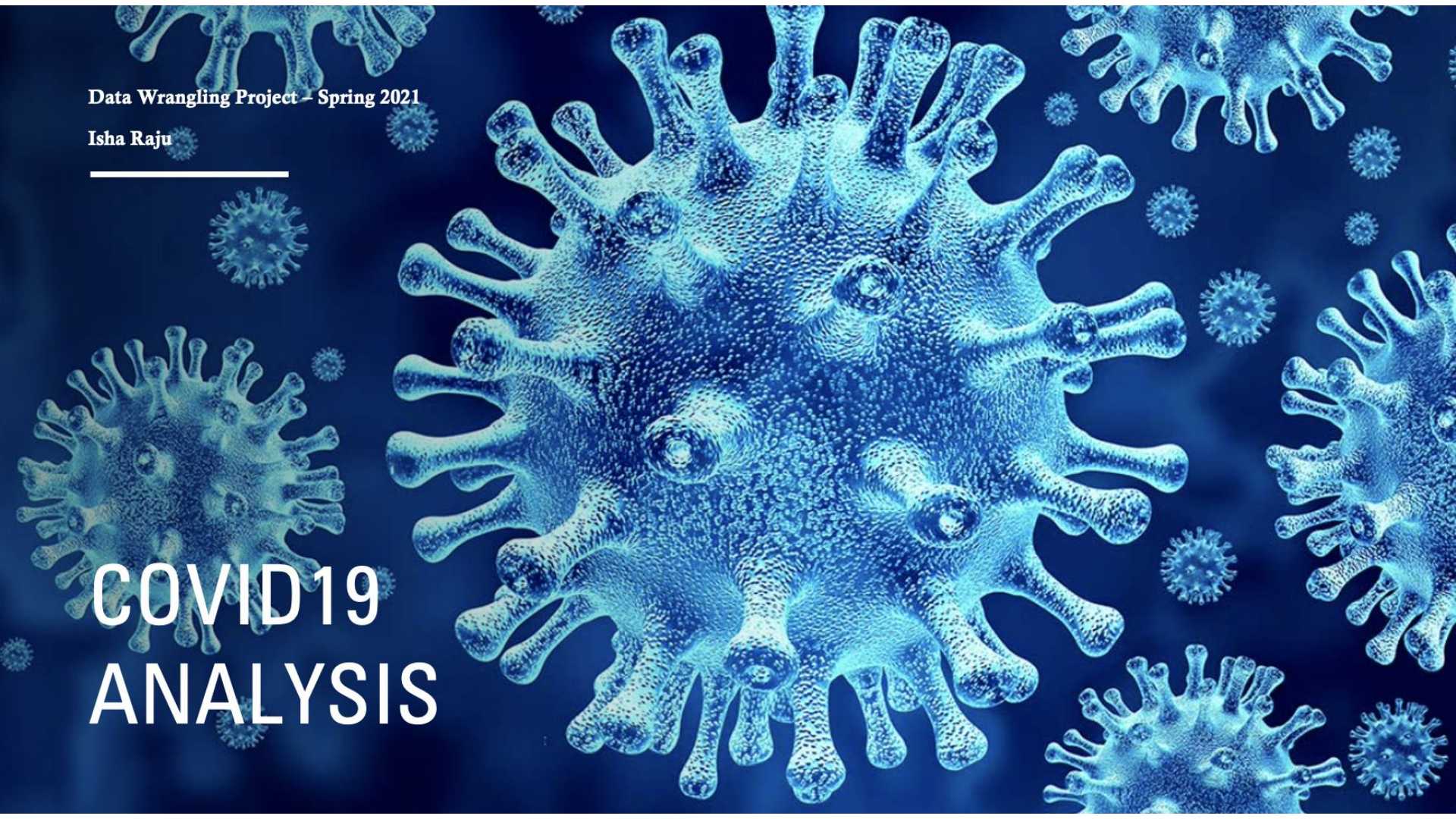
Data Wrangling Project – Spring 2021

Isha Raju



COVID19 ANALYSIS

Overview

The following file has two parts:

1. Data scraping and wrangling of COVID data from web pages As discussed during the presentation the web scraping took up a lot of time to get few data columns. To get more data for better analysis I would have required to scrape a lot more web pages. As I wanted to apply the visualizations learnt in class I decided to go ahead with readily available data.

2. Data wrangling and visualization of readily available COVID vaccine data from Kaggle

Data scraping and Wrangling

```
# A tibble: 2,460 x 7
# Groups:   Region, Year, Month_num, Date [2,340]
   Continent Region  Date                count  Year Month_num   Day
   <chr>     <chr>   <dttm>              <dbl> <dbl>     <int> <dbl>
 1 africa    algeria 2020-01-01 00:00:00     0  2020         1     1
 2 africa    algeria 2020-02-01 00:00:00     0  2020         2     1
 3 africa    algeria 2020-03-01 00:00:00     0  2020         3     1
 4 africa    algeria 2020-04-01 00:00:00    35  2020         4     1
 5 africa    algeria 2020-05-01 00:00:00   450  2020         5     1
 6 africa    algeria 2020-06-01 00:00:00   653  2020         6     1
 7 africa    algeria 2020-07-01 00:00:00   912  2020         7     1
 8 africa    algeria 2020-08-01 00:00:00  1210  2020         8     1
 9 africa    algeria 2020-09-01 00:00:00  1510  2020         9     1
10 africa    algeria 2020-10-01 00:00:00  1736  2020        10     1
# … with 2,450 more rows
```

Vaccine Data

	country <fctr>	iso_code <fctr>	date <fctr>	total_vaccinations <dbl>	people_vaccinated <dbl>	people_fully_vaccinated <dbl>
12965	Zimbabwe	ZWE	2021-04-14	262128	231632	30496
12966	Zimbabwe	ZWE	2021-04-15	285881	252932	32949
12967	Zimbabwe	ZWE	2021-04-16	304701	269732	34969
12968	Zimbabwe	ZWE	2021-04-17	311901	276095	35806
12969	Zimbabwe	ZWE	2021-04-18	314446	278583	35863
12970	Zimbabwe	ZWE	2021-04-19	316991	280568	36423

daily_vaccinations_raw <dbl>	daily_vaccinations <dbl>	total_vaccinations_per_hundred <dbl>	people_vaccinated_per_hundred <dbl>
13773	11816	1.76	1.56
23753	13172	1.92	1.70
18820	14071	2.05	1.81
7200	12738	2.10	1.86
2545	12993	2.12	1.87
2545	11773	2.13	1.89

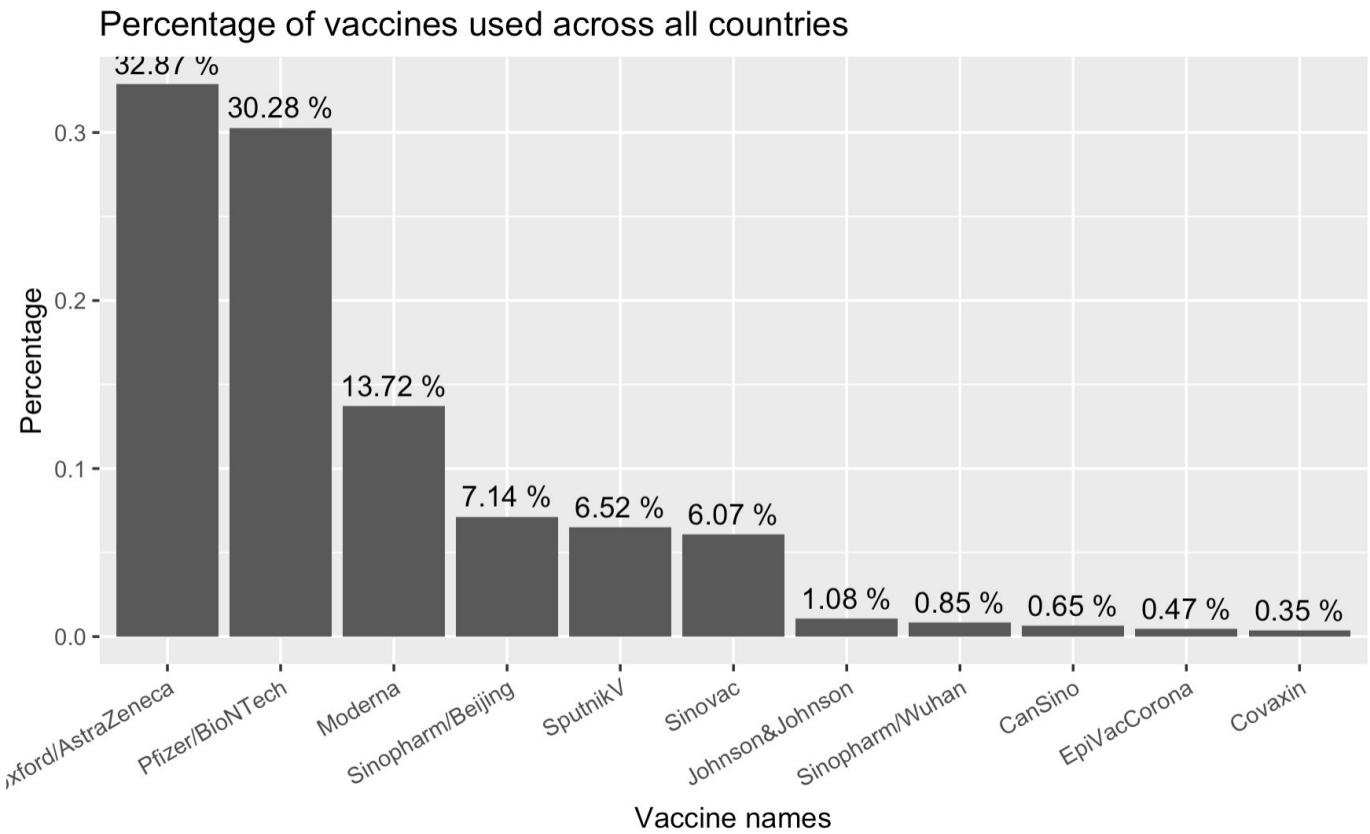
people_fully_vaccinated_per_hundred <dbl>	daily_vaccinations_per_million <dbl>	vaccines <fctr>
0.21	795	Sinopharm/Beijing
0.22	886	Sinopharm/Beijing
0.24	947	Sinopharm/Beijing
0.24	857	Sinopharm/Beijing
0.24	874	Sinopharm/Beijing
0.25	792	Sinopharm/Beijing

Content

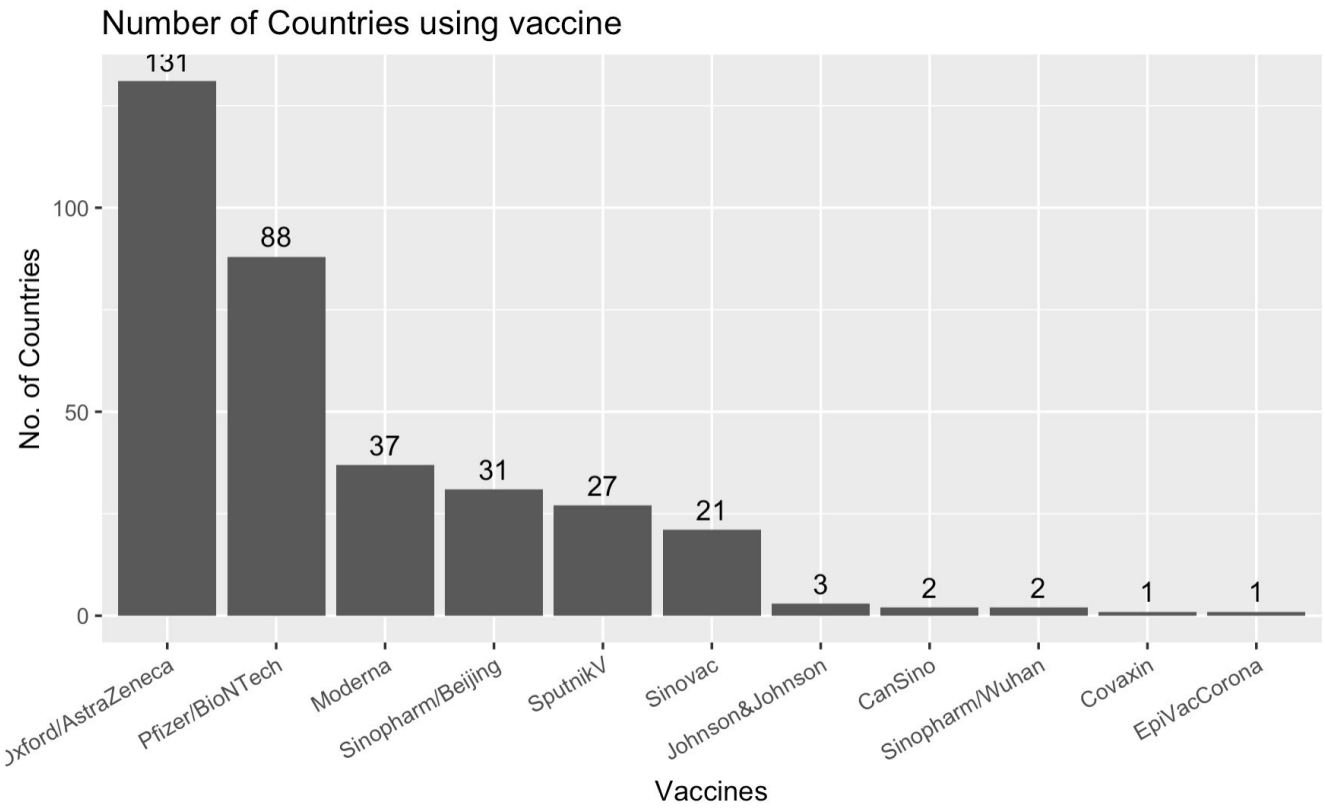
The data (country vaccinations) contains the following information:

- **Country**- this is the country for which the vaccination information is provided;
- **Country ISO Code** - ISO code for the country;
- **Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
- **Total number of vaccinations** - this is the absolute number of total immunizations in the country;
- **Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people;
- **Total number of people fully vaccinated** - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme;
- **Daily vaccinations (raw)** - for a certain data entry, the number of vaccination for that date/country;
- **Daily vaccinations** - for a certain data entry, the number of vaccination for that date/country;
- **Total vaccinations per hundred** - ratio (in percent) between vaccination number and total population up to the date in the country;
- **Total number of people vaccinated per hundred** - ratio (in percent) between population immunized and total population up to the date in the country;
- **Total number of people fully vaccinated per hundred** - ratio (in percent) between population fully immunized and total population up to the date in the country;
- **Number of vaccinations per day** - number of daily vaccination for that day and country;
- **Daily vaccinations per million** - ratio (in ppm) between vaccination number and total population for the current date in the country;
- **Vaccines used in the country** - total number of vaccines used in the country (up to date);
- **Source name** - source of the information (national authority, international organization, local organization etc.);
- **Source website** - website of the source of information;

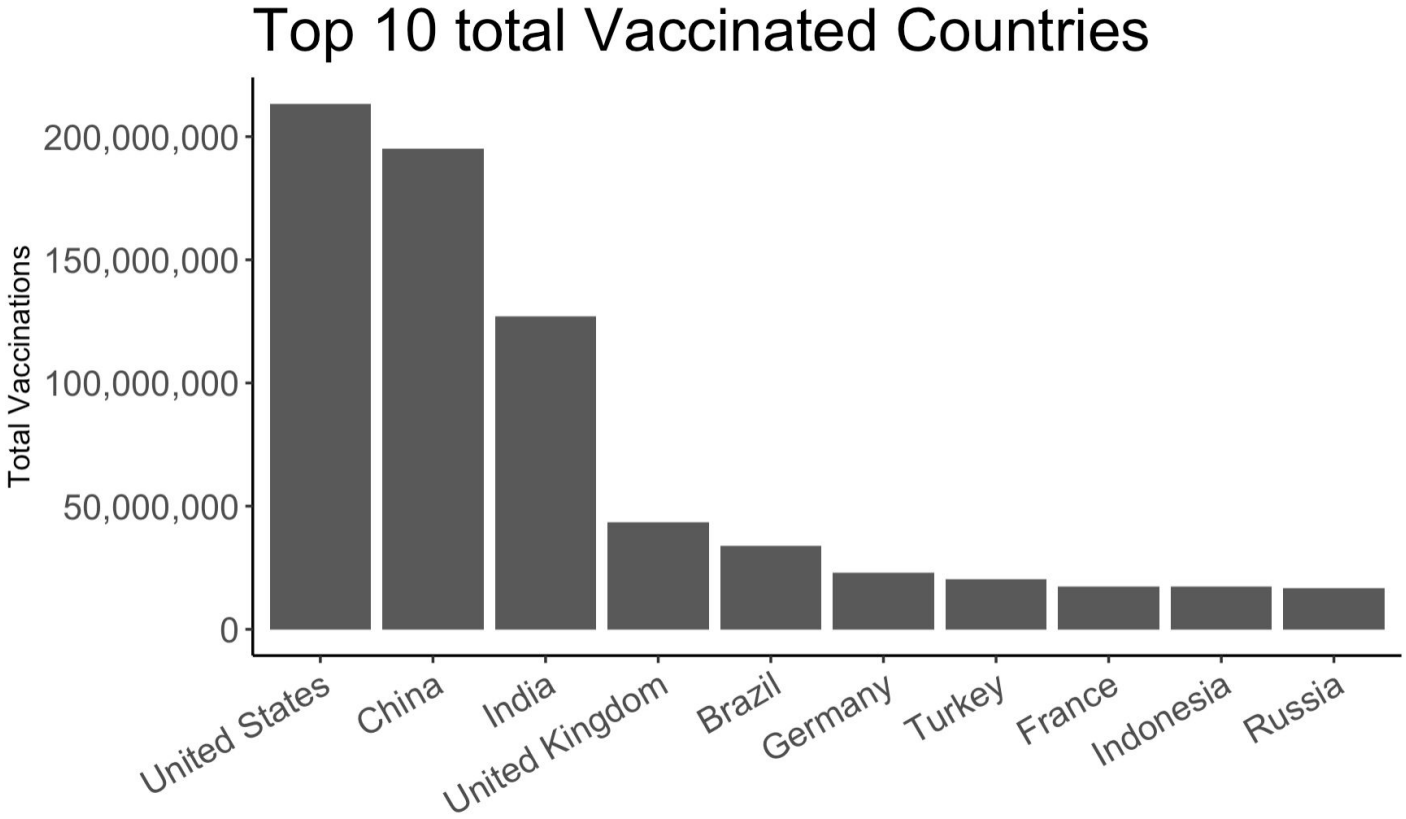
Of all the vaccines administered in the world, Astrazenica seems to be the most administered vaccine(approx 33%) followed by Pfizer(approx 30%) and Moderna(approx 14%)



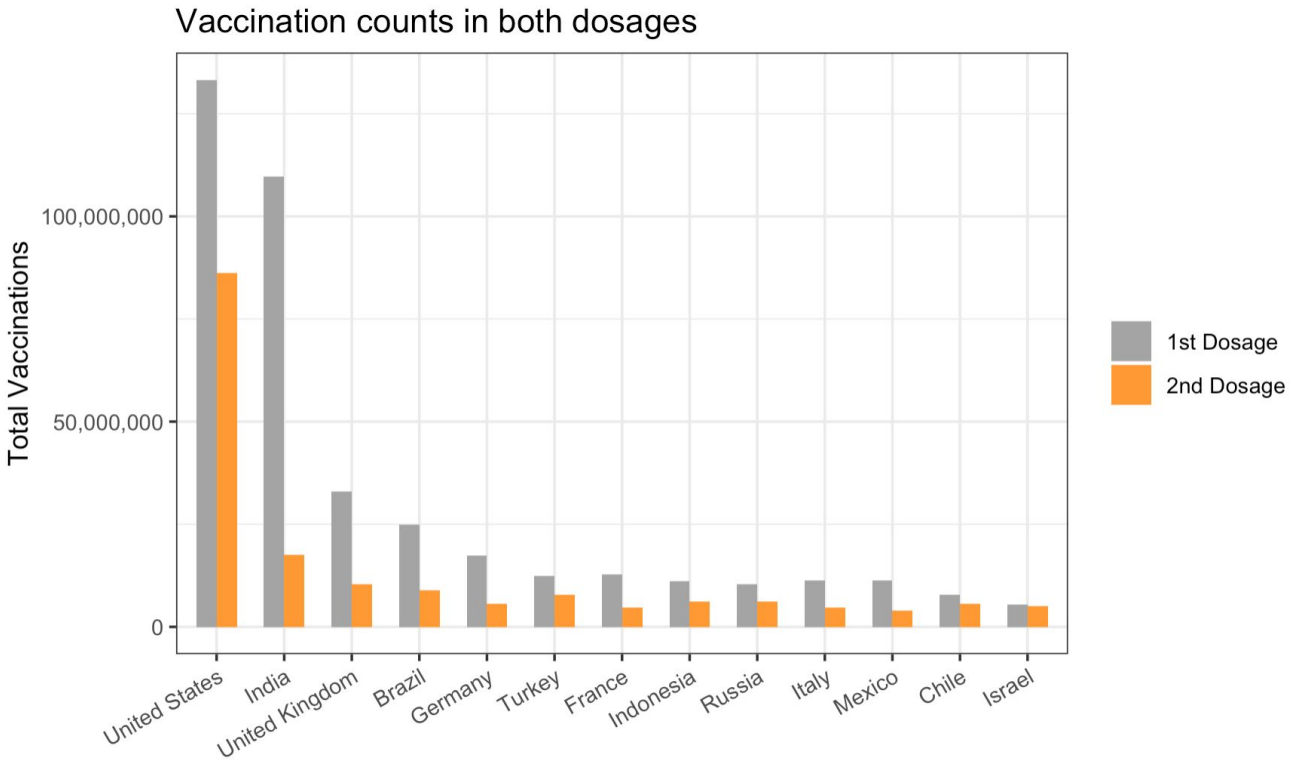
Of all the vaccines Astrazeneca is being used in the most of the countries 131 to be precise, followed by Pfizer which is administered in 88 countries and then with a huge drop with moderna which is administered in 37 countries.



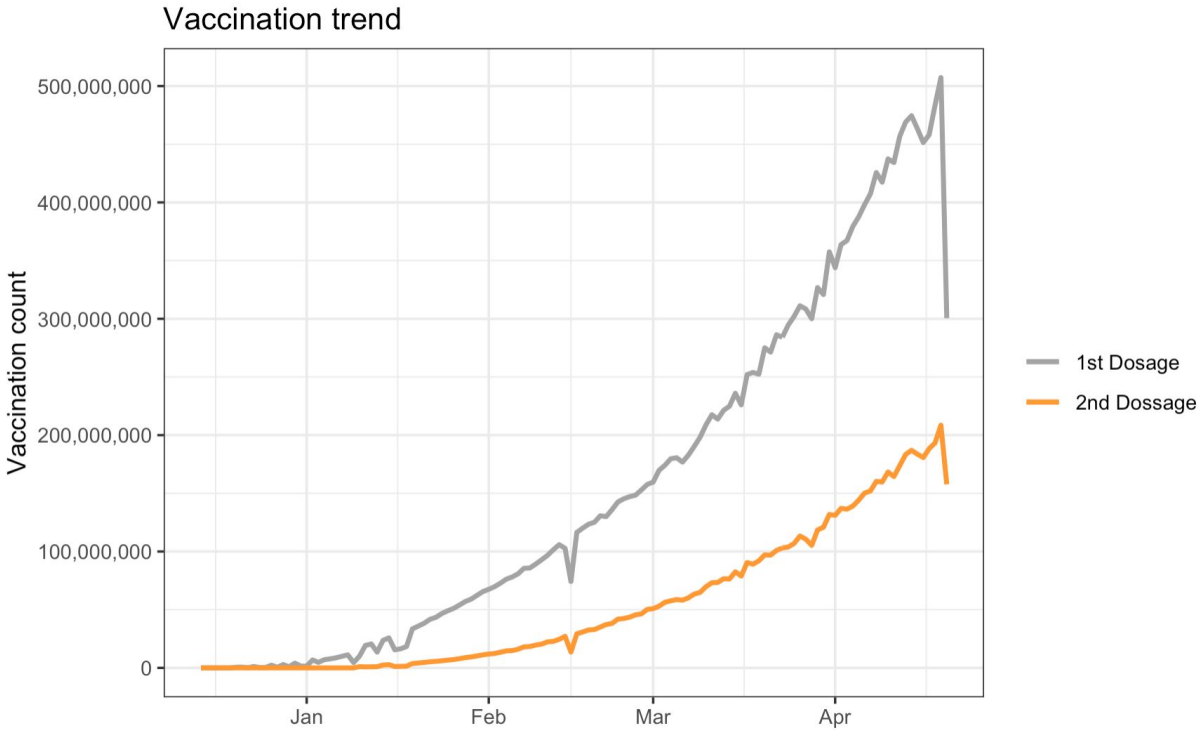
Based on total vaccinations data, US and China are leading followed by India for the maximum number of total vaccinations. This could also be because of the high population in some countries.



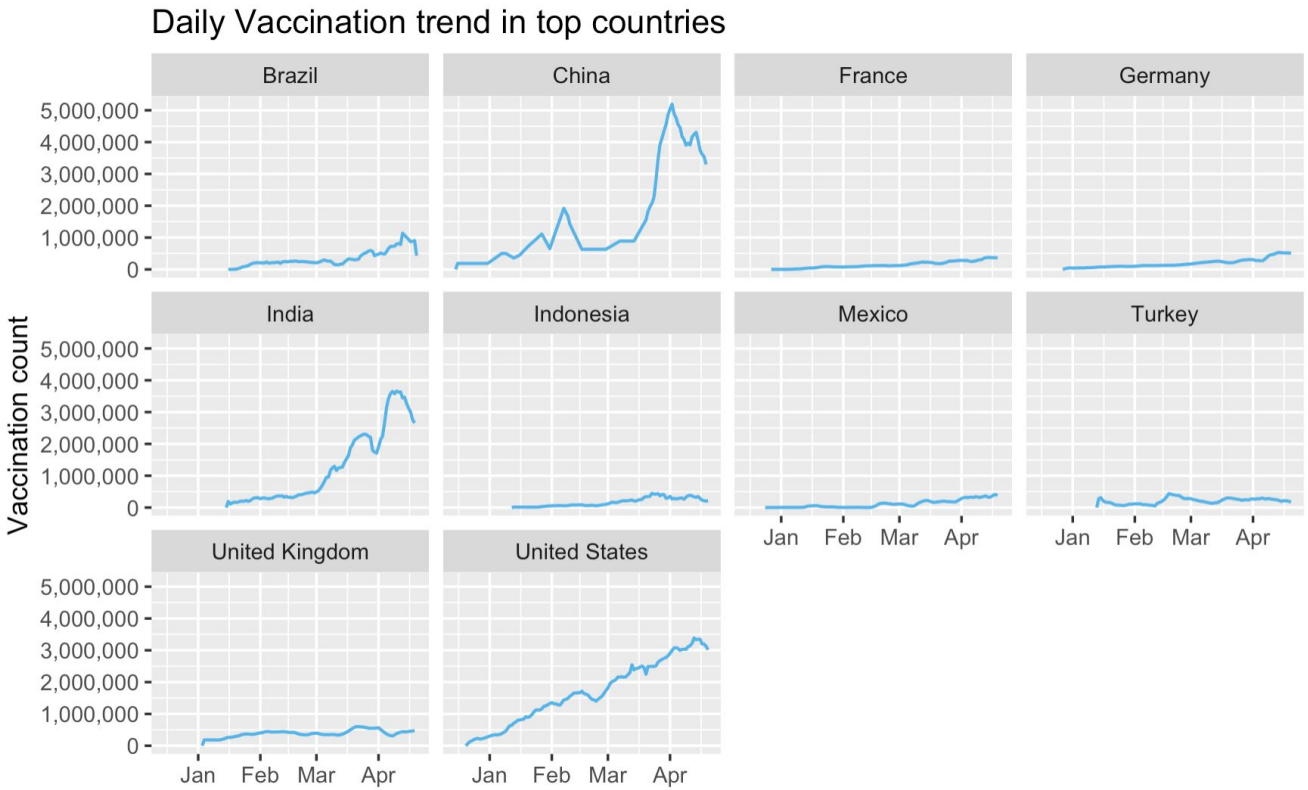
In the below plot it can be seen that US is leading in second dose vaccination. Other countries are still lagging behind. India has a large number of people who have received the 1st dose however not many have received the 2nd dose.



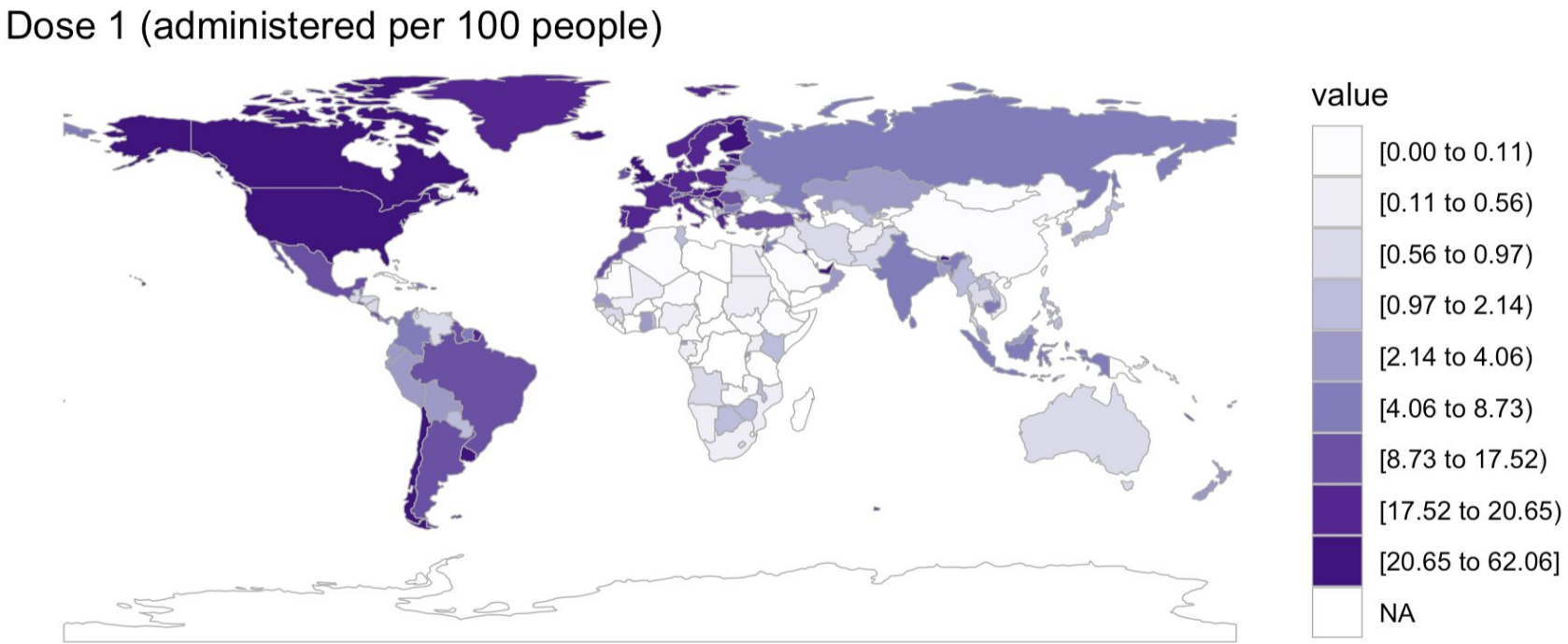
The below plot tells us the trend of both vaccination doses across the world. This is expected as 1st dose is leading and 2nd dose is catching up with the 1st dose.



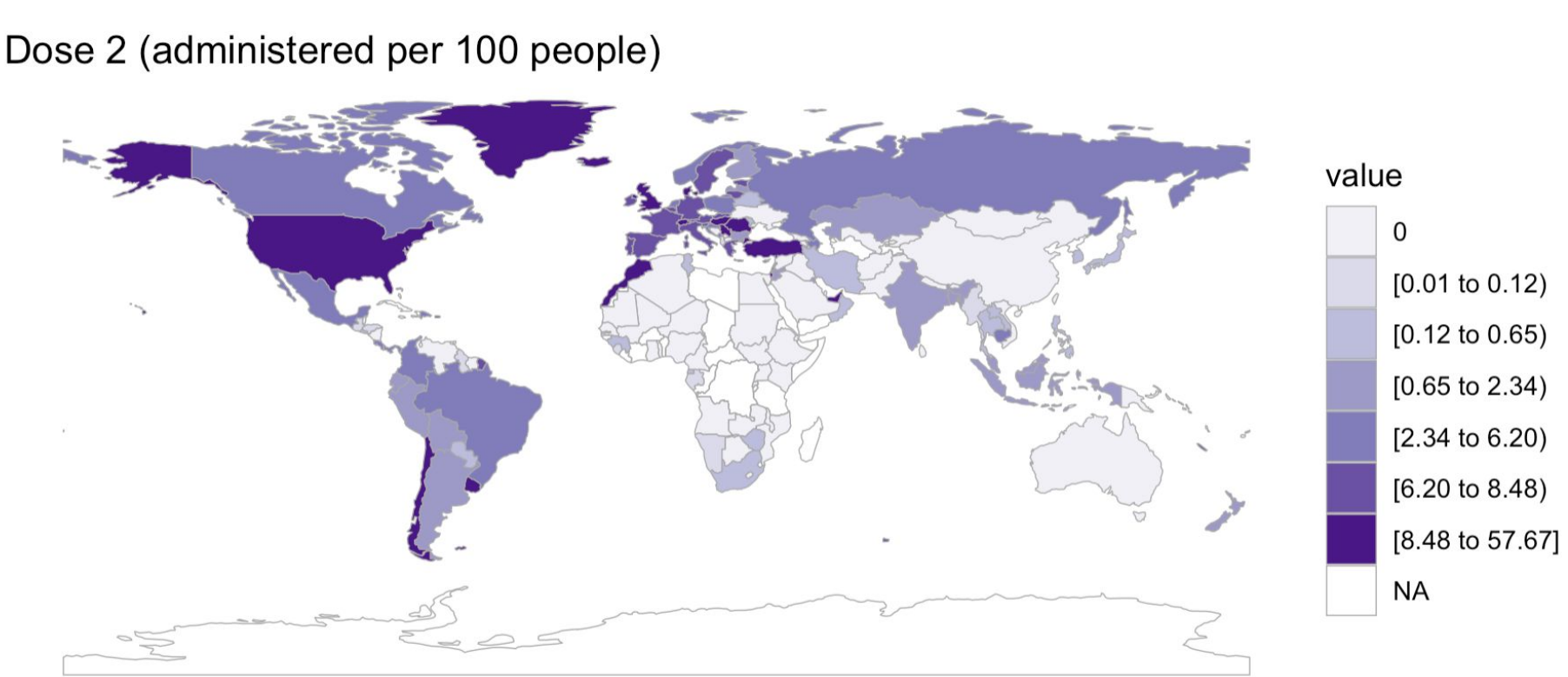
This plot gives the daily vaccination trend broken up for top countries where it is highest. Based on this it can be seen that China, US and India have higher daily vaccinations compared to other countries.This could again be affected by the population. Please not I was unable to scale the data as population of the country was not available in this data.



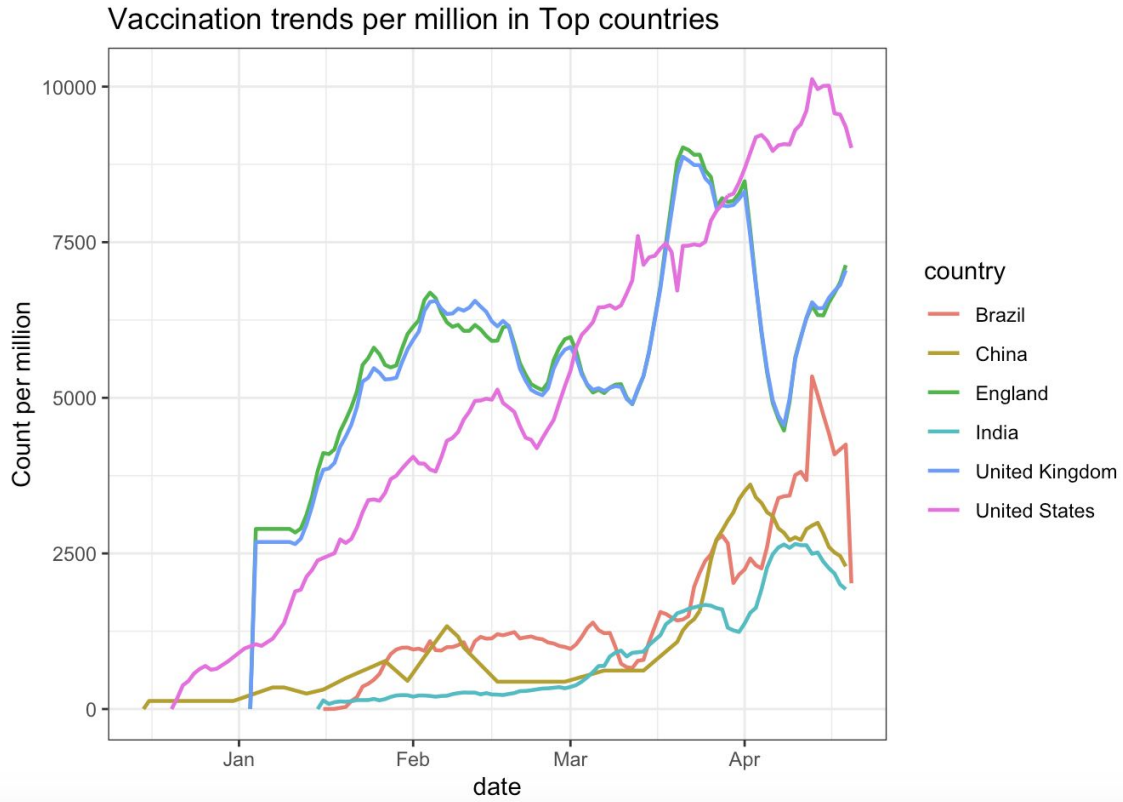
Below is a plot to visualize the status of Dose 1 vaccinations on the map. Please note here the information is per 100 people.The darker the color, the more percentage of vaccinations. US and Canada show the highest dose1 vaccinations



Below is a plot to visualize the status of Dose 2 vaccinations on the map. Please note here the information is per 100 people.The darker the color, the more percentage of vaccinations. US shows the highest dose2 vaccinations. Note that on comparing this with the previous map, it can be seen that Canada had more Dose1 and much less Dose2, its easily seen due to the light color visualization.



Below plot shows the vaccination trend per million people. It can be seen that US , UK and England are the highest. India brazil and china are comparitively lower. Based on the plot, the trend for england and UK seem to be exactly same which is weird, I plan to investigate whether these are duplicate records.



Thank you!