The COVID-19 Vaccines And Its Progress

Project Report

COSC3000 - Visualization and Data Analysis

I Made Gede Surya Nirvana

46638876

Bachelor of Computer Science

School of Information Technology and Electrical Engineering

The University of Queensland

2021

Table of Contents

List of Figures

Chapter 1

Introduction

In December 2019, the World Health Organization, or WHO, announced that there are dozens of people in Wuhan, China suffers from an unknown pneumonia-like disease. These people share a similarity between them: they previously had visited the live animal market in Wuhan [1]. Later in January 2020, the root cause is identified, a novel coronavirus temporarily named 2019-nCoV and subsequently named COVID-19 virus.

Seeing the increase of the infection, the WHO officially declared a public health emergency to the world by the end of January 2020 [2]. Several countries are already affected by that time, but the overall case outside China is still pretty minimum. Hence some people are still pretty chill about it. After two months, there are still cases, and it rises pretty significantly, which makes WHO declared the outbreak of the COVID-19 virus as a global pandemic [2]. Ever since then, the popularity of face protection tools, especially face masks, has skyrocketed. Due to the increase in demand for face protection tools, WHO mentioned that some countries suffered from a shortage.

At that time, there is currently no vaccine to treat the COVID-19 virus. Scientists then started the development of the vaccine. They predicted that the vaccine should be ready by the end of 2020. This prediction came true because in summer 2020 (between July and August), there are already vaccines that have gone through the first clinical trial [3]. By December 2020, several vaccines have been approved to be used, and countries worldwide start order and use the vaccine.

1.1 Aims

It has been several months since the first COVID-19 virus vaccine first used. Since then, several questions have surfaced throughout the days, such as:

- What is the current progress of people who have been both vaccinated and fully vaccinated in each country worldwide?
- How many doses of vaccines were given to the people in each country worldwide?
- How many kinds of vaccines for the COVID-19 virus currently exist today worldwide, and which country uses which vaccines?
- On average, how many types of vaccines used for each country worldwide?

The main objective of this project is to explore, highlight patterns and trends, and present the data in a visualized form that is targeted towards everyone in the world. The trends and patterns can be helpful as it shows the insights of the current vaccination progress as well as the vaccines used in each country between December 2020-April 2021. It is also useful to predict the number of total vaccinations, people who have been vaccinated and fully vaccinated based on daily basis. It creates more awareness for people to be more prepared to handle this pandemic.

1.2 Methodologies

The project is mainly done using Tableau, Tableau Prep, and Python with extra help from some libraries such as `pandas` and `BeautifulSoup`. The reason being the familiarity compared to other tools, and also Tableau is the perfect software for data analysis and visualization. Tableau Prep is very useful to clean and preprocess the data. On the other hand, Python is helpful for web scraping, which will be explained in the next section. Visualization techniques that are used for the project will be discussed throughout the report.

This project can only be accessed in my GitHub repository (https://github.com/suryanirvana/covid-19-vaccinations) since I don't have any license for neither Tableau Online nor Tableau Server.

Chapter 2

Data Analysis

2.1 Data Collection

Sourcing a dataset is the first challenge for this project. But today, there are currently tons of websites that already have credible datasets that can be used, from government websites to even Kaggle. Kaggle is a website for data science and machine learning practitioner that allows us to find and use a dataset, analyze, and build a machine learning model for it. The first dataset that is used is for the countries vaccination. The dataset is listed as a CSV file format consisting of 12970 rows of countries that have vaccination already [2.1]. The dataset is a mixture of number and text entries, where the majority of the text data is already written in formal English language. Hence, it doesn't add an extra challenge to take care of the text data. Each data record or row contains

- Country name,
- ISO Code for each country,
- Date when vaccination occurred,
- Doses of vaccines (labeled as Total Vaccinations) that have been given in each country,
- Numbers of People who have been Vaccinated in each country,
- Number of People Fully Vaccinated in each country or people that have received the entire set of immunization,
- Daily Vaccinations that occurred in each country,
- Types of Vaccines used,
- Source of information, and

- Website of the source of information.

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_h
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	NaN	1367.0	
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	NaN	1367.0	
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	NaN	1367.0	

Figure 2.1: Countries Vaccination Dataset

The second dataset came from the Worldometer website. The website provides live statistics of almost everything, from population numbers, government, economics, up to the number of Google searches that day. The population numbers in each country are the data used in this project, and to retrieve it, we can scrape the website. Web scraping is the process of harvesting or extracting data from a website, where later can be used to predict, analyze the data. It may seem intimidating, but it's easy to do using Python because it already provides a library called `BeautifulSoup` that allows to do that. The population dataset consists of 235 rows of countries with several other values, but only two features that we're interested in, the **country name** and its **population number** [2.2].

	Country	Population
0	China	1439323776
1	India	1380004385
2	United States	331002651
3	Indonesia	273523615
4	Pakistan	220892340

Figure 2.2: Population Dataset

2.2 Data Preprocessing

Before analyzing and visualizing the dataset, transforming the data into an understandable or appropriate format is needed. The first step is Data Preprocessing, cleaning, and preparing our data for data analysis and visualization. This step is crucial because it ensures that higher quality data is used to get higher accuracy of insights. Based on the characteristics and the requirements of the data, the following are the preprocessing stages:

1. **Column Removal**

   As listed in the previous section, there are lots of columns or features in both the countries vaccination and the population dataset. Since this project will not use all of them, they can be removed or hide. For the countries vaccination dataset, Tableau Prep provides an easy way to remove the feature. After loading the dataset, simply uncheck the feature to remove it. On the other hand, for the population dataset, since Python is already used to retrieve the data, removing the feature also conducted using Python. The `drop()` function that is built-in in the `pandas` library can do the job.

2. **Text Value Separation**

   Values that are needed for the visualization process must be in an independent form or a single form. Consider this; some countries use several vaccines such as **Pfizer** and **Sinovac**, the data will be recorded as **'Pfizer, Sinovac'**. We can't count them as a single value because we will end up with too many values (or combinations of value). In consideration of that, we must separate those two values to count them separately (e.g., **Pfizer** and **Sinovac**). Tableau Prep can help to split the value and then reorganize (pivot) the dataset. After this step, the number of rows increases to 64850.

Note that some rows in the countries vaccination dataset have a null value in a particular feature. For that, it is necessary to delete the empty rows, but Tableau already provides a feature to filter the null data. Therefore, removing null rows is not needed. After these preprocessing steps above, the number of rows in countries vaccination dataset increases to 64850 [2.3].

country	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations	vaccines	vaccines combined
Afghanistan	02/22/2021	0	0	*null*	*null*	Oxford/AstraZeneca	Oxford/AstraZeneca
Afghanistan	02/22/2021	0	0	*null*	*null*	*null*	Oxford/AstraZeneca
Afghanistan	02/22/2021	0	0	*null*	*null*	*null*	Oxford/AstraZeneca
Afghanistan	02/22/2021	0	0	*null*	*null*	*null*	Oxford/AstraZeneca
Afghanistan	02/22/2021	0	0	*null*	*null*	*null*	Oxford/AstraZeneca
Afghanistan	02/23/2021	*null*	*null*	*null*	1,367	Oxford/AstraZeneca	Oxford/AstraZeneca
Afghanistan	02/23/2021	*null*	*null*	*null*	1,367	*null*	Oxford/AstraZeneca
Afghanistan	02/23/2021	*null*	*null*	*null*	1,367	*null*	Oxford/AstraZeneca
Afghanistan	02/23/2021	*null*	*null*	*null*	1,367	*null*	Oxford/AstraZeneca
Afghanistan	02/23/2021	*null*	*null*	*null*	1,367	*null*	Oxford/AstraZeneca

Figure 2.3: Clean Countries Vaccination Dataset

Chapter 3

Results

3.1 The COVID-19 Vaccine Types

Three types of visualizations are created for all of the COVID-19 vaccine types. The stacked bar graph use to visualize different types of vaccines used in each country. As seen from figure 3.1, there are currently (April 2021) 11 types of vaccines produced within a different country. The United Arab Emirates, Mexico, and Hungary being the countries that use the most types of vaccines, which are five types.

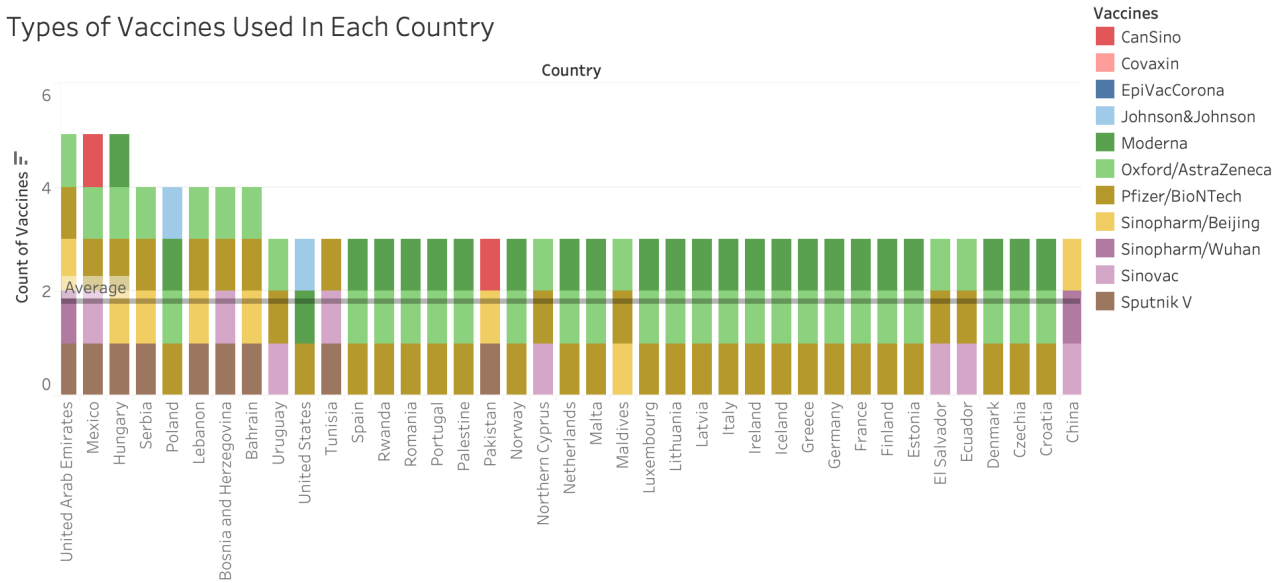

Figure 3.1: A stacked bar graph representing the different types of vaccines used in each country

Both Oxford/AstraZeneca (from the United Kingdom) and Pfizer/BioNTech (from the United States) have become the most used type of vaccine globally, with the number of countries using them shown in figure 3.2. The distribution across these two widely used vaccines around the globe are shown in

figure 3.3. From these two vaccines, we can see a pattern that most countries using the vaccine from a certain country are the neighboring ones, or they both have a relationship in politics, economics, etc. like the United Kingdom and Australia.

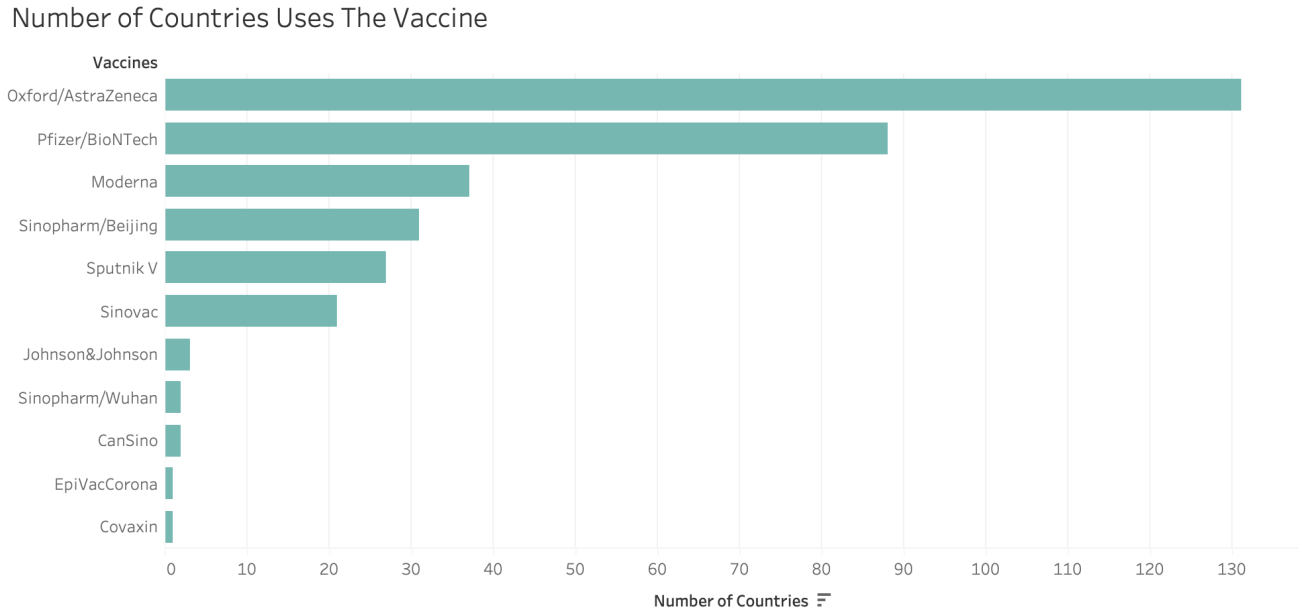

Figure 3.2: A bar graph representing the number of countries using the vaccine

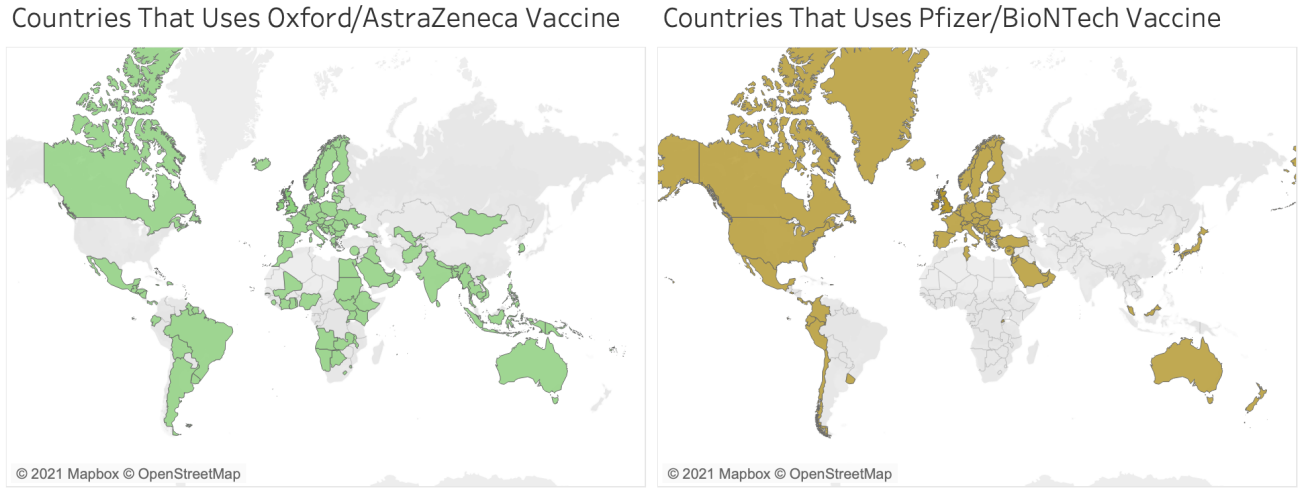

Figure 3.3: A world map representing the country that uses Oxford/AstraZeneca and Pfizer/BioNTech vaccine

On average, there are 1,81 types of vaccines used overall, which indicates that each country can use more than one type of vaccine. Thus, using a categorical world map in figure 6.1 where it assigned one type of vaccine to each country, will make no sense.

3.2 The Progress of People Who Have Been Vaccinated

Figure 3.4 shows the mapping for the number of people who have been vaccinated in each country, where the darker color shows a higher number. The United States and India have become the countries with the most number of people who have been vaccinated, with 133,2 million and 109,7 million respectively. At the same time, several countries such as China and Saudi Arabia have no data regarding the number of vaccinated people. Most of these countries that have no data are located in Africa. It is unclear whether there is actually no data or they have not yet given any vaccines in those countries.

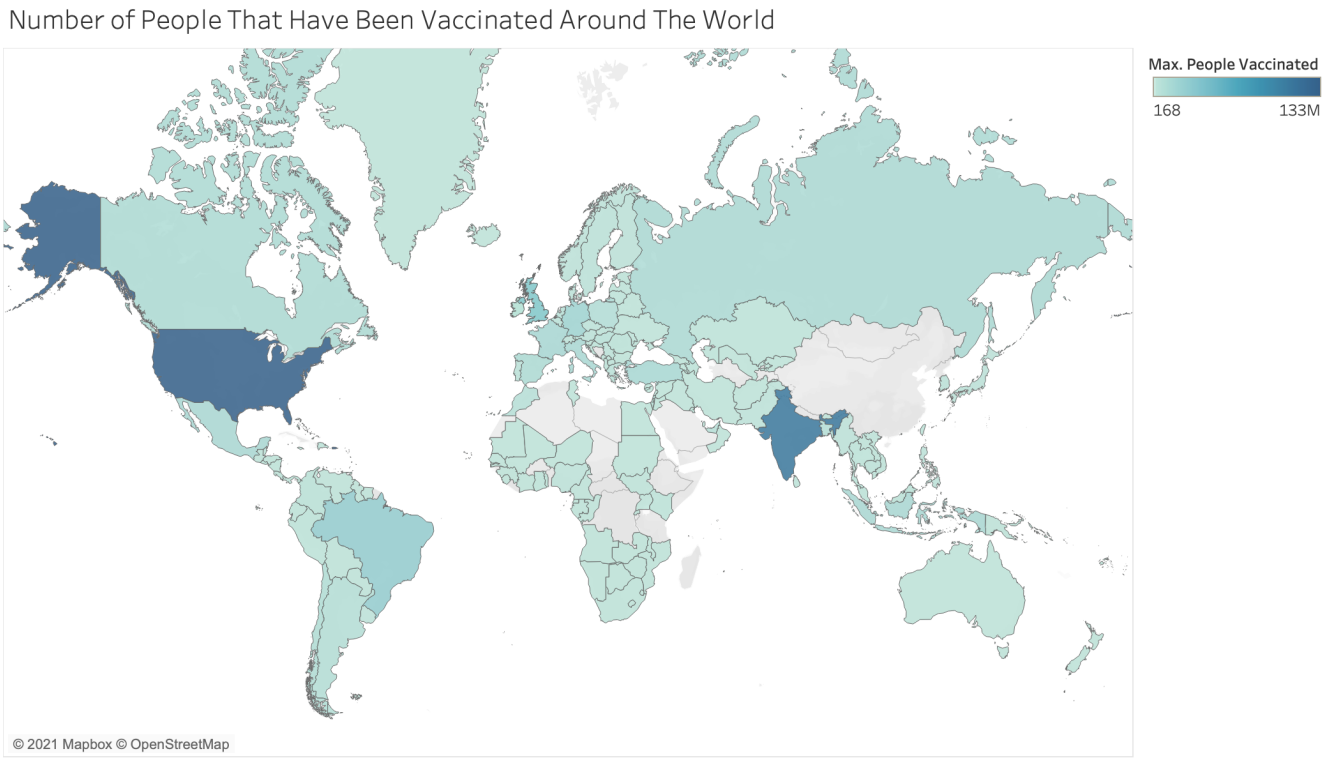

Figure 3.4: A color-coded world map representing the number of people who have been vaccinated from December 2020 to April 2021

It seems like the United States and India have covered a lot of people in their country, but it turns out they haven't. Figure 3.5 shows how many percent it covers in each country. At only a glance, the world map shows not much difference between countries; it looks all the same because small countries are hardly visible in Tableau. It's a challenge to determine which country has covered most of the population. Hence the presence of the bar graph to clarify this.

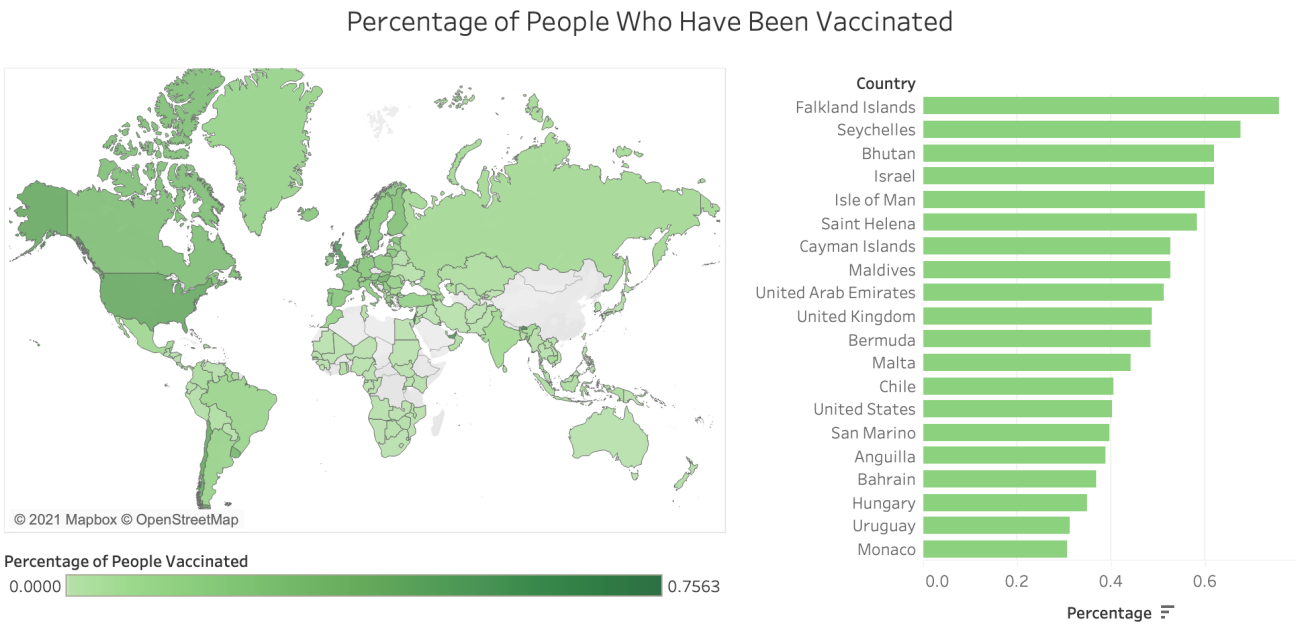

Figure 3.5: A color-coded world map and a bar graph representing the percentage of people who have been vaccinated from December 2020 to April 2021

From the bar graph, the people who have been vaccinated in the United States cover only 40% of the population. In contrast, small countries, such as the Falkland Islands, already cover up to 76% of its population. This shows that even though the country has many people that are already vaccinated, it doesn't indicate that the country already covered a large percentage of people.

As also seen from the bar graph, there are only nine countries that have covered half of their population for vaccinated people. This shows that we're still far from covering the whole population in each country. It could take months before everyone is vaccinated.

3.3 The Progress of People Who Have Been Fully Vaccinated

A vaccinated person doesn't mean that this person has received the entire set of immunization to the disease since, depending on the type that they received, they are required to take more than one dose [4]. Using the same methodologies as before to compare the difference between those who have been vaccinated and fully vaccinated.

Figure 3.6 shows the frequency of people who have been fully vaccinated in each country. The United States stands out the most with 86,2 million people. The difference in number (133,2 million for

vaccinated people) shows the truth that people who have been vaccinated don't mean that they received the complete immunization; there is a difference. Compared to before (figure 3.4), more countries have no data on fully vaccinated people. One of the possibilities being these countries haven't yet received any vaccines.

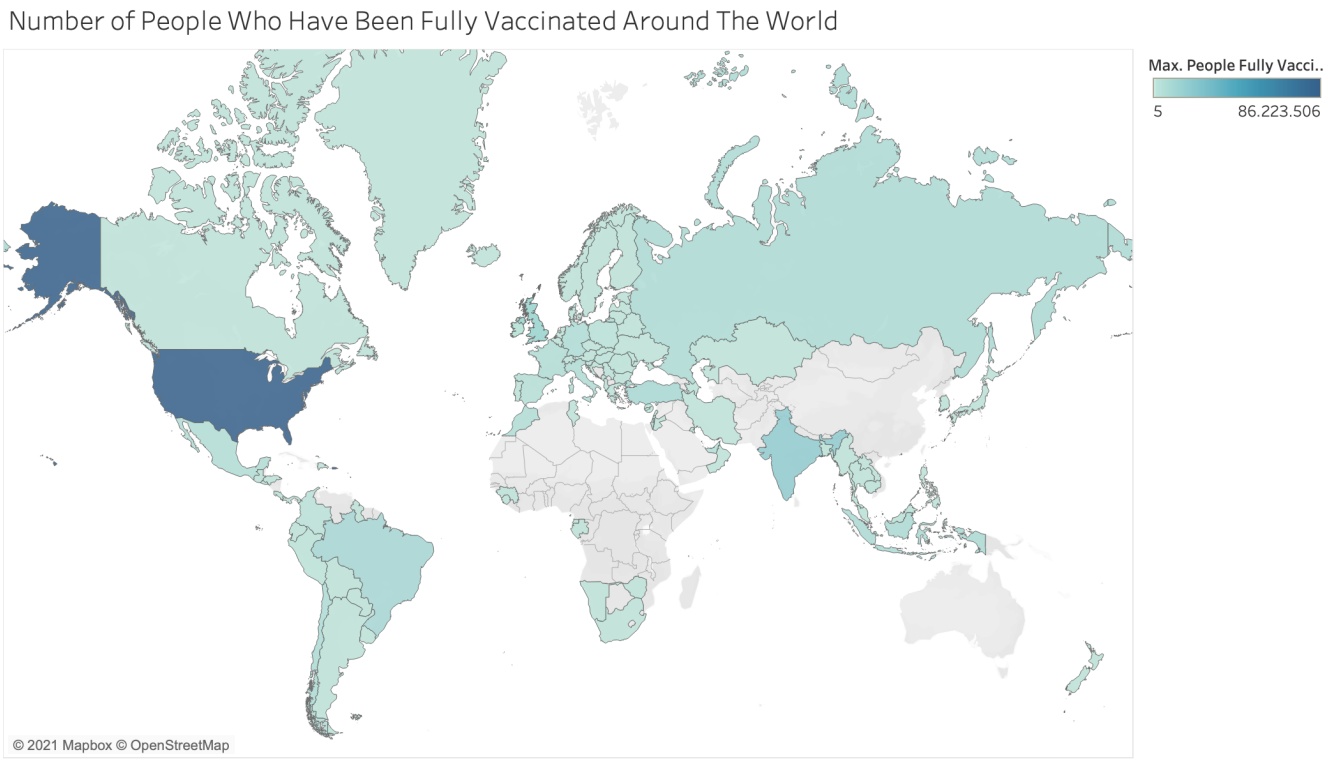

Figure 3.6: A color-coded world map representing the number of people who have been fully vaccinated from December 2020 to April 2021

Figure 3.7 shows the percentage that it covers over the population in each country. We can see that even though some small countries have a much lower number of people who have been fully vaccinated, they have covered a large proportion of their population. For example, Gibraltar already covered 91,4% compared to the United States, where it only covered 26% of its population. This has the same meaning as before; the larger number of people doesn't indicate a larger proportion that has been covered in that country.

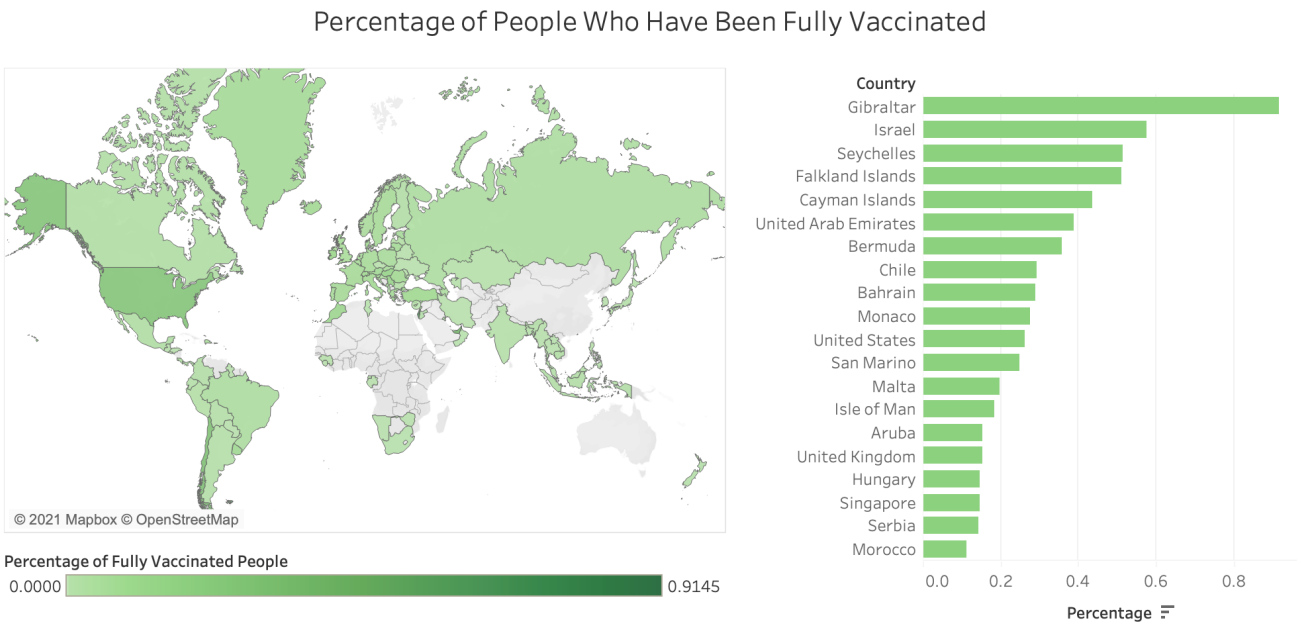

Figure 3.7: A color-coded world map representing the number of people who have been fully vaccinated from December 2020 to April 2021

Similar to the previous one, there are only now four countries that have covered half of their population for fully vaccinated people. This number decreases from nine to four and clearly shows that we're still at the beginning, which makes sense since the vaccinations only available from December 2020. As stated before and as seen in the figure 3.8, the decrease in number also shows that vaccinated people don't mean they received a complete immunization.

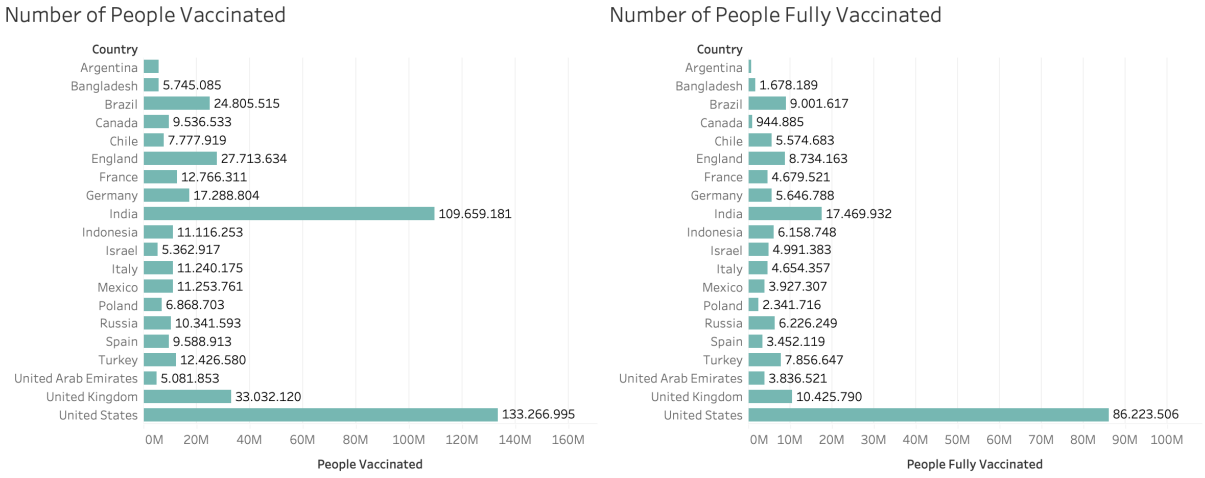

Figure 3.8: A graph representing the number of people who have been vaccinated and fully vaccinated from December 2020 to April 2021

3.4 Total Vaccinations Worldwide

Previously we know that one person can take more than one dose of vaccine depending on the type. This suggests that there are already a lot of vaccinations that occurred around the world.

A world heat map is used for this data — mapping out the total number of total vaccinations or total doses of vaccines that have been given in each country worldwide, color-coded so that the darker color shows a higher number. Figure 3.9 shows the frequency of total vaccinations in each country. Several countries that don't have any data on vaccinated people actually have data regarding the total vaccinations. This means that these two data are not related; the total vaccinations data doesn't represent the number of those who have been vaccinated.

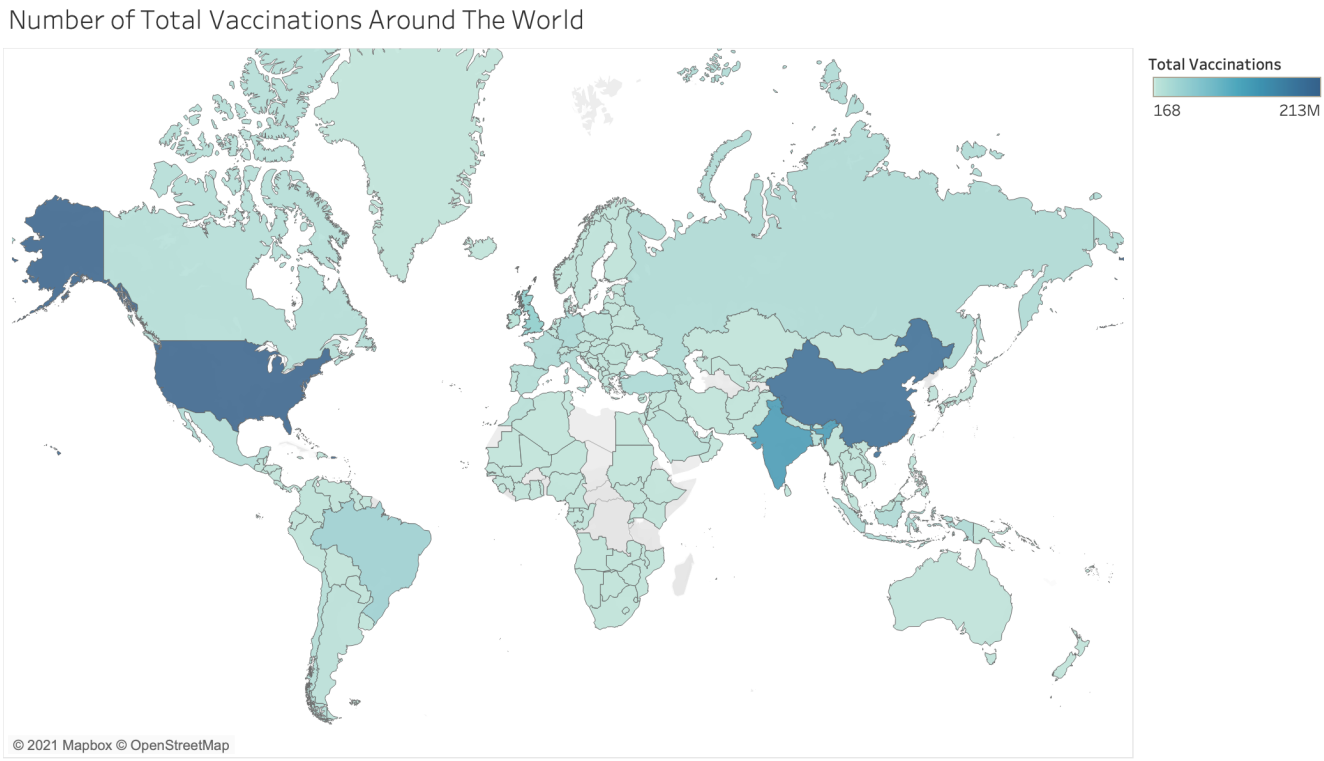

Figure 3.9: A color-coded world map representing the total vaccinations from December 2020 to April 2021

As shown, the most apparent countries with the highest frequencies are the United States, China, and India, with the frequency of 213,3 million, 195 million, and 127,1 million, respectively. The fact that China becomes one of the countries makes sense since the disease originated from that country [2]. For the United States and India, who made it to that list also makes sense for different reasons.

Those two countries have the most positive cases compared to others, so no wonder the doses of vaccines that have been given are a lot [5].

3.5 Trends and Forecast on Daily Total Vaccinations, Vaccinated and Fully Vaccinated People

To show the trend for the total vaccinations that have been conducted, people who have been both vaccinated and fully vaccinated, three scatter plots are used. Figure 3.10 shows the trend on a daily basis on a normalized value since the number on all of these data is not evenly scaled. As expected, as more days passed, the number also increases as well. Compared to others, the increase in fully vaccinated people is much more significant. This shows that more people are receiving complete immunization throughout days. There are also several outliers that could indicate that some data for some days may be missing.

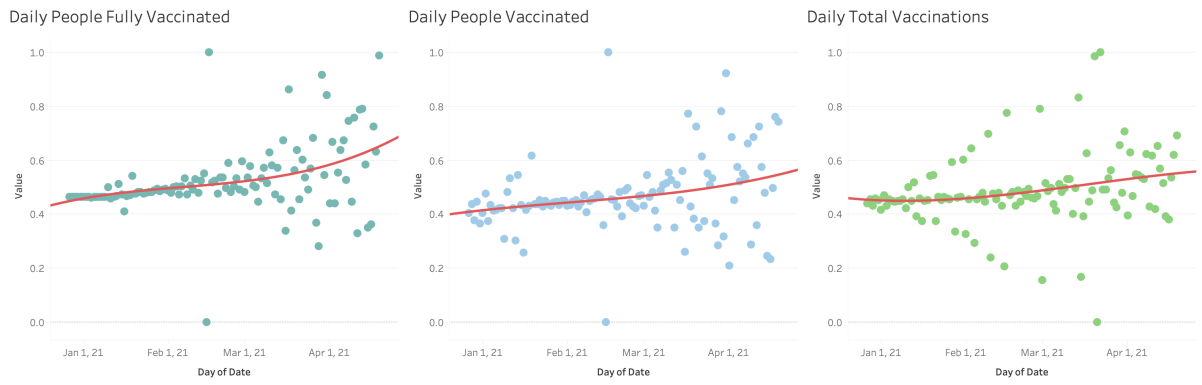

Figure 3.10: A scatter plots epresenting daily total vaccinations, vaccinated people, and fully vaccinated people from December 2020 to April 2021

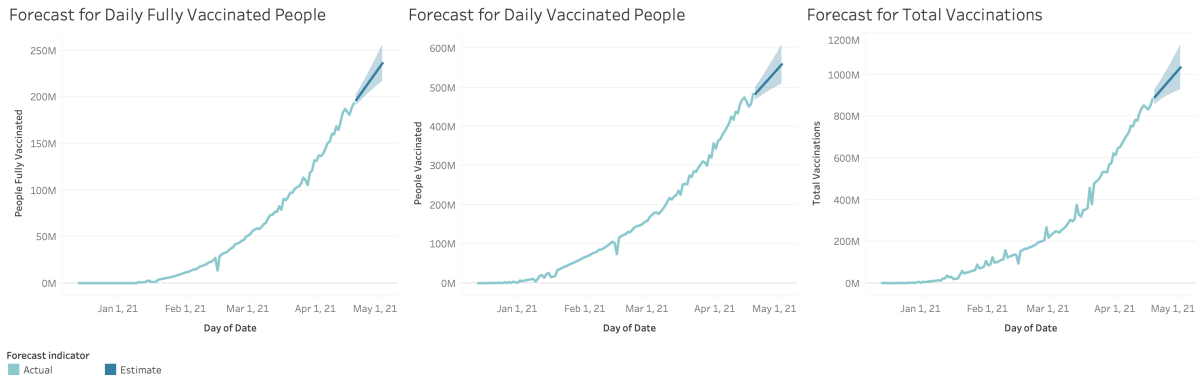

Figure 3.11: A graph representing the forecast daily total vaccinations, vaccinated people, and fully vaccinated people from December 2020 to May 2021

Using the actual value, we could also forecast for the next couple of days. As seen from figure 3.11 and as expected, since the number always increases, it will also increase for the number of total vaccinations and the number of people who have been vaccinated and fully vaccinated. With the increase of total vaccinations, this means that even more people will receive and immunize from the disease. As the 1st May 2021, the number of total vaccinations, vaccinated and fully vaccinated people are predicted to reach 1 billion, 560 million, and 236 million respectively

Chapter 4

Conclusions

Through my analysis of the progress of the COVID-19 vaccinations, this project has revealed the informative trends of the vaccination data. Using the Tableau software, the data was successfully visualized in color-coded world maps (world heat maps), bar charts, line graphs, and scatter plots. This helped identify the progress of the COVID-19 vaccinations and trends and forecasts the number of total vaccinations, people who have been both vaccinated and fully vaccinated. The visualizations covered several of the methods and techniques in the lectures.

These are the current progress, trends, and forecasts found in the visualizations:

- There are currently 11 types of vaccines that are widely used globally, where the majority are using Oxford/AstraZeneca and Pfizer/BioNTech
- These countries that use the Oxford/AstraZeneca and Pfizer/BioNTech are countries that manufacture those vaccines and their neighboring ones
- Countries that have a higher number in vaccinated or fully vaccinated people doesn't mean that these countries already covered a larger proportion of their population
- The United States and India have the most vaccinated people in the world, but only the United States has the most fully vaccinated people
- There's a decrease in the number between the fully vaccinated people data compared to the vaccinated people data that shows people are required to take more than one vaccinations
- The United States, China, and India have done a lot of vaccinations in their country, but it doesn't mean that they have progressed far enough
- The world is still in the beginning phase of the vaccinations, only nine countries where 50%

of their populations are vaccinated and only four countries where 50% of their populations are fully vaccinated

- Throughout days, the number of total vaccinations, vaccinated and fully vaccinated people increases
- Forecasted, the number of total vaccinations, vaccinated and fully vaccinated people will still increase

Chapter 5

Reflective Analysis

Based on what I did throughout this project, I believed that I deserved a 90 or above out of 100 (or since 20 points possible, 18 or above). Here are my justifications:

5.1 Reflections

Through this project, I learned how to apply the basic concepts that are taught in the class with other different tools besides MATLAB. Creating a visualization is easy, but creating the good one is a challenge. Sure using Tableau is easy; we can just drag and drop the fields, but then what? What are we trying to do here? What are we trying to analyze here? That's the challenge. Using the basic concepts combined with "then what" aspects helped me create a meaningful visualization for this project. This meaningful visualization helped me to gain insights, trends, and even forecasts from the dataset.

A world-heat map is my first choice of visualization to make it more meaningful for readers. A bar graph can also be used for the entirety of this project, but we would like to analyze the data in different countries for the most part. We already know that there is not just one country globally; there are 200 more countries. Hence, using a bar graph would not make sense since it will make it harder to see. But several visualizations actually needed a bar graph to clarify ambiguity like the percentage visualization on the figure a and b. At first glance, there are no differences between countries, which means that each country has the same percentage. It turns out that small countries that are hardly visible have a larger percentage, where they supposedly should have a darker color. Thus, having a bar graph helps with this issue.

Scatter plots also my other choice of visualization to highlight trends and to forecasts. Since I have learned about regression both inside and outside this course, using a scatter plot is the most suitable one combined with a line.

5.2 Design Aspects

Design-wise, using a brighter color scheme is my choice. I felt like if I use colors such as black or red (other than the trendlines), it seems that the visualizations have a hidden 'danger' message since these two colors are usually represented for extreme, danger, anger, etc. For example, if I use a red-colored world map, I'm afraid that the reader might think they're in "danger" etc. A Brighter-color scheme can help enlighten the readers. Also, using an evenly scaled image of visualization can help the readers to understand the analysis fully.

5.3 Limitations

The downside of using Tableau as a visualization and data analysis tool is that, up to April 2021, Tableau doesn't support more than two dimensions of visualizations. This is not a limitation, but it will be better to visualize in a 3D manner. For example, when creating a heat map, we could transform it into a 3D graph. Using this 3D graph, we could also forecast and highlight the trends.

Chapter 6

Appendices

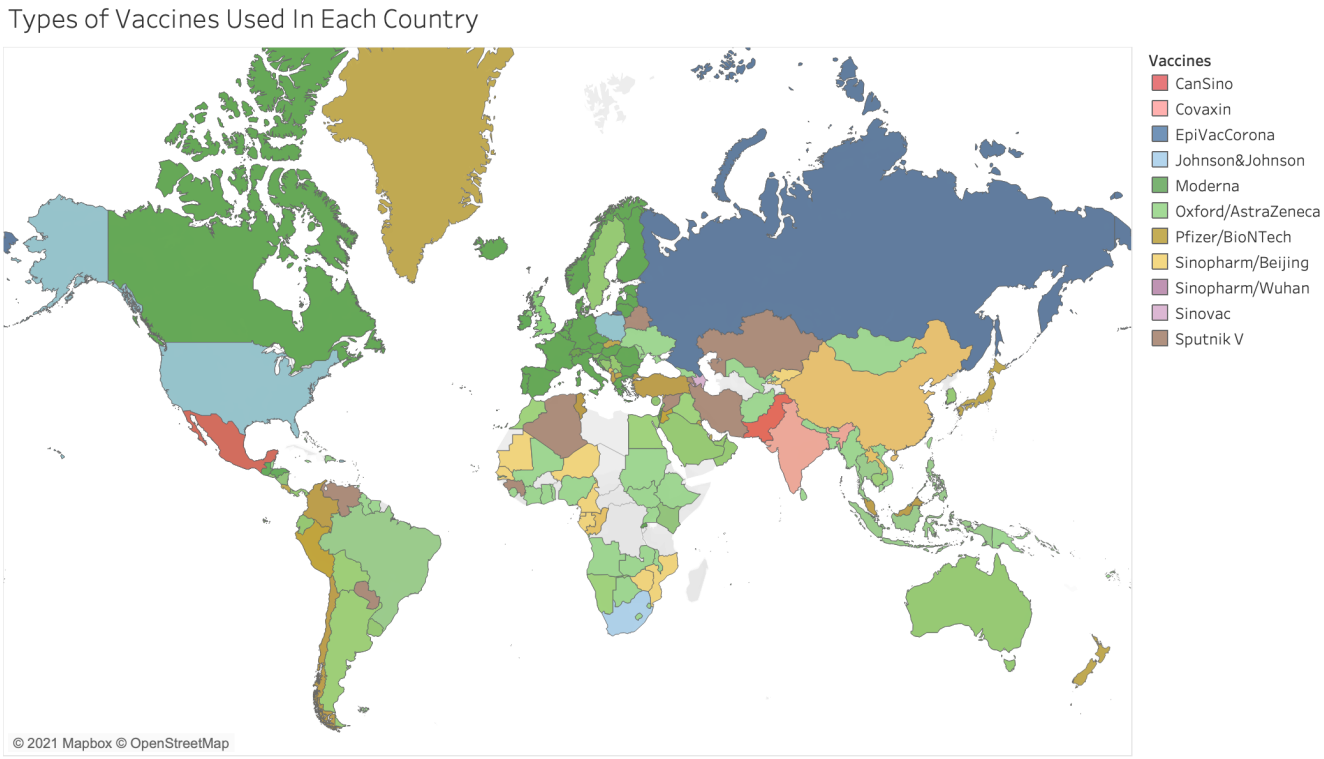

Figure 6.1: A color-coded world map representing types of vaccines used in each country

References

[1] Erin Schumaker. *Timeline: How coronavirus got started*. URL: https://abcnews.go.com/Health/timeline-coronavirus-started/story?id=69435165.

[2] *Coronavirus (COVID-19) Pandemic*. URL: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov.

[3] Will Brothers. *A Timeline of COVID-19 Vaccine Development*. URL: https://www.biospace.com/article/a-timeline-of-covid-19-vaccine-development/.

[4] Centers for Disease Control and Prevention. *Frequently Asked Questions about COVID-19 Vaccination*. URL: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/faq.html.

[5] *Countries Where COVID-19 Has Spread*. URL: https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/.