vaccinations per hundred of age groups less than 50 in different countries.

This is also a sunburst chart representing the Vaccinations per hundred in different countries of age groups that are less than 50. Same as the other two charts this chart also has the age group in the center.

This visualization uses the same method as the rest two, sunburst of plotly.express. When we hover the mouse on the chart we can see a legend that gives the values that are being represented on the map.

We are using the write_html method to save the plot and represent the charts in a separate html file.

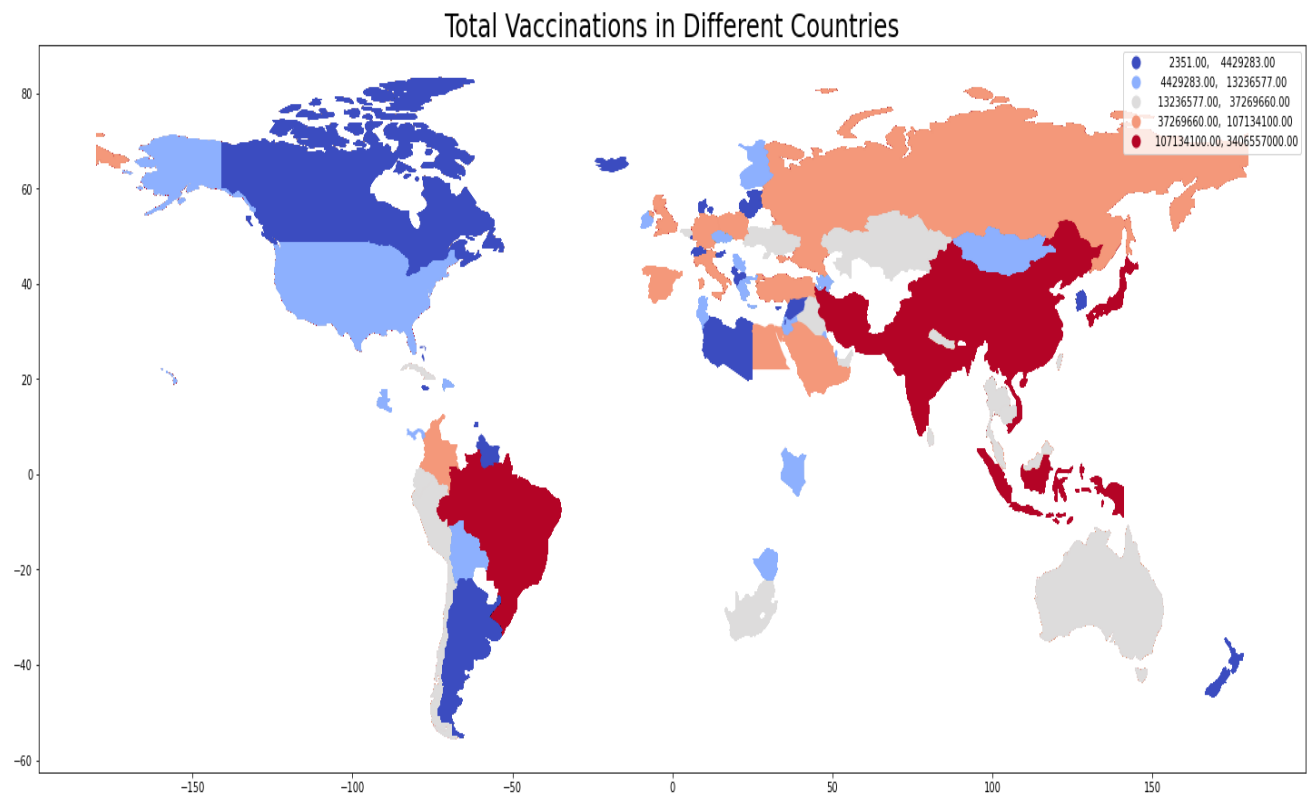


Fig 14. Map plot of Total Vaccinations in Different Countries.

The above plot is the Total Vaccinations in different countries till date. For this visualization we used the Vaccinations.csv data. The vaccinations that are administered for the latest date are being plotted on this map. The total vaccinations are the count of the vaccinations that are being administered incrementing the value everytime a dose is administered irrespective of it being a recommended dose or a booster dose.

This chart is drawn using the geopandas package. This package has a dataset with the world data with its geometry. The dataset of vaccinations and world are merged on iso_codes.

Then we are sorting the dataset on the total_vaccinations column in descending order. This helps in getting the cumulative value of the latest date. Then this merged dataset is plotted representing the total vaccinations in different countries. The legend can be seen on the top right of the chart.

Discussion, Future Work, and Conclusion

Covid-19 has been affecting billions of people around the world since 2020. As of July 2022, over half a billion people were infected and tens of millions died because of Covid-19. Our objective for this project is to provide world analysis on Covid-19. For our project, we analyzed data from a world view, to a state view, and to a manufacturer view. Based on our visualization, we can see that Pfizer is the most used vaccine in the United States and throughout the world.

For our future work, we plan on finding data related to mental health, obesity, and divorce rate to perform correlation analysis to see if Covid-19 had any static increases on those measures. Since many people had to quarantine and isolate themselves from other people, we believe that it affected many people's mental health.

In conclusion, with many experiments we performed regarding visualizations, we believe our analysis we performed on Covid-19 will help the viewers understand more about what happened during the last couple of years. Most of our visualizations outlined the effects of Covid-19 from 2020 through 2022. With hundreds of millions people affected, analysis on this pandemic will help stop the next one.

,

References

Bloomberg. (n.d.). Bloomberg.com. Retrieved July 20, 2022, from

<https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/>

Centers for Disease Control and Prevention. (2022, June 16). *What you should know about covid-19 and pets*. Centers for Disease Control and Prevention. Retrieved July 20, 2022, from <https://www.cdc.gov/healthypets/covid-19/pets.html>

Centers for Disease Control and Prevention. (n.d.). *Basics of covid-19*. Centers for Disease Control and Prevention. Retrieved July 20, 2022, from <https://www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19/basics-covid-19.html>

Covid-19 United States cases by county. Johns Hopkins Coronavirus Resource Center. (n.d.). Retrieved July 20, 2022, from <https://coronavirus.jhu.edu/us-map>

Covid-19 vaccine doses administered by manufacturer. Our World in Data. (n.d.). Retrieved July 20, 2022, from <https://ourworldindata.org/grapher/covid-vaccine-doses-by-manufacturer>

Infoplease. (n.d.). *State population by Rank*. Infoplease. Retrieved July 21, 2022, from <https://www.infoplease.com/us/states/state-population-by-rank>

Mayo Foundation for Medical Education and Research. (n.d.). *U.S. coronavirus map: Tracking the trends*. Mayo Clinic. Retrieved July 20, 2022, from <https://www.mayoclinic.org/coronavirus-covid-19/map>

Nirmita Panchal, R. K., & 2021, F. (2022, April 28). *The implications of COVID-19 for mental health and substance use*. KFF. Retrieved July 20, 2022, from <https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/>

