Data Source

The primary source of vaccination data is downloaded from the Kaggle website: https://www.kaggle.com/gpreda/covid-world-vaccination-progress

The secondary source of vaccination data is downloaded from the CDC website: https://covid.cdc.gov/covid-data-tracker/#vaccinations

The tertiary source of vaccination data is downloaded from the California of Public Health website: https://data.chhs.ca.gov/dataset/vaccine-progress-dashboard

Data Description

The primary data, country_vaccinations, is consisted of 15 columns (Table 1) and 6518 rows on March 15, 2021. The data is constantly updated to reflect the current count of the day.

Table 1: Column Description

Variables	Data Type
Country	object
iso_code	object
Date	object
total_vaccination	float64
people_vaccinated	float64
people_fully_vaccinated	float64
daily_vaccinations_raw	float64
daily_vaccinations	float64
total_vaccinations_per_hundred	float64
people_vaccinated_per_hundred	float64
people_fully_vaccinated_per_hundred	float64
daily_vaccinations_per_million	float64
Vaccines	object
source_name	object
source_website	object

The secondary data, covid19_vaccinations_in_the_united_states, is consisted of 51 columns (Table 2) and 67 rows on March 20, 2021. The data is constantly updated to reflect the current count of the day.

Table 2: Column Description

Variables	Data Type
State/Territory/Federal Entity	object
Total Doses Delivered	Int64
Doses Delivered per 100K	float64
18+ Doses Delivered per 100K	float64

Total Doses Administered by State where Administered	Int64
Doses Administered per 100k by State where Administered	float64
18+ Doses Administered by State where Administered	Int64
18+ Doses Administered per 100K by State where Administered	float64
People with at least One Dose by State of Residence	Int64
Percent of Total Pop with at least One Dose by State of Residence	float64
People 18+ with at least One Dose by State of Residence	Int64
Percent of 18+ Pop with at least One Dose by State of Residence	float64
People Fully Vaccinated by State of Residence	Int64
Percent of Total Pop Fully Vaccinated by State of Residence	float64
People 18+ Fully Vaccinated by State of Residence	Int64
Percent of 18+ Pop Fully Vaccinated by State of Residence	float64
Total Number of Pfizer doses delivered	Int64
Total Number of Moderna doses delivered	Int64
Total Number of Janssen doses delivered	Int64
Total number of doses from unknown manufacturer delivered	Int64
Total Number of Janssen doses administered	Int64
Total Number of Moderna doses administered	Int64
Total Number of Pfizer doses adminstered	Int64
Total number of doses from unknown manufacturer administered	Int64
People Fully Vaccinated Moderna Resident	Int64
People Fully Vaccinated Pfizer Resident	Int64
People Fully Vaccinated Janssen Resident	Int64
People Fully Vaccinated Unknown 2-dose manufacturer Resident	Int64
People 18+ Fully Vaccinated Moderna Resident	Int64
People 18+ Fully Vaccinated Pfizer Resident	Int64
People 18+ Fully Vaccinated Janssen Resident	Int64
People 18+ Fully Vaccinated Unknown 2-dose manufacturer Resident	Int64
People with 2 Doses by State of Residence	Int64
Percent of Total Pop with 1+ Doses by State of Residence	float64
People 18+ with 1+ Doses by State of Residence	Int64
Percent of 18+ Pop with 1+ Doses by State of Residence	float64
Percent of Total Pop with 2 Doses by State of Residence	float64
People 18+ with 2 Doses by State of Residence	Int64
Percent of 18+ Pop with 2 Doses by State of Residence	float64
People with 1+ Doses by State of Residence	Int64
People 65+ with at least One Dose by State of Residence	Int64
Percent of 65+ Pop with at least One Dose by State of Residence	float64
People 65+ Fully Vaccinated by State of Residence	Int64
Percent of 65+ Pop Fully Vaccinated by State of Residence	float64
People 65+ Fully Vaccinated_Moderna_Resident	Int64

People 65+ Fully Vaccinated_Pfizer_Resident	Int64
People 65+ Fully Vaccinated_Janssen_Resident	Int64
People 65+ Fully Vaccinated_Unknown 2-dose Manuf_Resident	Int64
65+ Doses Administered by State where Administered	Int64
Doses Administered per 100k of 65+ pop by State where Administered	Int64
Doses Delivered per 100k of 65+ pop	Int64

The tertiary data, covid19vaccinesbycounty, is consisted of 17 columns (Table 3) and 5426 rows on March 19, 2021. The data is constantly updated to reflect the current count of the day.

Table 3: Column Description

Variables	Data Type
County	object
Administered Date	object
Total Doses	Int64
Cumulative Total Doses	Int64
Pfizer Doses	Int64
Cumulative Pfizer Doses	Int64
Moderna Doses	Int64
Cumulative Moderna Doses	Int64
Johnson & Johnson Doses	Int64
Cumulative Johnson & Johnson Doses	Int64
Partially Vaccinated	Int64
Total Partially Vaccinated	Int64
Fully Vaccinated	Int64
Cumulative Fully Vaccinated	Int64
At Least One Dose	Int64
Cumulative At Least One Dose	Int64
California Flag	object

Data Cleaning

The primary data cleaning was performed in Python3. Empty cells were replaced with "0". Wale, Scotland, England, and Northern Ireland in column country replaced with the United Kingdom. The secondary and tertiary data did not need further data cleaning. Then, the data proceed to structure for data exploration.

Data Exploration

Primary data-Country and vaccines were used to group by for easy analysis. Total_vaccination, people_vaccinated, people_fully_vaccinated, and vaccines variables were selected for data

exploration and analysis. The values reported in these variables a cumulative numeric vs. day. The current day (3/15/2021) is reported the max or cumulative value of people vaccinated at this moment.

	Country	Total vaccinations	Vaccines
0	Albania	23635.0	Pfizer/BioNTech
1	Algeria	75000.0	Sputnik V
2	Andorra	4914.0	Pfizer/BioNTech
3	Angola	6169.0	Oxford/AstraZeneca
4	Anguilla	3929.0	Oxford/AstraZeneca

Figure #1: Grouped by country for total vaccinations and type of vaccines

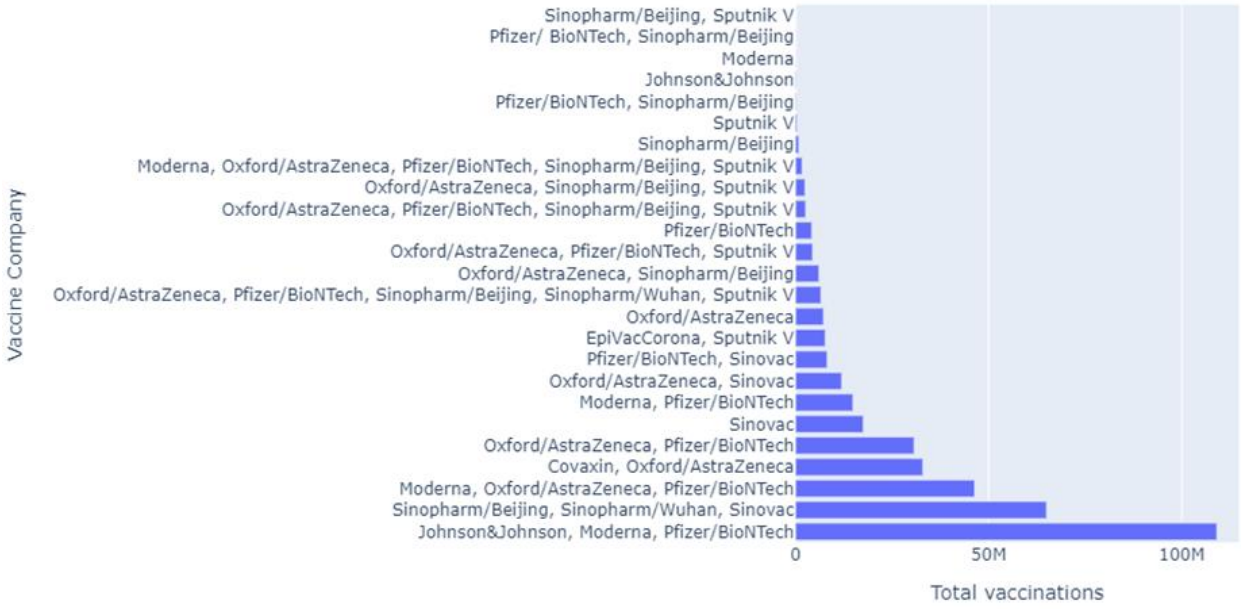


Figure #2: Group by type of vaccines to reflect total vaccinations.

Secondary data – US vaccination data was only focused on California. State/Territory/Federal Entity, Percent of Total Pop with at least One Dose by State of Residence, Percent of Total Pop with 2 Doses by the State of Residence were selected for data exploration and analysis. The values reported in these variables a cumulative numeric day. The current day (3/15/2021) is reported the max or incremental value of people vaccinated at this moment. Data is incorporated into the primary data.

Tertiary data – California county vaccination data was focused on Northern California counties. County, total_partially_vaccinated, cumulative_fully_vaccinated, cumulative_pfizer_doses, and cumulative_moderna_doses were selected for data exploration and analysis. The values

reported in these variables a cumulative numeric day. The current day (3/20/2021) is reported the max or incremental value of people vaccinated at this moment. Data is incorporated into the primary and secondary data.

Research Questions:

1. Which vaccine did each country use?
2. What is the percentage for total vaccination for the US up to date compared to certain countries (TBD) vs. time?
3. With the California vaccination site open, does the California trending reflects the US report from world vaccination progress data?
4. How the Northern California counties (Contra Costa, Alameda, Marin, and San Francisco) performed among themselves?

Brief description of the Jupyter program

The program is performed in Python3.8 via Jupyter editor. Packages used for the data are Pandas, Numpy, Matplotlib.pyplot, and plotly.express. Three data frames are structured by country, vaccine, counties, and state for ease of analysis. Top 10 viewed for data exploration.

Primary data - Sub data frames were generated to dive into the detail of five countries (US, Israel, Hungary, Denmark, and Malta). Data analysis was performed, and two bar charts were generated. Another sub-data frame were generated to analyze the vaccines used by 138 countries, and one figure was generated.

Secondary data – Sub data frame only pulls California data for the percentage of partially vaccinated and fully vaccinated people in the state. These results were incorporated with the primary data to generate one bar graph.

Tertiary data- Sub data frames were generated to dive into the detail of five counties (Marin, Contra Costa, Alameda, and San Francisco) based on number of people vaccinated and the type of vaccine. Data analysis was performed, and three bar charts were generated.

Description of the output files

Outputs consisted of:

Primary Data

1. df_covid_vaccination, the data frame showed the first ten rows.
2. df_covid_vaccination, the data frame showed the last ten rows.
3. Type of data consisted in the data frame.
4. The statistical description of the data
5. Group data frame by country, top 10, and total_vaccinations
6. Group data frame by country, top 10, and people_vaccinated
7. Group data frame by country, top 10, and people_fully_vaccinated

8. The subset of a data frame from the United States
9. Subset of data frame from Israel
10. Subset of data frame from Hungary
11. Subset of data frame from Denmark
12. Subset of data frame from Malta
13. Bar graph of population
14. Bar graph of vaccination
15. Bar graph of Percentage Vaccination of Partially vs. Fully
16. Subset of vaccine by country
17. Figure of vaccine by country
18. Subset of vaccine by total vaccination
19. Figure of vaccine by total vaccination

Secondary Data

20. df_covid_vaccinationUS the data frame showed the first 10 rows.
21. df_covid_vaccinationUS the data frame showed the last 10 rows.
22. Type of data consisted in data frame.
23. Statistical description of the data
24. Subset of data frame from California
25. Bar graph of Percentage Vaccination of Partially vs. Fully

Tertiary Data

26. df_covid_vaccinationCA the data frame showed the first 10 rows.
27. df_covid_vaccinationCA the data frame showed the last 10 rows.
28. Type of data consisted in data frame.
29. Statistical description of the data
30. Subset of data frame from Marin
31. Subset of data frame from Contra Costa
32. Subset of data frame from Alameda
33. Subset of data frame from San Francisco
34. Bar graph of Percentage Vaccination of Partially vs. Fully
35. Bar graph of Northern California county's population
36. Bar graph of Vaccines used by county.

Results

The recent top three countries in the European Union (EU), Denmark, Hungary, and Malta, have the highest vaccination rates. These countries are following the Israel vaccination model to achieve herb immunity. By comparison, United States still has the highest vaccination numbers of the other countries (figure 3). However, Israel has the highest percentage of its population

being vaccinated (figure 4). Results were normalized by population number. Total vaccination showed Hungary has the highest doses administered in the EU. However, Denmark has the highest percentage in population for partially and fully vaccinated.

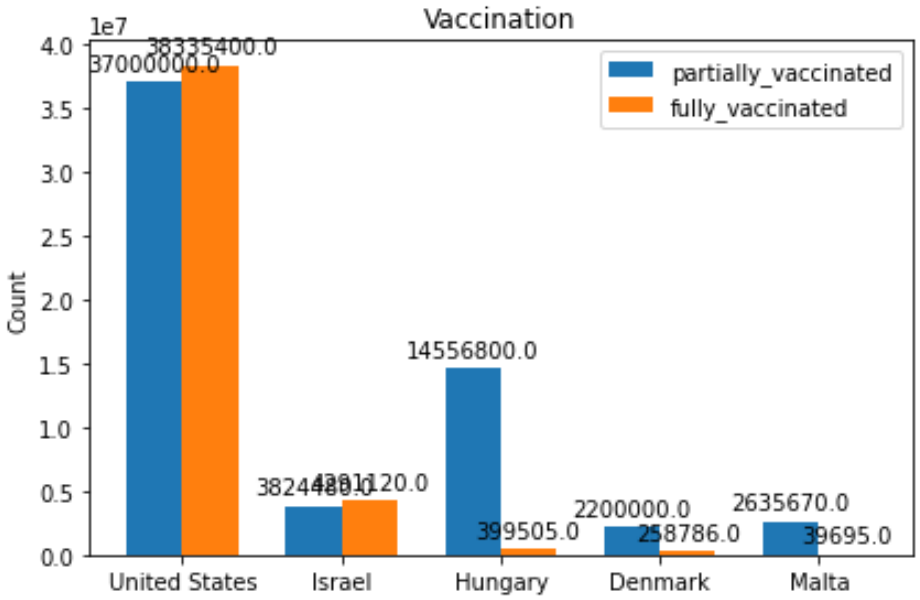


Figure 3: Countries based on total_vaccination

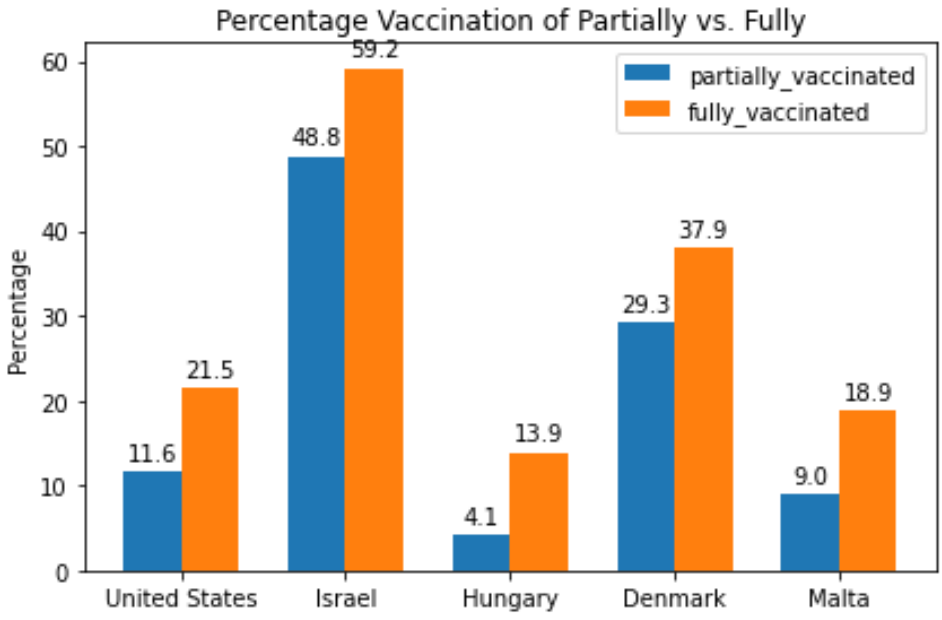


Figure 4: Countries based on the percentage of people_vaccinated and normalized by population

One hundred million doses of vaccines were dosed throughout the globe (figure 5). Table 4 shows which type of vaccines dosed at these countries.

Figure 5: Top Countries based on people_fully_vaccinated

Table 4: Type of vaccines dosed

Country	Vaccines
United States	Johnson&Johnson, Moderna, Pfizer/BioNTech
Israel	Moderna, Pfizer/BioNTech
Hungary	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Denmark	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Malta	Pfizer/BioNTech

California data incorporated to the primary data to compare against the US and the other countries to determine how the state performed in its vaccination program. California seemed to have a higher percentage of its population administered at 1st dose due to several mega vaccination sites (Figure 6). The rate of fully vaccinated people is lower than the overall US percentage and the other EU countries.

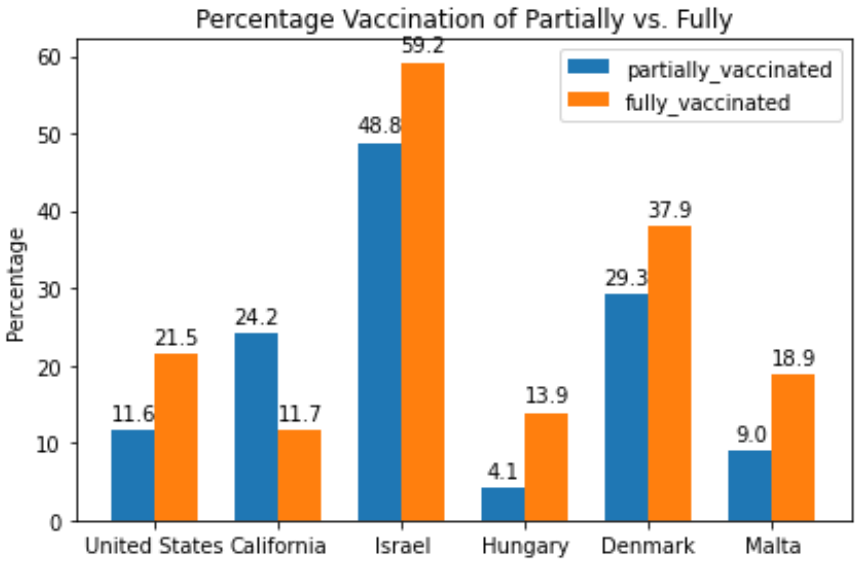


Figure 6: Countries and California comparison based on the percentage of people_vaccinated and normalized by population

Northern California data incorporated into the data to compare against the US, California, and the other countries to determine how the counties performed in their vaccination programs. Marin county seemed to have a higher percentage of its population administered partially and fully due to being a wealthy and smaller county (Figure 7) to the other counties. The pace of vaccination seemed to be comparable between Contra Costa, Alameda, and San Francisco. Compared to the US, California, Israel, and Denmark, the counties are still slow in rolling out their vaccination programs. But the counties performed better or comparable to Hungary and Malta.

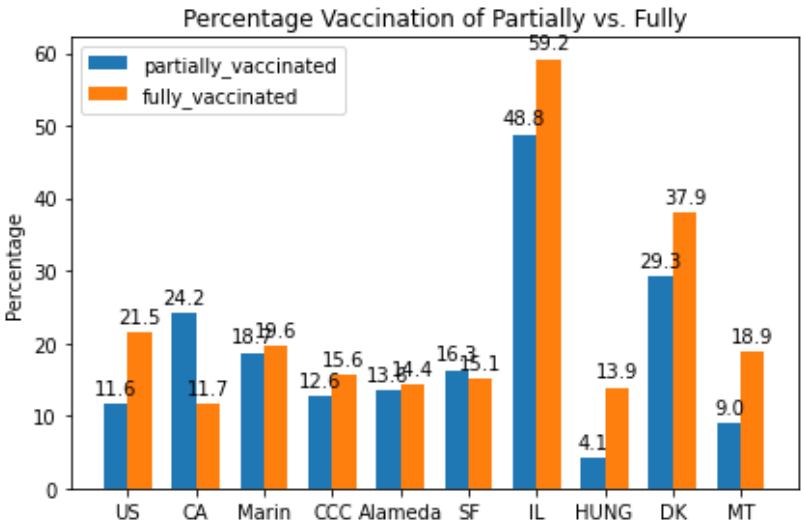


Figure 7: Countries, California, and counties comparison based on the percentage of people_vaccinated and normalized by population

Further comparison between the counties in the population (Figure 8), Alameda is the biggest county, and Marin is the smallest. The number of Pfizer or Moderna vaccines dosed seemed to be proportional to the population (Figure 9). There is no indication of one county has more vaccine than the other counties—Marin county's explanation for performing better due to its demographic. The county has a more significant well-educated and wealthy middle-class population than the other counties, in which people are more inclined to be readily vaccinated. Also, the county is small, and the vaccination program might be easier to roll out.

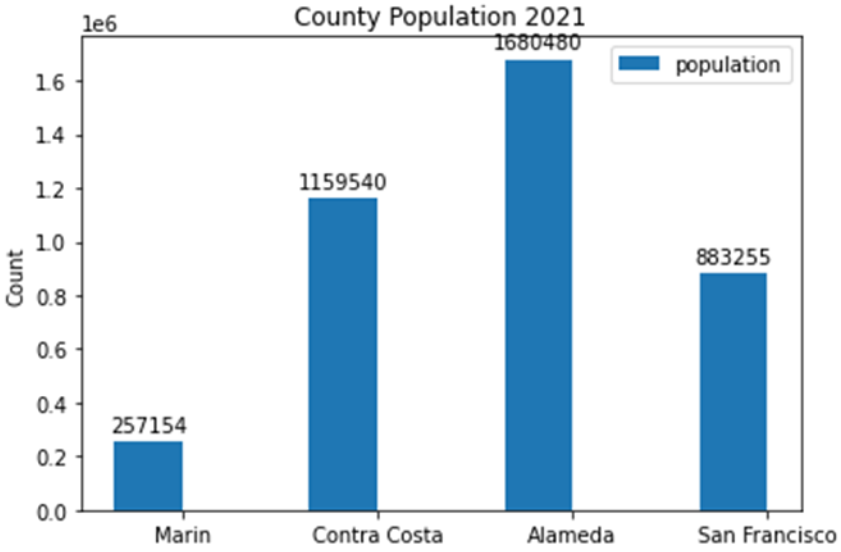


Figure 8: Northern California county population

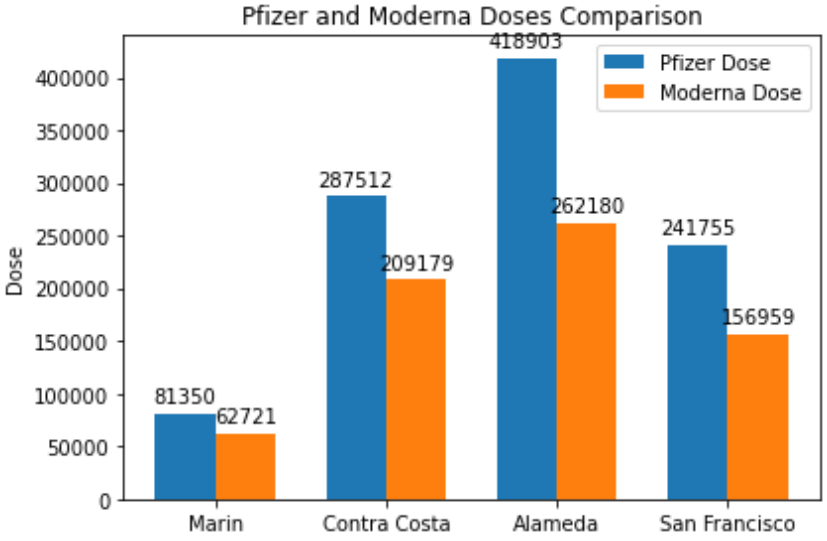


Figure 9: Number of Pfizer and Moderna doses by county

Conclusion

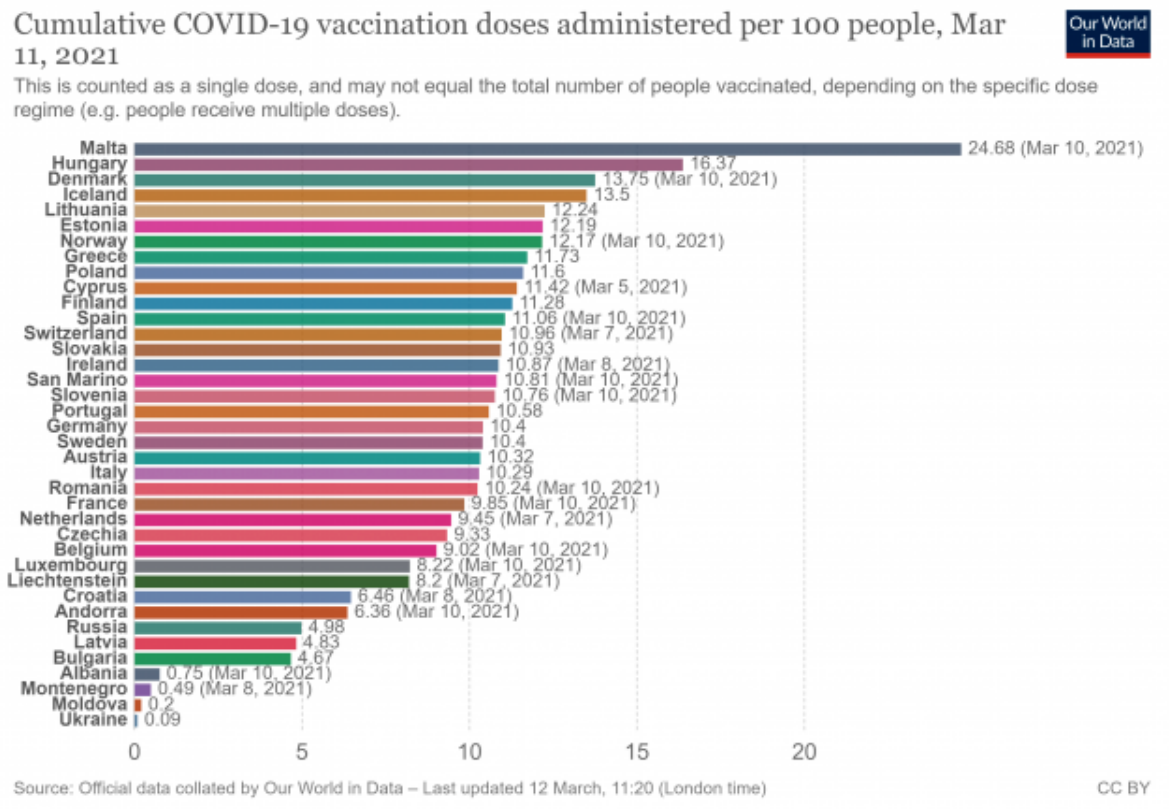
The United States seemed to be leading the number of doses for vaccines with immediate access to three approved vaccines, Pfizer, Moderna, and Johnson & Johnson. However, the percentage of people vaccinated remained slower than in Israel. Also, three EU nations, Denmark, Hungary, and Malta, adapt the Israel vaccination program to achieve herb immunity by rolling out their plans quickly. Compared to the US population, the other countries have a significantly smaller population to achieve their herb immunity goal promptly.

Compared to the US, California is slower than the US in the percentage of people fully vaccinated but is ahead in partially vaccinated. Northern California counties showed a higher percentage of people fully immunized in the state, which indicated that some counties have a lower rate. These results further suggested that the county does not reflect the trend of the state. Compared among the counties, a smaller county like Marin has a better vaccination program than the bigger counties. This finding might confirm the data shown in the US and Israel comparison.

In conclusion, the United States is rolling out the vaccination quickly and meeting two million doses per day. However, the slow rollout is still observed in California and some of its counties due to vaccine shortage. Hopefully, more vaccines will be available soon before a fourth surge occurred. The vaccination data generated from Israel and the EU countries might indicate how the US vaccination will be performed once 50% of the population is fully vaccinated.

Malta, Hungary and Denmark topping EU vaccine tables - what can we learn from them?

Different approaches to the roll-out have been taken across the EU.



IRELAND'S VACCINE ROLL-OUT has slowed down, with continued delivery shortfalls hampering the health service's progress.

During the week HSE chief Paul Reid expressed his frustration in particular with AstraZeneca, stating that repeated and sometimes last-minute reductions to deliveries had "rocked" confidence in the company.

Officials have stressed that Ireland is, by European Union standards, working its way through the population at a decent pace, ahead of some of the bloc's big players like France and Germany.

But a small number of countries in the EU have steamed ahead, despite the fact that they are part of the same vaccine pool, with an entitlement to doses proportionate to their population.

https://www.thejournal.ie/malta-hungary-denmark-vaccines-5379164-Mar2021/?utm_source=shortlink&utm_campaign=email_share