Visual Analytics of COVID-19 Immunization Data

Hima Poojitha Sai Sree Myla, Dheeraj Reddy Anugu, Singam Venkata Nishita , Srichand Medagani, Deepak Thondepu

University Of Maryland Baltimore County,MD,USA

Abstract-Each and every day, enormous amounts of data is produced, and its visual representation aids in giving individuals an instant understanding of the data. The Pandemic COVID-19 is one such source of data that has recently produced enormous amounts of data. Knowledge from this enormous amount of data helps the researchers, decision makers, epidemiologists to get a better understanding of the disease, vaccinations and awareness programs required. Globally, COVID has a lot of negative effects, and it is still continuing to be dangerous. We can all agree that prevention is better than cure, thus the WHO has developed vaccines to fend off covid and lessen its effects should someone contract it. Higher immunization dosages and better vaccinations, however, can help prevent the severity regardless of the aforementioned circumstances. In this paper, we portray the covid-19 data in the form of visualization dashboards on the density of covid instances and vaccines that may be used by the WHO officials or other authorities who take decisions on vaccinations. These visualizations are anticipated to be beneficial in helping people decide which additional immunizations are necessary.

I. INTRODUCTION

SARS-CoV-2 is the virus that causes COVID-19. It is a member of the coronavirus family, which also includes common viruses that cause more serious (but less common) disorders including Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS) (MERS). Coronaviruses, like many other respiratory viruses, spread swiftly through droplets that are emitted from the mouth or nose during breathing, coughing, sneezing, or speaking. With the spread of this deadly virus, the nation's hospital system and healthcare system are in jeopardy. The tracking of the virus among each and every person, as well as the gathering of data, has become crucial because the infection is contagious. With the collection of this data from 2020 to 2022 having enormous number of cases, deaths, vaccines has resulted in what is known as big data.

This tremendous information contain huge information and reasonable data. This requires the utilization of information science, which utilizes advances like AI, information examination, and visual investigation to extricate knowledge from tremendous information. The data that was found is useful. For example, the information acquired from these epidemiological information helps analysts, disease transmission specialists, and leaders in grasping the sickness, which might propel them to foster techniques for the recognition, avoidance, or potentially the executives of illnesses like viral illnesses.

II. RELATED WORK

Over the previous two years, a large number of dashboards and visualizers for the COVID-19 epidemic have been created. These dashboards primarily tracked the number of cases, fatalities, positive rate, and rate of global COVID-19 spread. Our depiction, in comparison, shows the specifics of immunization [2]. We have decided to choose our dataset to be Johns Hopkins dataset as the dataset consists of a lot of data points[8] and contains all the attributes which are suitable for our project. The dataset contains data from January 2020 to 30th October 2022. Not only the vaccination details but also the details regarding various health issues like smokers and people suffering from the cardiovascular disease [1]. In the process of validation of our project we have used the datasets of individual countries like India and USA for the compatibility check [3].

The main use case of our project is to make this work useful for the healthcare sector where the healthcare professionals get to know the details on what is the percentage of population that is vaccinated, how the production of the vaccinations should be controlled for the faster vaccination rate, how the awareness should be spread across among the people to get vaccinated [5][6][9].

III. IMPLEMENTATION

This research is being carried out to understand how coronavirus vaccinations are given and analysing the immunization data. The procedure begins with data collection from publicly accessible sources, followed by pre-processing and attribute selection. In Tableau, have used ZN function to convert the NULL numeric values to zero. It can handle both the aggregate and n

[ZN]: ZN([Numeric Field])

And with an aggregate:

[ZN]: ZN (SUM([Numeric Field]))

A. Covid-19 Vaccination rate on-aggregate values.

Fig.1 Vaccination rate across the world

The Vaccination count across the globe is represented in the above visualization making it easy to understand for the decision makers to take decisions on further immunizations. The above visualization have interactivity and is dynamic and the bubbles on the symbol map increases according to the month and year. The time series is taken into consideration.

B. Countries by Percentage of Population Fully Vaccinated.

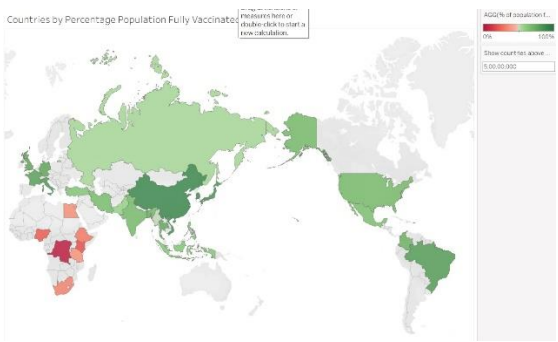

Fig.2 Details about Percentage of Population Fully vaccinated

The information about the rate of increase in the population getting vaccinated in the

country different in various countries which and is represented in a bubble map. The percentage of increase in the vaccination is represented in a choropleth map. To elaborate, as the vaccination rate increases in the country the size (radius) of the bubble increases. visualization describes the percentage of people or population in a country that are fully vaccinated. The map is the best way that we could come up with to represent the world data. We have created a new field that is percentage of population fully vaccinated from the fields: - People fully vaccinated & Population . We have used the following arithmetic formula to generate the desired field.

% People Fully Vaccinated= Max([People Fully Vaccinated])/MAX([Population]).

The marks used in the above visualization are Areas and the channels are color Hue,2D Area. We have made an attribute dynamic so as to ensure the user gets the

desired data. The search option on the top right of the dashboard is to search for the countries above a particular population The below is the image of the visualization n population above 300000000.

C. Top 20 countries in terms of vaccination.

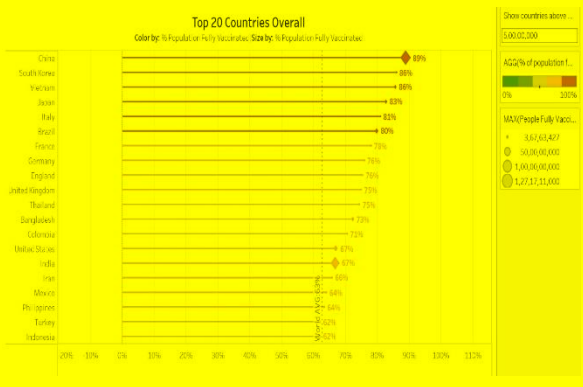

Fig.3 Visualization representing the top 20 Nations which has highest percentage of vaccinations

The top 20 nations with the highest immunization rates are represented in the visualization above. We have used the interaction mechanism which is the lollipop charts to represent the data. When comparing data sets or demonstrating how data changes over time, lollipop charts, which are quite similar to bar charts, can be utilized. The layout of the chart, which uses lines with circles at the ends to symbolize the data points that look like lollipops, is quite self-explanatory. When dealing with a large number of numbers and when the values are all high, such as in the 80–90% range (out of 100%), the lollipop chart is significantly more useful than a typical bar chart. Then a long row of towering columns could be overbearing. Channels like position, shape, color, and size have all been used. We used the position (horizontal lines), size (length varies the percentage), and color channel to indicate the aggregate proportion of the population fully vaccinated in order to depict the percentage of the population that is immunized. To depict the percentage, lines and points are utilized as marks. The population in each country is represented by the shape and size of the rhombus in the vis. To construct this visualization, we used data from the countries and the quantity of vaccinations received in each country. Furthermore, we employed quantitative and ordinal data, which are ordered qualities. We have used the ordinal scaling mechanism in our visualizations.

D. Top 20 Nations below world Average

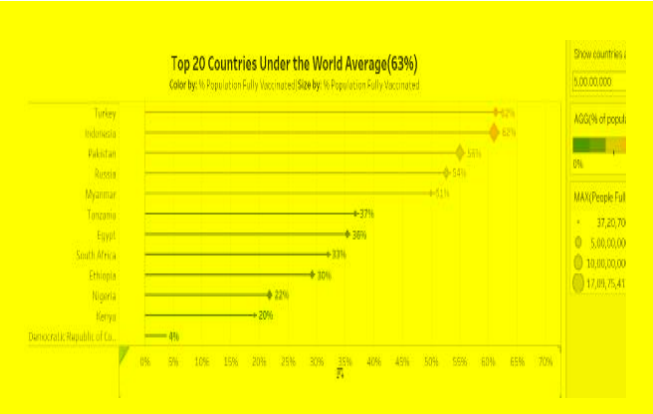

Fig 4.Visualization representing the top 20 Countries below the world average(63%)

The above visualization showcases the average vaccinations completed by different nations, where world average is 63%. From the visualization each country which has less than 63 % is showcased in the visualization. From this vis the user can identify that in which country the vaccinations should be increased and the stage where each nation is lacking in completing the vaccinations for each individual citizen. The size of rhombus in each nation varies as it represents the populations. As the size is big it represents more populations, and the small size represents less in populations. From the vis the colour changes according to the population size in the nation. Orange colour represents more population, yellow colour represents the ratio of average world vaccination to individual region, green colour showcases that it is in good ratio of covering the citizens in completing the vaccinations (Guatemala, Ukraine). The Lollipop chart representation is used to compare the vaccinations completed by nations and it is compared to world average (63%) of vaccinations. X Axis and Y axis represents the top 20 nations and the vaccinations completed by individual nation, percentages is represented on Y axis. Size of rhombus represents the population of nations.

E. Representation of Cardiovascular Death Rate.

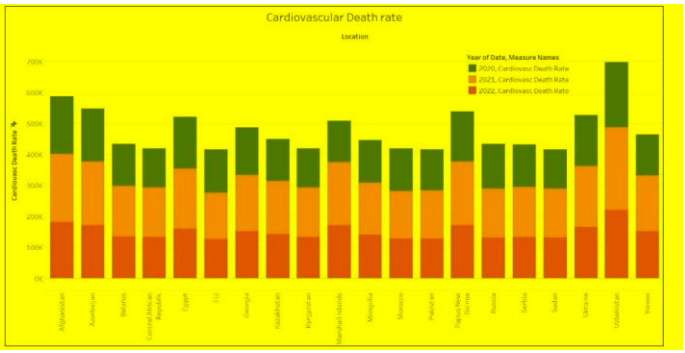

Fig.5 Visualization representing Cardiovascular Death rate .

The above Stacked bar charts visualization showcases cardiovascular deaths. Stacked Chart-A stacked chart is a type of bar graph that displays the composition and distinction of a few variables across time, either relative or absolute. Cardiovascular deaths- Deaths related to heart and blood vessels. The visualization represents the deaths which are caused by cardiovascular disease who are affected by covid-19. Cardiovascular diseases are related to heart and blood vessels. The stacked bar charts represent the death rate in the year 2020, 2021, 2022. Scale shows number of deaths in three years, every year is represented in different colour 2020 is displayed by blue, 2021 is displayed by orange, 2022 is displayed by red. X axis and Y axis describes deaths caused by cardiovascular in different nations. The user can get an overview of deaths which are caused by cardiovascular disease in different years. The required measures can be taken by individual nation to decrease there rate.

IV.EVALUATION

We have used qualitative validation approaches For checking the compatibility of the visualization that we have created, we used various Covid-19 data such as for United States and India[3]. Most COVID-19 visualizers or dashboards in use primarily concentrate on a single attribute (such as the number of cases, recoveries, or fatalities) at any given time. Our data visualization in comparison also illustrates

correlations various attributes like cardiovascular patients, smokers. How those with such medical issues were affected by the pandemic.

As part of validation, we considered the following evaluation forms, correction of truncated axes, the data can be twisted and mislead the viewer when a graph starts at a value other than zero. Changing the scale accordingly, in order for the reader to properly grasp any changes in the data, the scale for both the X and Y axis should rise accordingly. The scale can be compressed or expanded to make the changes in the data appear more or less substantial, respectively, than they actually are. Cleaning the dataset and manipulating the NULL values. Chosen the better idioms to represent our data visualizations. The representation is very user-friendly and easy to understand.

V.FUTURE DIRECTIONS

- Adding the vaccination information and types of vaccinations can be included if proper resources permitted.
- Showing the different types of vaccinations impact over the country
- Data can be used as which vaccine is more effective and in which country it worked well.
- So that user can choose vaccine which impacted more.
- Types of manufactures around the world and there efficiency which worked well.

VI.CONCLUSION

From the recent two years, many labs are working on developing and analysing different aspects of data related to COVID-19 Datasets. We have mainly focused on number of cases and total number of vaccinations with respect to their location. We have also worked on finding the Top-20% Covid-19 vaccinated countries and the next below average 20 countries. The world average vaccination rate is 63% as per the latest details. This information is very useful for people who runs health sectors as well as for normal people. We have created a dashboard which has all these details and also, we have dynamic searching feature where we can choose countries cases details as per the population selected and we will get different colors for each country according to the number of cases. we have created interactive visualizations where we also have details of cardiovascular patients affected by Covid-19.

By observing the percentage of vaccination for every country through the map we can easily conclude that the vaccination for many countries is still low. The Country with highest vaccination percentage as per the recent update is China with 89%. We also represent the number of vaccines according to specific date and month of particular year. This type of visualization is always helpful for the people who runs health organizations that can take necessary precautions to completely eradicate Covid-19 disease.

VII.REFERENCES

[1] C. K. Leung, Y. Chen, C. S. H. Hoi, S. Shang, Y. Wen and A. Cuzzocrea, "Big Data Visualization and Visual Analytics of COVID-19 Data," 2020 24th International Conference Information Visualisation (IV), 2020, pp. 415-420, doi: 10.1109/IV51561.2020.00073.

[2] M. Jamshidi et al., "Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment," in IEEE Access, vol. 8, pp. 109581-109595, 2020, doi:

10.1109/ACCESS.2020.3001973.

[3] A. Pravin, T. P. Jacob, K. M. Prasad, T. Judgi and R. Rajakumar, "Data Analysis on Detection and Prediction of Covid-19," 2022 3rd International Conference for Emerging Technology (INCET), 2022, pp. 1-4, doi: 10.1109/INCET54531.2022.9824539.

[4] B. Meng, S. Cheng and A. Kumar, "Big Data Visualization Analysis: Distribution of COVID-19 Mortality and Vaccination in the US," 2022 International Symposium on Electrical, Electronics and Information Engineering (ISEEIE), 2022, pp. 8-12, doi: 10.1109/ISEEIE55684.2022.00009.

[5] G. Wolfe, A. Elnashar, W. Schreiber and I. Alsmadi, "COVID-19 Candidate Treatments, a Data Analytics Approach," 2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA), 2020, pp. 139-146, doi: 10.1109/MCNA50957.2020.9264290.

[6] A. P. Hartono, C. R. Luhur and N. N. Qomariyah, "Forecasting Vaccination Growth for COVID-19 using Machine Learning," 2022 5th International Conference on Computing and Informatics (ICCI), 2022, pp. 356-363, doi: 10.1109/ICCI54321.2022.9756096

[7] Y. Dong and Y. -D. Yao, "IoT Platform for COVID-19 Prevention and Control: A Survey," in IEEE Access, vol. 9, pp. 49929-49941, 2021, doi: 10.1109/ACCESS.2021.3068276.

[8] S. Shang, C. K. Leung, Y. Chen and A. G. M. Pazdor, "Spatial Data Science of COVID-19 Data," 2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2020, pp. 1370-1375, doi: 10.1109/HPCC-SmartCity DSS50907.2020.00177.

[9] I Bhat, "Comparative analysis of Covid- 19 Data: A statistical approach for unraveling the relation between the awareness and healthcare system from Indian perspective," 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), 2022, pp. 1-4, doi: 10.1109/ICSTSN53084.2022.9761353