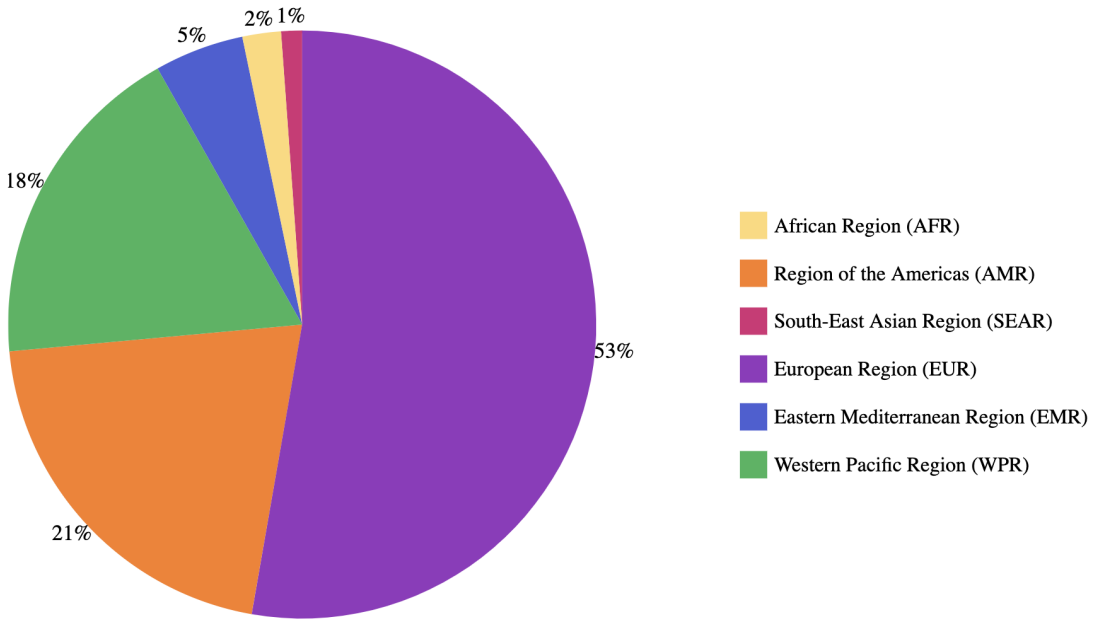
Data Driven Web Application Project 1

Design Rationale

Legends

We looked for a color palette for the WHO regions that had consistent luminosity and did our best to prevent any possible issues with color blindness. The WHO region abbreviations are not particularly well-recognized and their real names are rather long, so we decided to include the full names for the first legend and use only the abbreviations in the rest.

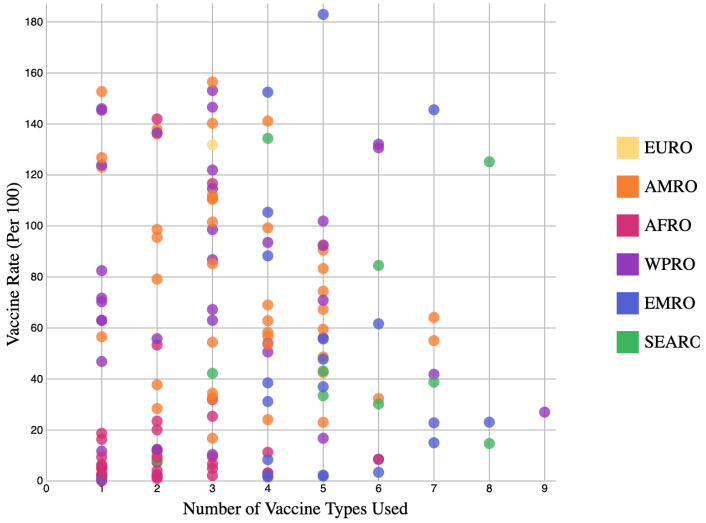
Percentage of Cumulative Vaccine Doses Administered for Each WHO Region by August 2021



<u>Design Rationale for First Visualization (Pie Chart):</u>

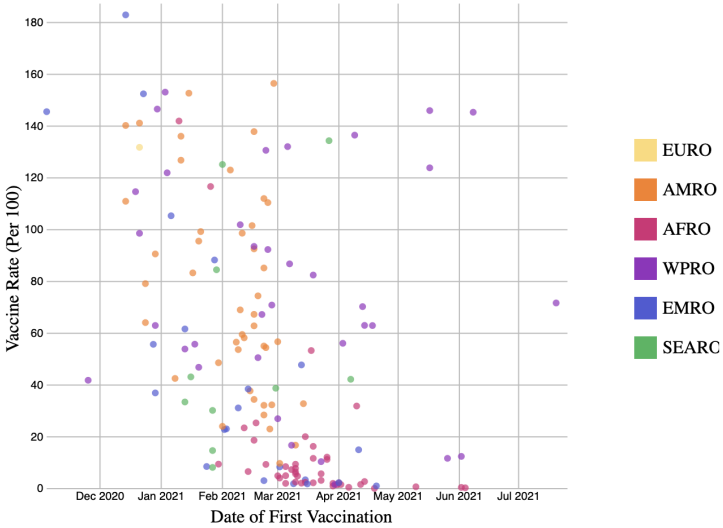
With a purpose to present the portion of each WHO region's cumulative vaccine administration looks like, a pie chart is used to illustrate the numeric proportion of each. The color hue is used as the channel of this visualization. A legend is created to indicate each WHO region's color on the pie chart. The pie chart is divided into slices based on each WHO region's total vaccine administration number per the total vaccine administration number of all WHO regions. The text of percentage of each WHO region is added around each arc, indicating each portion's percentage.

Vaccine Rate Compared to Number of Types of Vaccines Used per WHO Region



Design Rationale for Second Visualization (Scatter plot):
Using a scatter plot was the most optimal way to map the vaccination rate per 100 people, number of vaccine types used, and WHO region in the same chart. The goal was to determine if there was a relationship between the number of vaccines used and the vaccination rate. The marks are the circles, colored by hues based on the WHO region of the data point. The x axis (vaccination rate) and y axis (number of vaccine types) are scaled linearly since the data is highly scattered throughout the axis' extents. While starting the "Vaccine Rate" at 0 is intuitive, the result is a mark overlapping the "1" on the axis tick. Additionally, although there is one large clump of data points to the bottom left of the graph, users are able to see that they largely belong to the same WHO region. Nevertheless, to see overlap a bit easier, the marks are slightly transparent.

Vaccine Rate Compared to Date of First Vaccination per WHO Region



Design Rationale for Third Visualization (Scatter plot):
Using a scatter plot was the best way to map the vaccination rate per 100 people, the date of first vaccination, and WHO region in the same chart. The goal of the chart was to see if there was a relationship between the date of first vaccination and the most recently updated vaccination rat. The x- (date of first vaccination) and y-axes (vaccination rate per 100 people) are scaled linearly since the data are distributed fairly evenly through both axes extents, as well as the fact that representing time with anything but linear at this scale of time would be hard to understand. The x-axis ticks were put in the "month year" format to give a better sense of the time scale. The circles were given slightly reduced opacity to make the denser area near (Mar-Apr 2021, 0) a little easier to see that there were many points in the area, as opposed to only a few. The color scale for the WHO regions was used to maintain some consistency across the three charts in an attempt to show trends in individual WHO regions.
The marks used in this chart were the circles, each representing a specific country, and the channels used were hue and aligned vertical and horizontal positions, as well as opacity to the extent that it helped show the density in crowded parts of the chart.

The Dataset

The data we choose comes from a Kaggle dataset (https://www.kaggle.com/umeshkumar017/vaccination-data) that records the WHO-provided vaccination data summaries up to Aug 31st, 2021. The original dataset contains 227 distinct countries and corresponded 13 column variables, such as WHO regions, total vaccine doses, fully vaccinated count, .etc. In this project, we choose to specifically look at possible correlations between vaccination rates, number of vaccine types, first vaccination dates, and WHO regions. Therefore, we only keep 8 variables from the original data: country name, ISO3, WHO region, total vaccine doses, total vaccine doses per 100, vaccines used, number of vaccines used by type, and first vaccination date.

After the variable selection, we did data cleaning by dropping all data rows that contain any of these key variables null or empty. This gives us 156 countries left with no bad values and we use this for all later visualizations in this project. From the cleaned data, we also built a subset grouped by WHO regions and summed up vaccine doses in each. This pre-processing enables us to efficiently plot the percentage pie chart later.

The Story and Insight

COVID-19 vaccines have shown the effectiveness at helping protect against severe disease and death caused by COVID-19 and its variants. Overall, our visualizations try to examine the status of each WHO region's vaccine dose administration.

From the pie chart visualization, the data shows that the Western Pacific Region (WPRO) has the highest number of cumulative vaccine administrations, taking more than half of the total

proportion. On the other hand, the African Region (AFRO) only takes 1 percentage of the total. What is really surprising is that WPR and AFR regions have a similarly large population but the difference in the number of vaccine administration is tremendous. This indicates the disparity or inequality of vaccine resources distribution over low-income countries.

The second visualization shows a relationship between the number of vaccine types (Pfitzer, Moderna, etc) per country and the vaccine rate of cumulative vaccine doses administered per 100 population. The visualization shows that countries in the African Region (AFRO) concentrated in the number of vaccine types of 1, 2, and 3 with a really low vaccine rate. Surprisingly, many countries have a larger number of vaccine types but still have a lower vaccine administration rate.

The last visualization indicates a relationship between the start or launch date of the first vaccine administered in a country and the overall vaccine administration rate of that country. We can see a general pattern that the country who has a late vaccine administration launch date has a lower vaccine administration rate. Specifically, the countries with a late administration date and low vaccine rate are concentrated in the African Region (AFRO) and Eastern Mediterranean Region (EMRO). This pattern might be caused by the fact that some underdeveloped and developing countries usually have access to vaccine administration later than developed countries and they also have fewer vaccine resources than developed countries.

In conclusion, the visualizations want to convey a general pattern of each WHO region's vaccine administration progress. WHO regions in which more upper-middle income countries reside normally have a higher vaccine rate. Moreover, the discrepancy of vaccine resources is extremely large between lower-income countries and upper-income countries.

Contributions

The whole group collaborated on selecting the data set, coming up with the ideas for the different charts, and completing the final report, spending about 3 hours in total on these parts. Quinn (kw463) wrote the code for the first pie chart visualization, spending about 3 hours. Jackie (jmw499) wrote the code for the second scatter plot visualization, spending about 2 hours. Caroline (chz6) wrote the code for the third scatter plot visualization, spending about 2 hours. Nina (xz348) did the data cleaning and contributed to the color legends build up, spending about 3 hours.