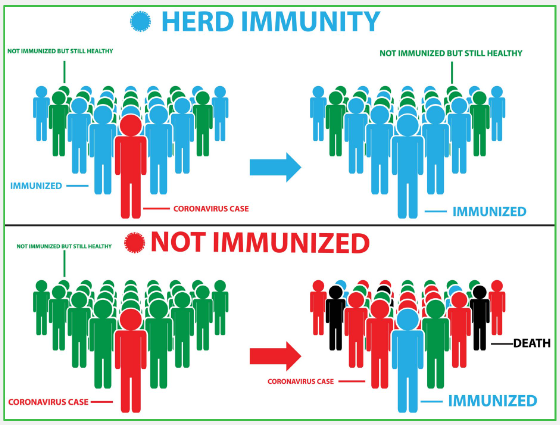
**COVID-19 Vaccination Progress – Worldwide**



Data Analysts: Paige Singleton, Sherali Yadav, Tara Polli, Varnika Rachupally, Yong Yang

Questions to Explore

The COVID-19 pandemic has been affecting every person throughout the world for the past year. Most countries are in the vaccination phase, making vaccine data readily available and current data. We decided to explore the vaccination process to answer our main question:

1. **Which** **country may reach herd immunity the soonest?** Furthermore,
2. Do countries who lead in distributing COVID-19 vaccinations experience a subsequent decline in the number of COVID-19 confirmed cases?
3. Which country is the most successful in rolling out the COVID-19 vaccine and what vaccine(s) are they using?

Data Sources

* **country\_vaccinations.csv**: Using this [link](https://www.kaggle.com/gpreda/covid-world-vaccination-progress) from Kaggle, data is collected daily from [Our World in Data](https://ourworldindata.org/) GitHub repository for COVID-19. Size: 1846 rows, 16 columns
* **owid-covid-data.csv** is a larger file containing similar data from country\_vaccinations, also available from [Our World in Data](https://covid.ourworldindata.org/data/owid-covid-data.csv). Size: 63,573 rows, 61 columns

Data exploration and cleanup process (see Jupyter notebook)

1. Merge two dataframes using a ‘left outer’ join. Save new csv file in resources folder.
2. As a group, review all columns in merged DataFrame to decide which columns we may want to use versus omit. The dataset contains similarly-named columns (ie, new\_cases, new\_cases\_smoothed).
3. Since population is such an important column of data, we confirmed this and other frequently-used data appears in the DataFrame before continuing.
4. Remove rows missing an iso\_code (England, Scotland, N. Ireland, Wales data stored in GRB is\_code representing the United Kingdom.). Save the new csv file in resources folder.
5. Reorganize the columns in the DataFrame, utilizing the columns we assume we will be using the most frequently.

*The above steps were conducted as a group, then a clean starter file was pushed to the main branch for everyone to pull into their own branches.*

Analysis (see Jupyter notebook)

* *Total Vaccines Per Hundred vs People Fully Vaccinated Per Hundred* were plotted and analyzed by different team members. When cross-comparing which countries appeared in the top five for each of these plots, only Israel appeared in same position (#1) in both. Why? Because fully vaccinated data means any country distributing & completing the 2-dose regimen are counted in the running total for people fully vaccinated.
* Because Israel appears to successfully vaccinate their population, two line graphs showing *Israel’s New Daily Cases* and *People Fully Vaccinated Per Hundred* were visualized. In both graphs, it appears as if the number of new cases is diminishing over time as more people are vaccinated.
* *Pharmaceutical companies* have been rushing to produce and distribute vaccines. We plotted the company showing the highest number distributed. Leading companies: Pfizer and Moderna.
* Lastly, we *categorized each country by population size* (small = < 4,000,000, medium up to 18,000,000, large up to 1.5 billion). Next, we compared the top five countries in each category. Perhaps the smallest countries have fewer people to vaccinate; therefore, they reach herd immunity the soonest? Please see our conclusions:

Conclusions

As of February 2021, it appears as if Israel (population of 4,937,796) is the most successful in vaccinating their people, and may reach herd immunity the soonest. We assume richer nations such as the United States, United Arab Emirates, and United Kingdom are well on their way to herd immunity, but you must factor in population and new cases. While the US ranks high in the fully vaccinated category, our population is simply too large to win this race.