

Data Visualization to better understand Covid Pandemic

Group 7

Carolina Sá, m20210693

Diogo Nobre, m20210677

Mariana Romão, m20210688

Mário Negas, m20210596

Introduction:

In this project, we worked with data regarding covid and covid vaccinations. The main goal is to study how covid affects the global population gaining an overall perspective of the countries and their cases, deaths and total vaccines.

On the dash it will be possible to interact with different types of variables and visualizations. At the end of the dash the user will be able to see the top countries concerning the three main variables: cases of covid, deaths and vaccines. Although vaccinations only began in December of 2020 it was thought best to incorporate this information too.

In order to get the data, we searched in kaggle and discovered these two datasets:

<https://www.kaggle.com/datasets/ognevdenis/covid-19-datasets>

<https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>

Dataset Description:

It was downloaded two different datasets, the first one containing information about covid between March 2020 and March 2022 worldwide and the second one regarding covid vaccinations and their type.

A brief description of the variables used on both datasets provided in the following tables correspondingly (Table 1 - Covid DataSet, Table 2 - Vaccine DataSet):

Table 1 - Covid DataSet

VARIABLE	TYPE	DESCRIPTION
country	object	A country is a distinct territorial body
day	object	YYYY-mm-dd
cases_new	object	The difference in relation to the previous record of all cases
cases1Mpop	object	The number of cases per million people
cases_total	int64	Records of all cases
deaths_new	object	The difference in relation to the previous record of all cases
deaths1Mpop	object	The number of cases per million people
deaths_total	float64	Records of all cases

Table 2 - Vaccine DataSet

VARIABLE	TYPE	DESCRIPTION
country	object	this is the country for which the vaccination information is provided;
date	object	date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
total_vaccinations	float64	this is the absolute number of total immunizations in the country;
total_vaccinations_per_hundred	float64	ratio (in percent) between vaccination number and total population up to the date in the country;

Inspiration:

During the pandemic the whole world was overwhelmed by the amount of data provided every day, so in order to auxiliate the understanding of the information provided it was thought that more and different visualizations were necessary in order to extract more knowledge.

We got inspired by searching through pages with visualizations of covid and so began to develop creative visualizations with the purpose of facilitating the comprehension of the dimension of all the data and the repercussions of the pandemic.

Although the vaccination only began in December of 2020, we are able to see the top countries as of the last recorded data. The top countries with regard to cases, deaths and vaccinations are an important visualization so the user can have a global understanding of the countries' situations.

Layout:

It was decided to divide the dash in three main parts regarding the amount of cases, deaths and vaccines.

Initially the user can choose if he wants to see the number of cases or deaths, after this decision the user has the option to focus on the new, the total or the number by 1M of habitants of the country. This decision can be made using 5 different buttons on the top of the screen. The data selected will be presented in a map where it is possible to select the month of the data, chosen by the user through a slider, as well as the continent.

It will be possible as well to see line plots that will illustrate the cases across time. The user shall choose the countries and then compare them individually or in combined plots. Combining the plots translates to the union of the y axes, it is interesting to see because with this alteration it is easy to compare the countries and their main statistics. These two options of visualization permit a real comparison of the data as well as the comparison of the seasonality, trends, and slope. For example, it is possible to see, as most of the population gets vaccinated, the restrictions got lifted, people became more carefree and so it had an impact on these main two variables, cases, and deaths, increasing the number of cases, but decreasing the number of deaths. There is a clear trend, increasing the number of cases in January, as a consequence of the holidays, as well as during the summer, a small increase.

Concerning the dataset, the user will be able to see the top ten of countries concerning cases, deaths, and vaccines in total and per million.

Data encoding:

In order to encode the data, firstly it was verified that some countries presented their new death's data with a minus and/or a plus, transforming the type of this variable into an object, so it was necessary to extract the signs in order to represent them as an integer. Secondly, in order to construct the map visualization, the data had to be transformed and so it was necessary to aggregate the dates and transform them into months, to facilitate the result of the slider.

This new transformation required an update on the other variables calculating the average in each month, in order to become a cleaner visualization.

Data filtering:

Data filtering has the main goal of choosing a smaller part of the data and then using that subset for a better viewing and analysis. Concerning this dataset, it is very important to take this into account because there are a lot of variables and a lot of entries to observe and interpret.

The user will be able to select which data he wants to see and compare among the countries available in the dataset. Within the figures it will be possible to zoom in, with the prospect of viewing the results better, as well as selecting just some of the variables.

Technical aspects:

It is possible to access our dash through this link :

<https://github.com/Dan20210677/DVProjectGroup7/>

This dash was implemented using languages such as python, html and css. Plotly was the main library used to implement the visualizations, and bootstrap in order to facilitate the styling of the final product.

In the first visualization, the map for the countries that do not possess data are presented in the color white, translating the absence of data. Adding to that when the user selects deaths per million and on the slider the first two months, due to the fact that there were not a lot of inputs it was thought best to replace it with an image saying that there is no data to be shown, moving forward to the third month onwards, the dash will display the geographic graph.

Concerning the second visualization, we paid a lot of attention to details and so if someone eliminates a country in the middle of the subplots displayed, this will be deleted without affecting the final result. Also, there is a possibility of removing all countries, presenting an empty graph. In this visualization, when combining the plots, the share axis button becomes unavailable.

In a general aspect, all the figures have been properly described informing the user of its main visualization goals.

Discussion:

With this project it was possible to represent a various number of variables related to covid in different ways.

While developing this dash it was encountered some limitations as the dataset did not possess all the variables necessary for the visualizations that we would have liked to develop. Although a great quantity of variables were provided by the dataset, these variables were not used due to the fact that it did not add any more valuable information in our perspective. One of our main concerns was to develop a clean and simple dashboard and so our choices concerning the variables were carefully thought in order to only give the necessary information in order to not overwhelm the user, but our visualizations could have been a lot more developed and with more information if we implement JavaScript in order to expand and display more details.

Line plots with different types of variables like new cases and deaths do not work combined because of the different magnitudes of values. The sliders had to be changed to monthly due to the fact that a daily or a weakly slider would not be possible due to the number of frames.

Regarding the future work it would be interesting to introduce more data concerning the impact on global pollution, emission of CO₂, discover the impact of covid in different diseases and many more situations that changed due to this pandemic, as well as create similar dash apps to other diseases, since nowadays all data is extracted and stored regularly. Additionally, much more information could be represented and visualized about the covid's dataset, such as testings, vaccinations, by country, and much more, with the use of JavaScript so that the dash would not become too cluttered with so many options and graphs.

Conclusion:

Our main goal was to make sure that the visualizations were simple and easy to understand, keeping this in mind creating an easy and intuitive to read so that the users could rapidly come to their conclusions. We also took into account colorblind and blind people, adding audio descriptions to the visualizations available.

We tried to make it as clean and minimalist as possible, though with the amount of data provided it became a challenge, buttons and a slider are the options to select and change any variables in order to facilitate the management of them.

The main conclusion that was taken from this project was that the pandemic affected the whole world, and the vaccines came to alleviate some of the disaster that came to some of the countries.

Bibliography:

1.

Plotly. Plotly.com. Published 2022. <https://plotly.com/python/>

2.

McCandless D. COVID-19 #CoronaVirus Infographic Datapack — Information is Beautiful. Information is Beautiful. Published March 9, 2020. <https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/>

3.

Otto M. Bootstrap. Getbootstrap.com. Published 2022. <https://getbootstrap.com/>

4.

Empowering the world to develop technology through collective knowledge. Stackoverflow.co. Published 2021. <https://stackoverflow.co/>

5.

Dash Bootstrap Components. Faculty.ai. Published 2022. <https://dash-bootstrap-components.opensource.faculty.ai/>

6.

W3Schools Free Online Web Tutorials. W3schools.com. Published 2022. <https://www.w3schools.com/>

7.

Denis O. COVID_19 Datasets. Kaggle.com. Published 2022. <https://www.kaggle.com/datasets/ognevdenis/covid-19-datasets>

8.

Preda G. COVID-19 World Vaccination Progress. Kaggle.com. Published 2022. <https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>