**Visualisation Project Plan**

Student Name: Wangjun Shen

Student ID: 110248810

文本

低可信度描述已自动生成

[Introduction 3](#_Toc134639028)

[Data Sources 3](#_Toc134639029)

[Project Plan 3](#_Toc134639030)

[Data Preparation 4](#_Toc134639031)

[Data Exploration 5](#_Toc134639032)

[References 7](#_Toc134639033)

## 

## Introduction

The proposed visualization aims to provide insights into the progress of COVID-19 vaccination worldwide. It helps answer the main question of how vaccination programs are progressing in different countries and how they compare to each other. The visualization can be used to understand which countries are using which vaccines, which countries have advanced vaccination programs, and which countries are vaccinating more people per day in terms of both absolute numbers and percentages of the population. This information is useful in monitoring global vaccination efforts and identifying countries that need more support to increase their vaccination rates. The visualization can also be combined with other COVID-19 related datasets to gain a more comprehensive understanding of the pandemic's dynamics.

## Data Sources

To develop a comprehensive understanding of the project, we will gather data from reliable sources such as the World Health Organization Coronavirus Dashboard, Our World in Data, official reports, the United Nations, World Bank, Global Burden of Disease, and Blavatnik School of Government. It's crucial to collect data from multiple sources to ensure its accuracy and reliability. Fortunately, Our World in Data has already collected and compiled data from these sources and has been maintaining it on GitHub for a while (Owid/Covid-19-Data, n.d.) . Hence, we'll use this dataset directly for our visualization exploration.

The proposed visualization project's dataset is quite large, with a size of 78.5mb and containing 67 variables. We've thoroughly analyzed the dataset and identified the variables that are most relevant to gain insights into the worldwide progress of COVID-19 vaccination. Based on this analysis, the following variables have been selected as useful: iso code, continent, location, date, total vaccinations, people vaccinated, people fully vaccinated, total vaccinations per hundred, people vaccinated per hundred, people fully vaccinated per hundred, new vaccinations, and new vaccinations smoothed per million. These variables are all related to COVID-19 vaccination, including the total number of vaccinations, people vaccinated, and people fully vaccinated, as well as the number of new vaccinations per day and per million people. The data types for these variables include numeric, date, and categorical. A processed sub-dataset that can be used as a reference comes from kaggle, but it stopped updating a year ago (Preda, 2021).

## Project Plan

The chosen dataset can be used to answer elementary, intermediate, and overall level questions related to COVID-19 vaccination progress worldwide. Elementary level questions include the number of people vaccinated worldwide, the most common vaccines used globally, and daily new vaccination trends worldwide, while intermediate level questions include the vaccination rates by country and continent, the relationship between vaccination rates and COVID-19 cases, and vaccination rate differences between countries with different income levels. Overall level questions address the effectiveness of COVID-19 vaccines, the impact of vaccination programs on the pandemic's trajectory, the likelihood of new COVID-19 variants emerging, and the ethical implications of unequal access to COVID-19 vaccines globally.

To provide a comprehensive overview of COVID-19 vaccination progress, a visualization project will be created using a subset of the original dataset that focuses on the most relevant variables related to vaccination. This visualization will enable users to compare vaccination rates and trends across countries and continents, identify countries that need more support in increasing their vaccination rates, and understand the impact of vaccination programs on the pandemic's dynamics.

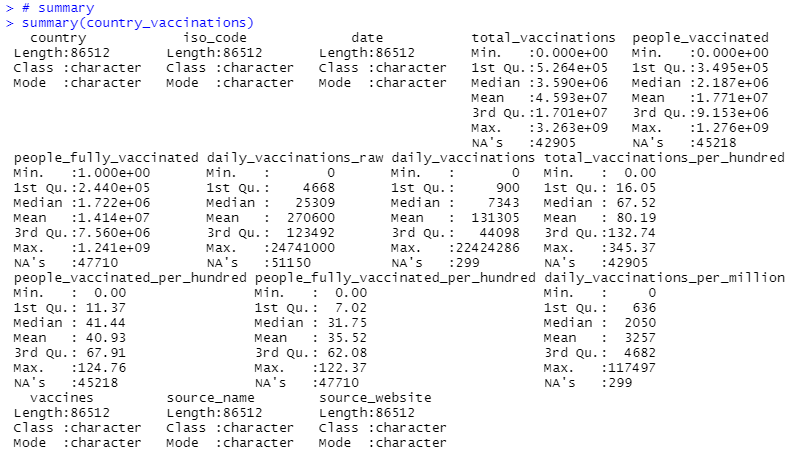
Potential challenges in the project include data quality issues, such as missing or inaccurate data, that may impact the visualization's accuracy and reliability. The size of the dataset may also pose challenges in processing and analyzing the data efficiently. Additionally, creating an accessible visualization that can be understood by a diverse audience with varying levels of technical expertise and familiarity with COVID-19 and vaccines is another potential challenge.

To address these challenges, it is essential to ensure data quality by cleaning and validating data, imputing missing data, and cross-referencing with other sources. Documenting data sources and processing steps can also help. To handle large datasets, use techniques like data aggregation, sampling, and data reduction. Specialized software like R or Python can help manage and analyze the data. To create an accessible visualization, use clear language, provide context, and avoid jargon. Interactive features like filtering can engage users with varying technical expertise. Testing and feedback can identify areas for improvement.

## Data Preparation

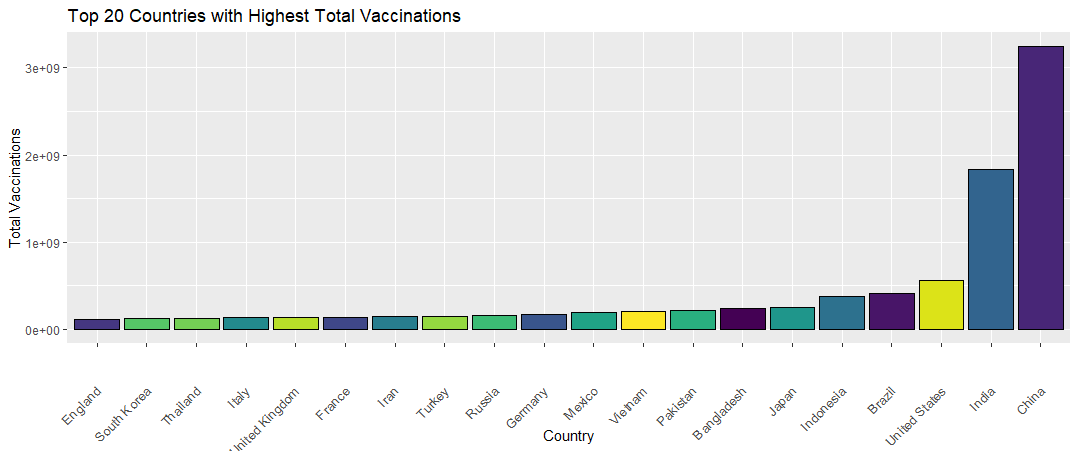
Before creating data visualizations for the proposed project, several data preparation tasks need to be completed on the selected dataset. These tasks include checking for missing or incomplete data and deciding how to handle them, removing unnecessary columns, filtering the dataset to include only relevant variables, sorting the dataset by date and location, and ensuring the data types are correct and consistent. No file joining is necessary since the selected dataset contains all necessary variables for the proposed visualization.

## Data Exploration



**Figure 1. Descriptive Statistics for Variables**

The summary statistics of the "country\_vaccinations" dataset offer insightful information regarding the worldwide vaccination efforts. The dataset includes 86,512 observations and features variables such as country, date, total vaccinations, people vaccinated, people fully vaccinated, daily vaccinations, and vaccine types. The average total vaccinations and people vaccinated are approximately 45.93 million and 17.71 million, respectively. However, the median total vaccinations and people vaccinated are significantly lower at 3.59 million and 2.19 million. The dataset also reveals that the highest number of people fully vaccinated is 1.24 billion, and the highest daily vaccination rate is 22.42 million. Additionally, the data points out the presence of many missing values in the variables related to daily vaccinations.

**Figure 2: Top 20 Countries with Highest Total Vaccinations**

The Figure 2 suggests that some countries have made more progress in terms of total vaccinations than others. The top 20 countries in this plot have all administered at least 50 million doses of the COVID-19 vaccine, while the bottom countries have administered less than 30 million doses. This highlights the significant variation in the global distribution of COVID-19 vaccinations.

图形用户界面, 图表

描述已自动生成

**Figure 3: Vaccination status of 5 representative or influential countries**

The five selected countries are diverse and important players in the global COVID-19 outbreak and vaccination efforts. The visualization depicts their respective COVID-19 vaccination rates. The United States leads with the fastest vaccination rate, which started late in 2020 and accelerated in early 2021. China has been slow in vaccinating, starting mass vaccinations only in early 2021. India's vaccinations began at the same time, but due to the outbreak, the pace has increased rapidly. Brazil and Russia's vaccination rates have been slow, but they have picked up speed since early 2021. The speed of vaccination could be influenced by various factors, such as political, economic, and social factors, which will require further analysis.

Below is a display of the 10 most commonly used vaccines:

图表

描述已自动生成

**Figure 4: Top 10 COVID – 19 Vaccines Used**

The chart provides a clear overview of the usage of different vaccines, highlighting which ones are the most commonly used. It can also help answer some questions about the usage of COVID-19 vaccines, such as which vaccines are used most frequently and which are used less frequently. Based on the chart, it is evident that Pfizer/BioNTech has the highest usage rate among the vaccines, followed by AstraZeneca, Moderna, Sinovac, and Sputnik V. "Other" vaccines are used to a lesser extent and account for the smallest proportion. This data could provide valuable insights for policymakers and public health officials to monitor the worldwide distribution of COVID-19 vaccines and determine areas where vaccine availability may be restricted.

图表, 散点图

描述已自动生成

**Figure 5: Scatter Plot for Total Vaccinations and Daily Vaccinations by Country**

Scatter plots demonstrate a positive correlation between the total number of vaccinations and the daily number of vaccinations. As more people are vaccinated, there is also a corresponding increase in the number of vaccinations administered per day. However, there appears to be a saturation point where an increase in the total number of vaccinations does not result in a proportional increase in the daily number of vaccinations. This suggests that countries may encounter obstacles in scaling up vaccination efforts beyond a certain threshold, which could affect their ability to attain herd immunity.

## References

owid/covid-19-data. (n.d.). GitHub. <https://github.com/owid/covid-19-data/tree/master/public/data>

Preda, G. (2021). COVID-19 World Vaccination Progress. Retrieved from <https://www.kaggle.com/gpreda/covid-world-vaccination-progress>