# Visualization Project

## COVID-19 Vaccine Progress

1: Dashboard 1: Which Vaccines Are Used by Which Countries?

https://public.tableau.com/app/profile/kota.kobayashi7296/viz/Assignment_16803219176360/Dashboard1

2: Dashboard 2: What is the trend of infections since the global vaccination campaign began?

https://public.tableau.com/app/profile/kota.kobayashi7296/viz/Assignment_16803219176360/Dashboard2

# Contents

# 1. Introduction

This report outlines the design process for effective visualization of global COVID-19 vaccination progress and infection trends. Utilizing Munzner's (2014) 4 levels of validation, the report presents an iterative, user-centered approach for creating comprehensive visualizations for public health officials and researchers. It covers domain situation, user tasks, data/task abstraction, visual encoding, interaction idioms, algorithm selection, and implementation, culminating in an evaluation of the visualization's effectiveness.

# 2. Domain Situation and User Task

## 2-1. Domain Situation:

The COVID-19 pandemic has affected every aspect of life across the globe. As countries implement vaccination campaigns to curb the spread of the virus, it is crucial to monitor the progress and the impact of these efforts on the infection trend. However, a problem lies in the lack of accessible and comprehensive visualizations to understand the distribution of vaccinations across countries and the correlation between vaccination efforts and infection trends. This absence of easily digestible information hampers effective decision-making for policymakers and the general public, leading to potential misinformation and misconceptions about the vaccines' effectiveness and progress.

## 2-2. Target Users:

The target users for this visualization are primarily public health officials and researchers involved in tracking and analyzing the progress of vaccination campaigns. According to a study by Betsch et al. (2018), effective communication of vaccination data to these professionals can enhance decision-making and promote evidence-based strategies. By narrowing down the target audience, we ensure that the visualizations are tailored to meet the specific needs of these experts, who play a crucial role in shaping public health policies and managing the pandemic response.

## 2-3. User Tasks:

<u>1. Understand which countries have adopted which vaccines:</u>

By providing clear visualizations of vaccine distribution among countries, this user task addresses the problem by offering a straightforward way to understand the vaccination landscape globally. This information empowers decision-makers and the public to stay informed and make data-driven decisions.

2. Analyze the impact of vaccination campaigns on infection trends:

This user task aims to illustrate the impact of vaccination campaigns on the infection trends worldwide. By demonstrating the correlation between vaccination efforts and changes in infection rates, it helps stakeholders assess the effectiveness of vaccination strategies and make informed choices regarding public health policies and personal decisions.

# 3. Data and Task Abstraction

## 3-1 Selection of dataset and variables:

We use three datasets from Our World in Data's GitHub: 'locations.csv', 'owid-covid-data.csv', and 'vaccinations.csv'. These datasets offer information on COVID-19 vaccination progress, location details, and pandemic statistics. To visualize vaccine distribution and the relationship between vaccination and infection trends, we'll focus on variables like location, vaccine types, total vaccinations, daily vaccinations, and COVID-19 case and death counts.

## 3-2 Pre-processing step:

Each pre-processing action serves a specific purpose to ensure a clean, organized, and optimized dataset for visualization. Here's a concise explanation of each step:

1. Importing necessary libraries (pandas, datetime) and reading CSV files: This initializes the environment and loads datasets for processing.

```
[18]: import pandas as pd
      import datetime as dt

[19]: df_vaccinations = pd.read_csv("vaccinations.csv")
      df_country = pd.read_csv("locations.csv")
      df_covvid19 = pd.read_csv("owid-covid-data.csv")
```

2. Merging 'df_vaccinations' and 'df_country' datasets based on 'location': This combines vaccination data with country information, creating a comprehensive dataset for visualization.

```
[20]: df_vaccinations = pd.merge(df_vaccinations,df_country,on='location',how='left')
```

3. Dropping irrelevant columns and renaming for consistency: This simplifies the dataset by removing unnecessary information and standardizing column names, facilitating data manipulation

during visualization.

4. Removing last two rows of 'df_vaccinations': This eliminates potentially incomplete or erroneous data, ensuring data quality and accuracy.

5. Checking for missing values in 'iso_code': This identifies and addresses missing values in a critical column, guaranteeing data quality and consistency.

```
[21]: print(df_vaccinations.columns)

      Index(['location', 'iso_code_x', 'date', 'total_vaccinations',
             'people_vaccinated', 'people_fully_vaccinated', 'total_boosters',
             'daily_vaccinations_raw', 'daily_vaccinations',
             'total_vaccinations_per_hundred', 'people_vaccinated_per_hundred',
             'people_fully_vaccinated_per_hundred', 'total_boosters_per_hundred',
             'daily_vaccinations_per_million', 'daily_people_vaccinated',
             'daily_people_vaccinated_per_hundred', 'iso_code_y', 'vaccines',
             'last_observation_date', 'source_name', 'source_website'],
            dtype='object')

[22]: df_vaccinations = df_vaccinations.drop(['iso_code_y','last_observation_date','source_name','source_website'],axis=1)

[23]: df_vaccinations = df_vaccinations.rename(columns = {'iso_code_x':'iso_code'})

[24]: df_vaccinations = df_vaccinations.drop(df_vaccinations.tail(2).index)
```

6. Converting date strings to datetime objects and extracting year, month, and day: This organizes temporal data in a structured format, enabling easier analysis and visualization of trends over time.

```
[26]: yseries = df_vaccinations['date'].map(lambda str: dt.datetime.strptime(str, '%Y-%m-%d').strftime('%Y'))
      mseries = df_vaccinations['date'].map(lambda str: dt.datetime.strptime(str, '%Y-%m-%d').strftime('%m'))
      dseries = df_vaccinations['date'].map(lambda str: dt.datetime.strptime(str, '%Y-%m-%d').strftime('%d'))

[27]: year=[]
      month=[]
      day=[]

[28]: for i in yseries.tolist():
          year.append(int(i))
      for i in mseries.tolist():
          month.append(int(i))
      for i in dseries.tolist():
          day.append(int(i))
```

7. Creating a new DataFrame with cleaned data and exporting as 'vaccinations_new.csv': This consolidates the cleaned data into a new DataFrame, ready for visualization, and exports it as a separate file for clarity.

```
[29]: df = pd.DataFrame({'location':df_vaccinations['location'],'iso_code':df_vaccinations['iso_code'],'date':df_vaccinations['date'],'year' : year, 'month' : month,'day':day,
                         'total_vaccinations':df_vaccinations['total_vaccinations'],'people_vaccinated':df_vaccinations['people_vaccinated'],
                         'people_fully_vaccinated':df_vaccinations['people_fully_vaccinated'],'daily_vaccinations_raw':df_vaccinations['daily_vaccinations_raw'],
                         'daily_vaccinations':df_vaccinations['daily_vaccinations'],'total_vaccinations_per_hundred':df_vaccinations['total_vaccinations_per_hundred'],
                         'people_vaccinated_per_hundred':df_vaccinations['people_vaccinated_per_hundred'],'people_fully_vaccinated_per_hundred':df_vaccinations['people_fully_vaccinated_per_hundr
                        'daily_vaccinations_per_million':df_vaccinations['daily_vaccinations_per_million'],'vaccines':df_vaccinations['vaccines']})
```

```
[14]: df.to_csv("vaccinations_new.csv")
```

8. Exporting 'df_covvid19' DataFrame as 'covid19.csv': This exports the cleaned COVID-19 data into a new file for organized data storage and seamless integration into the visualization process.

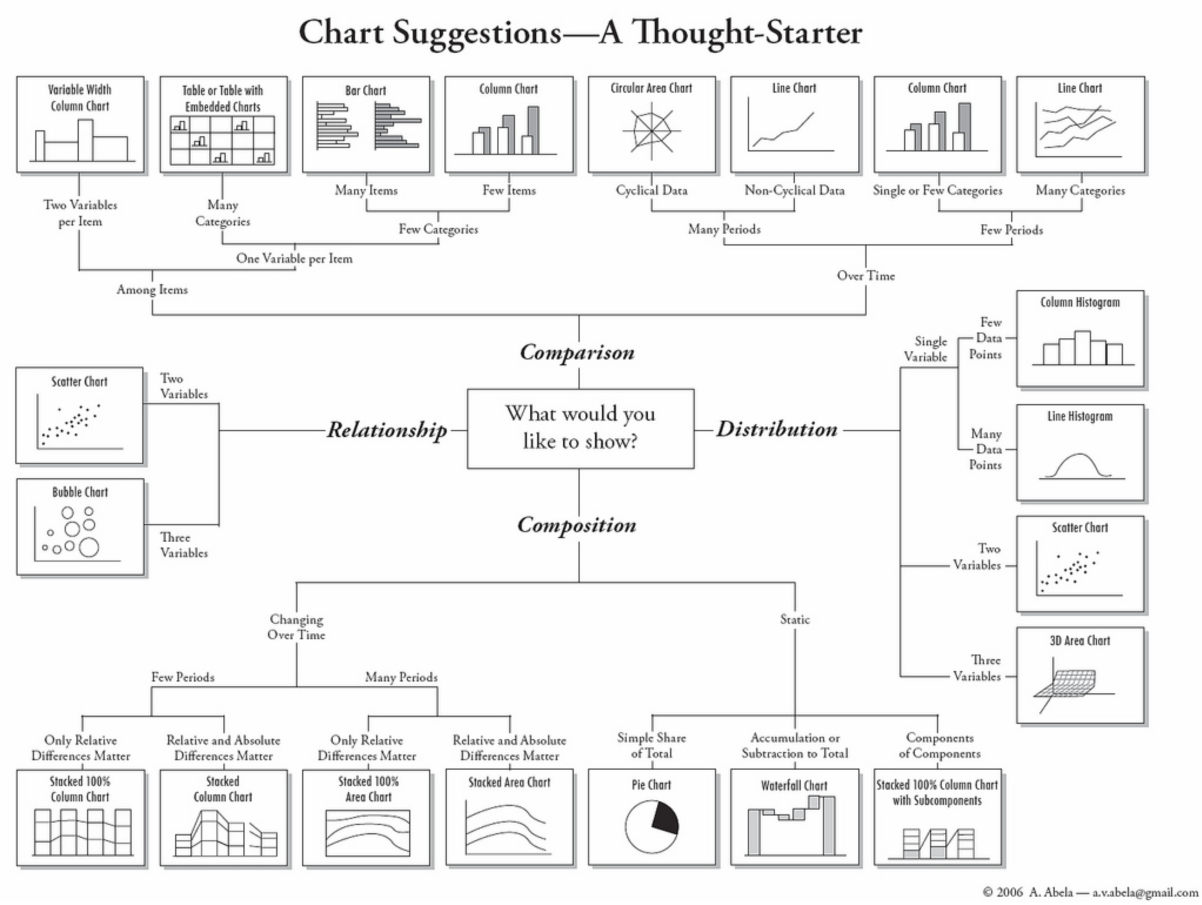
```
[15]: df_covvid19.to_csv("covid19.csv")
```

# 4. Visual Encoding and Interaction Idioms

## 4-1 Selection of type of chart:

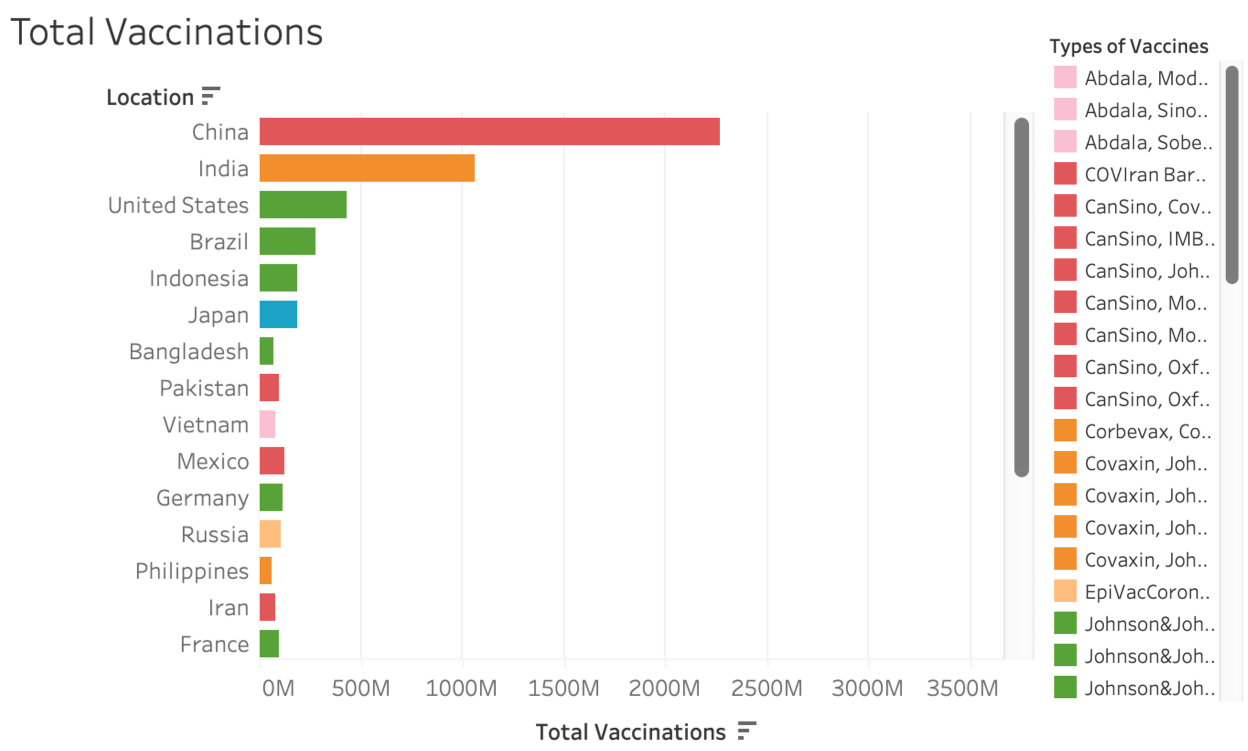
This section discusses the chart type selection process based on fundamental principles. We refer to this chart (Figure 1) to select appropriate chart types throughout this section (Adela, 2006).

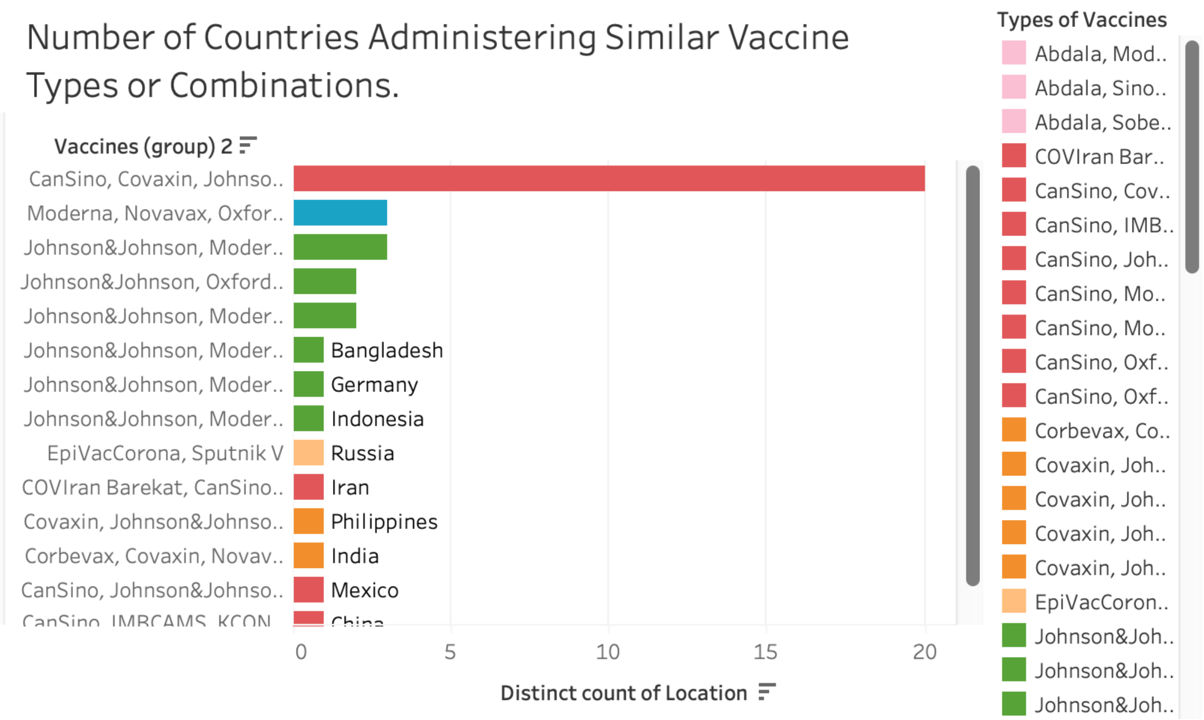
Figure 1



4-1-1 Bar Graphs in the First Dashboard

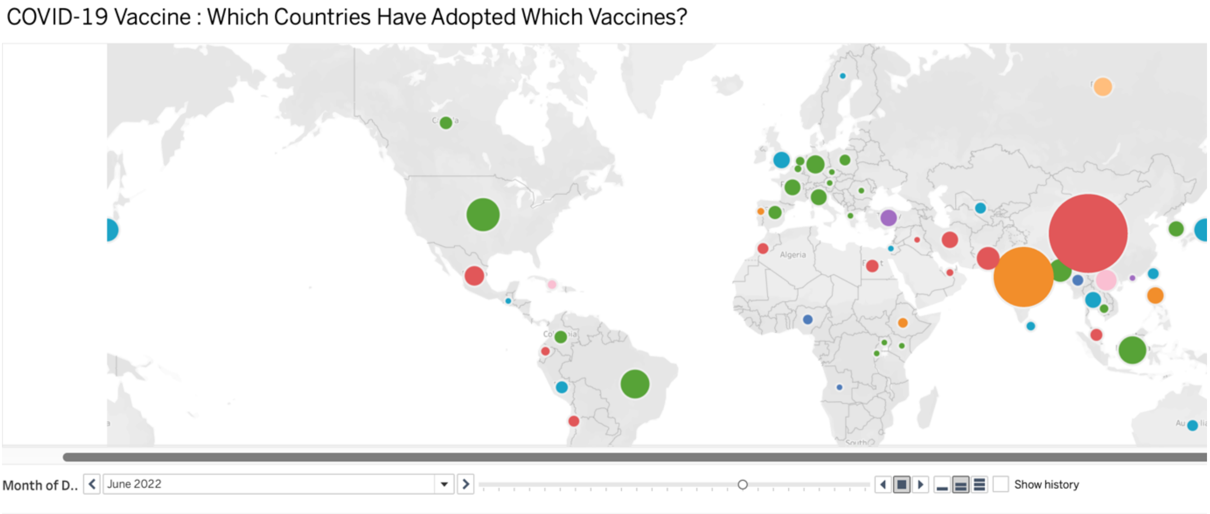
Figure 2




To easily compare total vaccinations and vaccine types across different countries, the two bar graphs included in the first dashboard are effective. This is comparing the overall number of vaccinations by country and the number of countries administering specific types or combinations of vaccines. This type of graph was chosen over pie charts, which are more suitable for displaying proportions, and line graphs, which are better suited for representing continuous data over time. As demonstrated in Figure 1, bar graphs are particularly well-suited for making comparisons among items that have one variable per item.

**Figure 3**




4-1-2. Geographical Bubble Map in the First Dashboard

In the first dashboard, a geographical bubble map is preferred over alternative visualization techniques, such as choropleth maps and proportional symbol maps, for representing countries' vaccine adoption and the number of vaccinated individuals, indicated by bubble size. Although choropleth maps effectively convey data variations within regions, they may inadequately represent point-based information like vaccination numbers. Proportional symbol maps, though functionally similar to bubble maps, might not provide equivalent visual clarity. Wilkinson (2012) underscores the significance of selecting suitable visualization techniques for geographical data. A bubble map enables users to promptly discern the geographical distribution of vaccine adoption and compare bubble sizes across countries. This visualization approach efficiently addresses the user's task of comprehending global vaccine adoption patterns in a lucid and succinct manner.

4-1-3. Line Graphs in the Second Dashboard

**Figure 4**

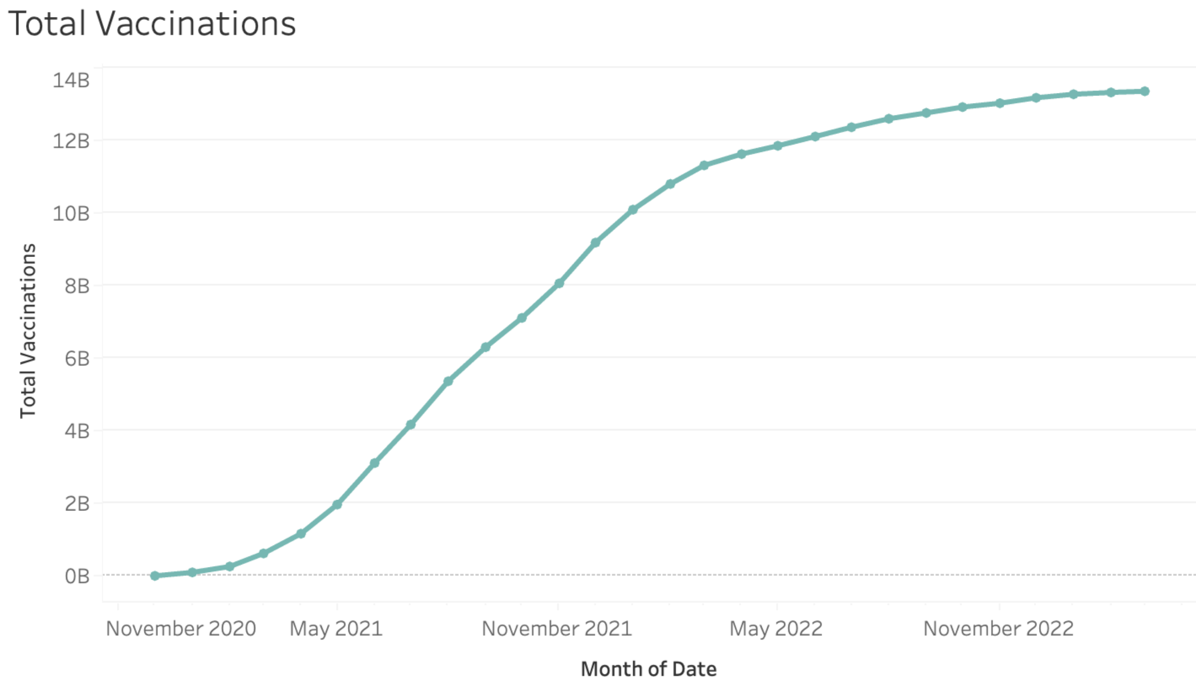
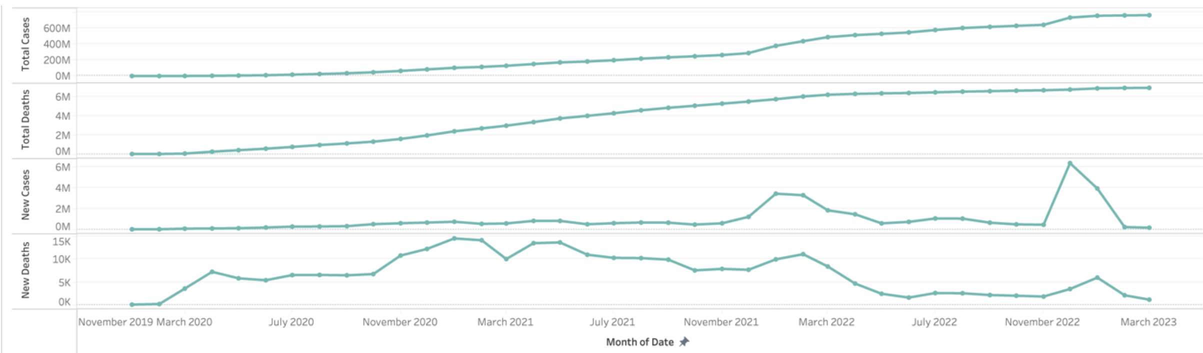



**Figure 5**

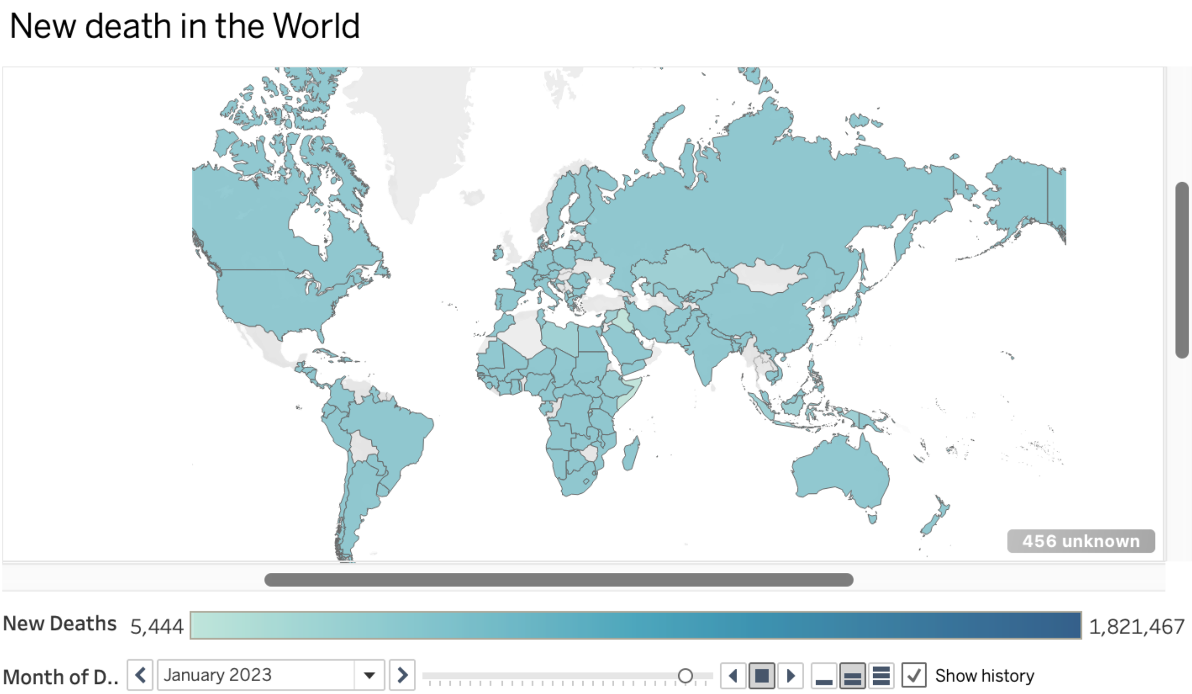



In the second dashboard, two line graphs are employed over alternatives such as bar graphs and stacked area charts. Figure 4's horizontal axis represents the timeline in monthly increments, while the vertical axis indicates the number of people vaccinated worldwide. Figure 5 illustrates trends for "total COVID cases," "total COVID deaths," "new cases," and "new deaths."

Line graphs are selected due to their efficacy in comparing temporal data, particularly with numerous periods and non-cyclical data. In comparison, bar graphs may appear cluttered, and stacked area charts might not clearly exhibit changes in individual categories. Line graphs adhere to Shneiderman's information-seeking mantra of overview, zoom, and filter, providing a data trends overview, enabling users to concentrate on specific timeframes (zoom), and comparing distinct categories (filter). This facilitates users in effortlessly visualizing and contrasting vaccination and infection rate trends over time, supporting their task of comprehending the evolving infection trend since vaccine distribution commenced.

### 4-1-4. Choropleth Map in the Second Dashboard

**Figure 6**

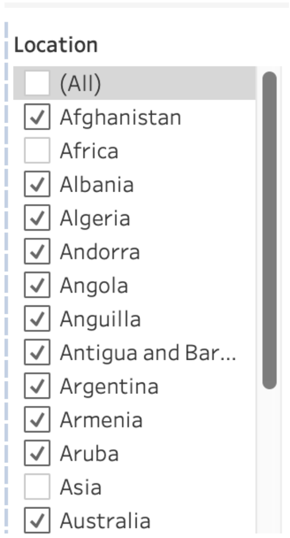



The second dashboard employs a choropleth map to display worldwide COVID-19-related new death distribution. Choropleth maps effectively represent data variations using color, while alternative visualizations may be less clear or distort geographic shapes. Alternative visualizations, such as symbol maps or cartograms, may not clearly convey the data's spatial distribution or could distort geographic shapes, making it harder to interpret. This map adheres to Tufte's (2001) principles of clarity, precision, and efficiency, allowing users to quickly assess mortality rates, compare regions, and identify areas requiring additional attention, promoting informed strategies to address the pandemic.

## 4-2. Interaction idioms

### 4-2-1. Location Interaction in the First Dashboard:

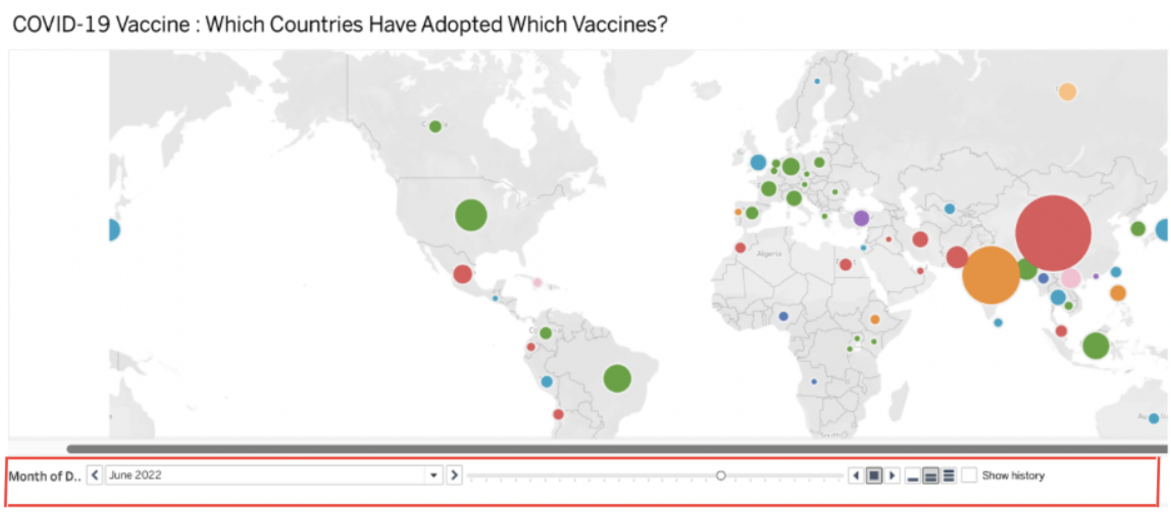
Figure 7




The first dashboard incorporates a location interaction feature, which allows users to select specific countries or regions they want to focus on. This interaction empowers users to narrow down the data based on their interests or needs, enabling them to analyze and compare vaccination efforts and vaccine types in different locations more effectively. By providing this location-based interaction, users can engage with the visualizations more actively, explore country-specific details, and gain a deeper understanding of the global vaccine landscape.

4-2-2. Time Series Animation in the First Dashboard:

Figure 8



The first dashboard also features a time series animation that displays the trend of vaccination data from 2020 to 2023. Users can play the animation to observe how the data has evolved over time or pause it to focus on a specific month. This interaction helps users identify patterns and trends in vaccination efforts, such as the acceleration or deceleration of vaccination campaigns, and the adoption of different vaccine types. By enabling users to interact with the data temporally, they can gain valuable insights into the progression of global vaccination efforts and their potential impact on public health.

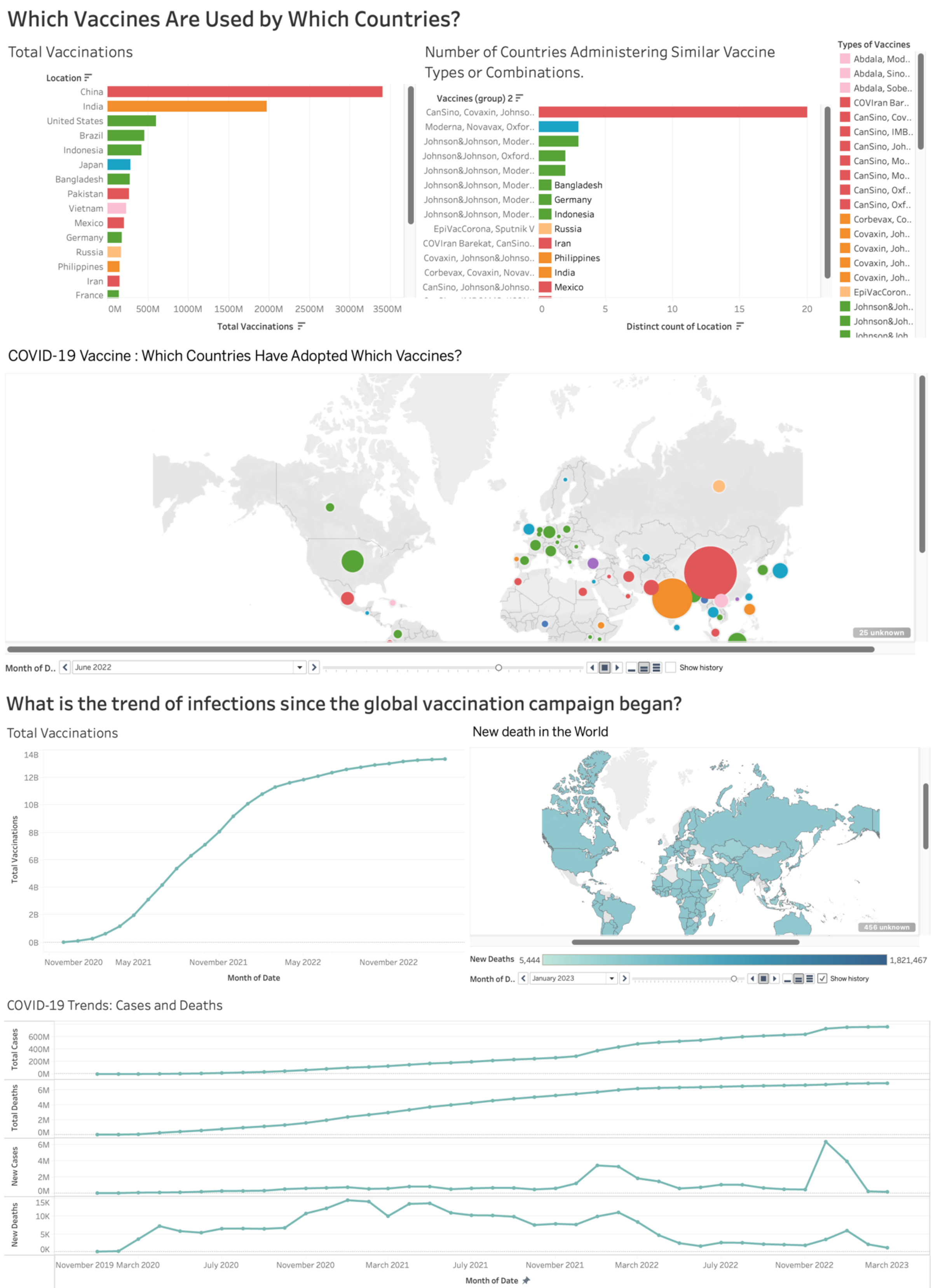
4-2-3. Interactive Filters in the Second Dashboard:

**Figure 9**


The second dashboard features a time series animation, illustrating the trend of new COVID-19 deaths worldwide from 2020 to 2023. Users can activate the animation to observe the change in mortality data or pause at specific points to concentrate on particular months. This interaction allows users to detect patterns and shifts in the number of new deaths, such as periods of increase or decrease, and the relationship with vaccination efforts. By providing users the ability to explore data over time, they can acquire crucial understanding of the progression of the pandemic's impact and the influence of global vaccination strategies on public health.

# 5. Algorithm selection and implementation

Figure 10 First and Second Dashboard



### 5-1. Why dashboard?

To resolve the user tasks, a dashboard is used (Figure 10) because using a dashboard to resolve the two user tasks enables efficient visualization and comparison of multiple data sets. It provides an interactive platform that consolidates relevant information, allowing users to explore, compare, and contrast vaccine adoption patterns and the impact of vaccination campaigns on infection trends simultaneously, leading to better-informed decision-making.

### 5-2. Why tableau?

Furthermore, Tableau is chosen for its ability to create dynamic, interactive visualizations, including time series animations. It allows users to explore temporal changes in vaccine adoption and infection trends easily, while its user-friendly interface and extensive features facilitate visually appealing dashboards.

## 6. Evaluation

This project's evaluation assesses the effectiveness of visualizations and interactions in addressing user tasks and meeting the target audience's needs. Clarity is achieved through appropriate chart types and interactive features, enhancing the interpretability of visualizations. Relevance is maintained by aligning visualizations with user tasks and selecting pertinent datasets and variables. Efficiency is attained by consolidating relevant information within an interactive dashboard, enabling users to explore and compare data effectively. User engagement is fostered through interactive features, promoting active exploration and in-depth analysis. Overall, the project successfully addresses user tasks, providing valuable support for public health officials and researchers in managing the COVID-19 pandemic.

## 7. Conclusion

In summary, this report demonstrates the development of a comprehensive and effective visualization that addresses the needs of public health officials and researchers in managing the COVID-19 pandemic. By employing Munzner's (2014) 4 levels of validation and adopting an iterative design process, the visualization facilitates data-driven decision-making and promotes a deeper understanding of global vaccination progress and its impact on infection trends. This user-centered approach ultimately contributes to informed public health policies and strategies in tackling the ongoing pandemic.

## 8. References


Adela (2006). Chart Suggestions. Retrieved from https://img.ctrlq.org/di/choosing_a_good_chart2.pdf

Betsch, C., Böhm, R., & Korn, L. (2018). Inviting free-riders or appealing to prosocial behavior? Game-theoretical reflections on communicating herd immunity in vaccine advocacy. Health Psychology, 37(9), 817-826. https://doi.org/10.1037/hea0000642

Munzner, T. (2014). Visualization analysis and design. CRC Press.

OWID. (2023). COVID-19 Data Repository by the Our World in Data team. Retrieved March 26, 2023, from https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In Proceedings of the 1996 IEEE Symposium on Visual Languages (pp. 336-343). IEEE. https://doi.org/10.1109/VL.1996.545307

Tufte, E. R. (2001). The visual display of quantitative information. Graphics Press.

Wilkinson, L. (2012). The grammar of graphics. Springer Science & Business Media.