CSCI – 6406 - Visualization

Project Report Winter 2021-2022

- Sumadhur Ramarao Vaidyula – B00871907

1] Introduction:

With the reappearance of the covid-19 virus across the globe, the world has been ravaged by an epidemic since the beginning of the year 2020. The globe has been harmed by its emergence because it is a virus that spreads quickly from humans and has resulted in numerous deaths. This pandemic posed a new set of challenges for data scientists, as it generated a massive amount of data globally. Keeping track of all the virus's effects, as well as analysing patterns and understanding the intensity of the spread across the globe, became critical in order to prevent further outbreaks and educate the public.

The goal of the project is to create a dashboard for visualizing various aspects of Covid-19 data, which is critical for tracking the disease around the world, analyzing situations in different regions, tracking the progress of improvement made by various countries, and predicting any potential future consequences. The project's purpose is to make the visualization dashboard simple to understand for visible eyes, so that anyone without a background in data science may come to a conclusion by looking at it.

The data for the project will come from John Hopkins University, which is constantly updated to reflect changing global trends in covid-19. John Hopkins University [5] has made various covid-related data available on Github for application and research purposes.

2] Motivation for developing Covid-19 Dashboard:

Related Work:

Due to rise in Covid-19, any dashboards were developed by many organizations. While few developed few literature articles on Covid-19, and few focuses on economic impacts of covid-19. [7]

Motivation:

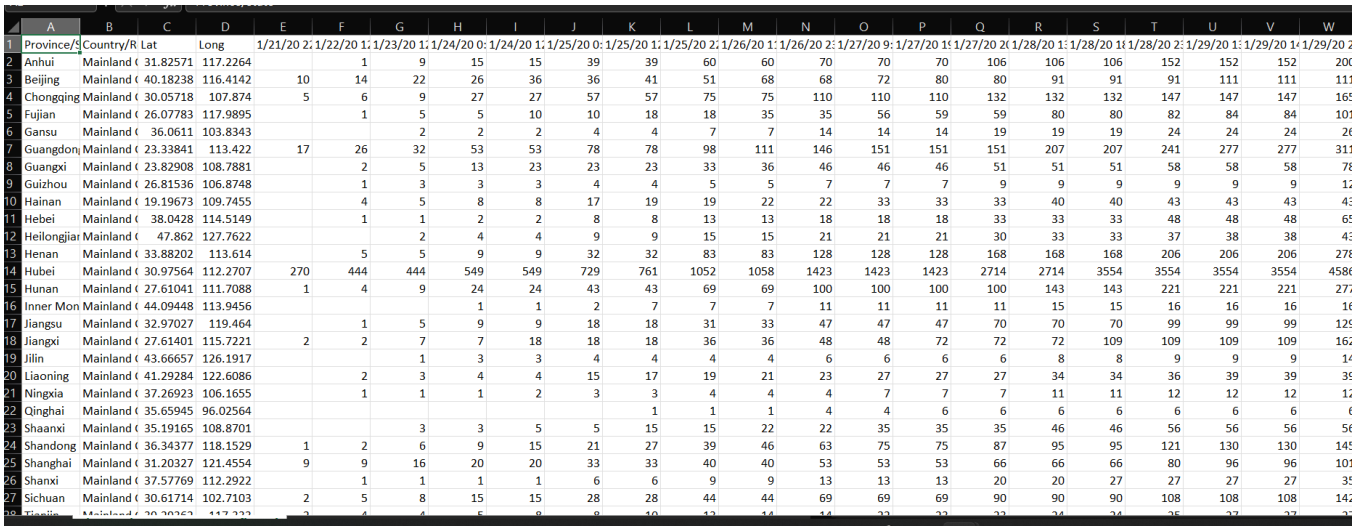
Due to increasing effects od Covid-19 across globe, a lot of industries varied in their approach. Also, few industries, suffered extreme losses due to the impact of covid-19. So, in order to monitor the global pattern regarding the covid-19 situation, this dashboard was implemented. John Hopkins University[5] is one of the ideas behind implementing the dashboard. The university highlights constant updates on the dashboard. Many decisions can be influenced based on the covid-19 situation across globe. Few of the examples include, the investment patterns, the travel industry situation can be monitored as well, and many more.

Project Requirements:

The project is implemented using D3.js [12], which is a JavaScript library for data-driven document manipulation. Using HTML, SVG, and CSS, D3 allows to bring data to life. With powerful visualization components and a data-driven approach to DOM manipulation, D3 combines powerful visualization components and a data-driven approach to DOM manipulation to give the full capabilities of modern browsers without tying a proprietary framework.

3] Datasets Used in the Project:

1) Dataset used for Confirmed cases Dashboard on choropleth Map



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Province/S	Country/R	Lat	Long	1/21/20 2:	1/22/20 1:	1/23/20 1:	1/24/20 0:	1/24/20 1:	1/25/20 0:	1/25/20 1:	1/25/20 2:	1/26/20 1:	1/26/20 2:	1/27/20 9:	1/27/20 1:	1/27/20 2(	1/28/20 1:	1/28/20 1:	1/28/20 2:	1/29/20 1:	1/29/20 1:	1/29/20 2
2	Anhui	Mainland (	31.82571	117.2264		1	9	15	15	39	39	60	60	70	70	70	106	106	106	152	152	152	200
3	Beijing	Mainland (	40.18238	116.4142	10	14	22	26	36	36	41	51	68	68	72	80	80	91	91	91	111	111	111
4	Chongqing	Mainland (	30.05718	107.874	5	6	9	27	27	57	57	75	75	110	110	110	132	132	132	147	147	147	165
5	Fujian	Mainland (	26.07783	117.9895		1	5	5	10	10	18	18	35	35	56	59	59	80	80	82	84	84	101
6	Gansu	Mainland (	36.0611	103.8343			2	2	2	4	4	7	7	14	14	14	19	19	19	24	24	24	26
7	Guangdon	Mainland (	23.33841	113.422	17	26	32	53	53	78	78	98	111	146	151	151	151	207	207	241	277	277	311
8	Guangxi	Mainland (	23.82908	108.7881		2	5	13	23	23	23	33	36	46	46	46	51	51	51	58	58	58	78
9	Guizhou	Mainland (	26.81536	106.8748		1	3	3	3	4	4	5	5	7	7	7	9	9	9	9	9	9	12
10	Hainan	Mainland (	19.19673	109.7455		4	5	8	8	17	19	19	22	22	33	33	33	40	40	43	43	43	43
11	Hebei	Mainland (	38.0428	114.5149		1	1	2	2	8	8	13	13	18	18	18	33	33	33	48	48	48	65
12	Heilongjiar	Mainland (	47.862	127.7622			2	4	4	9	9	15	15	21	21	21	30	33	33	37	38	38	43
13	Henan	Mainland (	33.88202	113.614		5	5	9	9	32	32	83	83	128	128	128	168	168	168	206	206	206	278
14	Hubei	Mainland (	30.97564	112.2707	270	444	444	549	549	729	761	1052	1058	1423	1423	1423	2714	2714	3554	3554	3554	3554	4586
15	Hunan	Mainland (	27.61041	111.7088	1	4	9	24	24	43	43	69	69	100	100	100	100	143	143	221	221	221	277
16	Inner Mon	Mainland (	44.09448	113.9456				1	1	2	7	7	7	11	11	11	11	15	15	16	16	16	16
17	Jiangsu	Mainland (	32.97027	119.464		1	5	9	9	18	18	31	33	47	47	47	70	70	70	99	99	99	129
18	Jiangxi	Mainland (	27.61401	115.7221	2	2	7	7	18	18	18	36	36	48	48	72	72	72	109	109	109	109	162
19	Jilin	Mainland (	43.66657	126.1917			1	3	3	4	4	4	4	6	6	6	6	8	8	9	9	9	14
20	Liaoning	Mainland (	41.29284	122.6086		2	3	4	4	15	17	19	21	23	27	27	27	34	34	36	39	39	39
21	Ningxia	Mainland (	37.26923	106.1655		1	1	1	2	3	3	4	4	4	7	7	7	11	11	12	12	12	12
22	Qinghai	Mainland (	35.65945	96.02564							1	1	1	4	4	6	6	6	6	6	6	6	6
23	Shaanxi	Mainland (	35.19165	108.8701			3	3	5	5	15	15	22	22	35	35	35	46	46	56	56	56	56
24	Shandong	Mainland (	36.34377	118.1529	1	2	6	9	15	21	27	39	46	63	75	75	87	95	95	121	130	130	145
25	Shanghai	Mainland (	31.20327	121.4554	9	9	16	20	20	33	33	40	40	53	53	53	66	66	66	80	96	96	101
26	Shanxi	Mainland (	37.57769	112.2922		1	1	1	1	6	6	9	9	13	13	13	20	20	27	27	27	27	35
27	Sichuan	Mainland (	30.61714	102.7103	2	5	8	15	15	28	28	44	44	69	69	69	90	90	90	108	108	108	142

Figure 3.1 – Confirmed Cases Dataset

Data source: https://github.com/CSSEGISandData/COVID-19/blob/master/archived_data/archived_time_series/time_series_2019-ncov-Confirmed.csv [5]

Data processing:

The Sum of time series data was calculated, using excel formulas, and a single column was presented with all the cumulation of all the data.

For the “Code” column, as shown in the figure, the dataset of World population was used, to extract country code of each country.

	A	B	C
1	name	code	pop
2	Antigua an	ATG	83039
3	Algeria	DZA	32854159
4	Azerbaijan	AZE	8352021
5	Albania	ALB	3153731
6	Armenia	ARM	3017661
7	Angola	AGO	16095214
8	American S	ASM	64051
9	Argentina	ARG	38747148
10	Australia	AUS	20310208
11	Bahrain	BHR	724788
12	Barbados	BRB	291933

Figure 3.2 – World Population dataset

Data Source: https://raw.githubusercontent.com/holtzy/D3-graph-gallery/master/DATA/world_population.csv

	A	B	C
1	Country	code	Total
2	Antigua an	ATG	4632
3	Algeria	DZA	17648
4	Azerbaijan	AZE	7015
5	Albania	ALB	61
6	Armenia	ARM	10197
7	Angola	AGO	2767
8	American S	ASM	3681
9	Argentina	ARG	1771
10	Australia	AUS	1682
11	Bahrain	BHR	456
12	Barbados	BRB	579
13	Bermuda	BMU	160974
14	Bahamas	BHS	4497
15	Banglades	BGD	2118
16	Belize	BLZ	2632

Figure-3.3 [Covid_confirmed_cases_data.csv] Final Dataset for visualization [5]

2) Dataset used for Vaccination record for Bubble Map:

Covid-19 vaccination

country	iso_code	date	total_vaccinations	people_vaccinated
Country	ISO Code	Date	Total vaccinations	People vaccinated
	223 unique values	1Dec20 – 28Mar22	0 – 3.26b	0 – 1.28b
Afghanistan	AFG	2021-02-22	0.0	0.0
Afghanistan	AFG	2021-02-23		
Afghanistan	AFG	2021-02-24		
Afghanistan	AFG	2021-02-25		
Afghanistan	AFG	2021-02-26		
Afghanistan	AFG	2021-02-27		
Afghanistan	AFG	2021-02-28	8200.0	8200.0
Afghanistan	AFG	2021-03-01		
Afghanistan	AFG	2021-03-02		

Figure 3.4 Vaccination Dataset

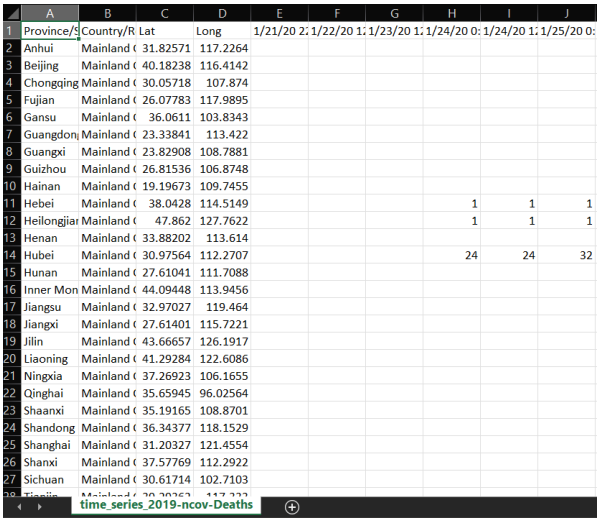
Data Source: https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress?resource=download [6]

	A	B	C	D	E
1	country	iso_code	total_vaccinations	Lat	Long
2	Afghanista	AFG	118041426	33	65
3	Albania	ALB	295540465	41.1533	20.1683
4	Algeria	DZA	238946062	28.0339	1.6596
5	Andorra	AND	2447643	42.5063	1.5218
6	Angola	AGO	388740243	-11.2027	17.8739
7	Anguilla	AIA	806657		
8	Antigua an	ATG	6840405	17.0608	-61.7964
9	Argentina	ARG	18678951288	-38.4161	-63.6167
10	Armenia	ARM	34983105	40.0691	45.0382
11	Aruba	ABW	38528732		
12	Australia	AUS	9733094247	-33.8688	151.2093
13	Austria	AUT	596758436	47.5162	14.5501
14	Azerbaijan	AZE	2364582887	40.1431	47.5769
15	Bahamas	BHS	8527171	25.0343	-77.3963
16	Bahrain	BHR	778833949	26.0275	50.55
17	Banglades	BGD	13205078414	23.685	90.3563
18	Barbados	BRB	76587881	13.1939	-59.5432
19	Belarus	BLR	198340415	53.7098	27.9534
20	Belgium	BEL	5762415299	50.8333	4
21	Belize	BLZ	21555588		
22	Benin	BEN	24937864	9.3077	2.3158

Figure – 3.5 Processed Dataset

The required columns were identified and the irrelevant columns from the data were removed from the dataset.

3) Dataset used for Deaths and Recovered Data for plotting Lollipop Chart:



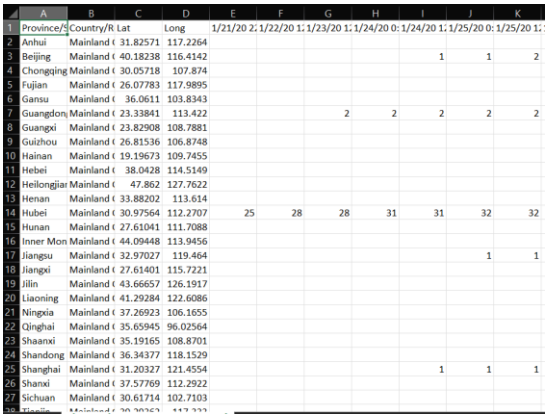


Figure 3.6 – Deaths and Recovered dataset

Data Source:

Deaths: https://github.com/CSSEGISandData/COVID-19/blob/master/archived_data/archived_time_series/time_series_2019-ncov-Deaths.csv [5]

Recovered: https://github.com/CSSEGISandData/COVID-19/blob/master/archived_data/archived_time_series/time_series_2019-ncov-Recovered.csv [5]

Data Processing:

For the processing of data, the top 20 records of the Recovered and Deaths data were grouped and merged, to plot the data. Python and Jupyter Notebook was used for the processing. Initially, the top 20 records of Deaths were identified from the Deaths dataset and then top 20 records from Recovered dataset were identified from the recovered dataset. (The notebook of processing of data is included in the project zip folder).

	A	B	C
1	country	Deaths	Recovered
2	US	5.99E+08	4115
3	Brazil	4.34E+08	3.41E+09
4	India	3.09E+08	72140011
5	Mexico	2.2E+08	1203220
6	Peru	1.69E+08	9755165
7	Russia	1.43E+08	48133031
8	United Kin	1.26E+08	2654988
9	Italy	1.18E+08	641874
10	France	1.02E+08	1.11E+08
11	Colombia	88715074	41026
12	Iran	87387298	167935
13	Spain	80752712	3350829
14	Argentina	78805084	7.12E+08
15	Germany	75479402	5.22E+08
16	Indonesia	75055282	28445567
17	South Afric	60886832	2307589
18	Poland	58368788	1469023
19	Ukraine	48242682	28288270
20	Turkey	46070288	515923
21	Romania	30958286	56875095

Figure 3.7 – Final Dataset of Deaths and Recovered Cases

4) Dataset used for Case Fatality Ratio for plotting Bar graph:

Top 10 countries with highest Case fatality Ratio were identified and plotted on Bar graph.

	A	B	C
1	Country_R	Case_Fatality_Ratio	
2	US	63.98822	
3	Yemen	29.03379	
4	MS Zaanda	22.22222	
5	Mexico	15.78575	
6	Peru	14.7885	
7	United Kin	7.692308	
8	China	6.620787	
9	Ecuador	6.585028	
10	Syria	6.220718	
11	Japan	6.169666	

Figure 3.8 – Death- Recovered Dataset

Data Source: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_daily_reports/01-01-2021.csv [5]

4] Implementation:

1] Choropleth Map using D3.js

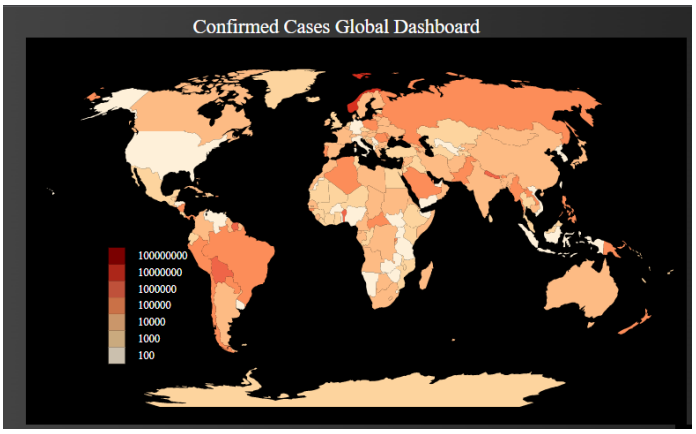


Figure 4.1 – Choropleth Map [1]

Initially, a map structure was needed to be generated, to get a layout of the coordinates, to plot the relevant data on the map. The dataset used for generating map data is the WorldGeo.jspon, available from, https://raw.githubusercontent.com/holtzy/D3-graphgallery/master/DATA/world.geojson . Once the Map was plotted, the processed data was parsed using d3.js, and then it was plotted on the map. The map data gets updated, on changing the values in the CSV file. Choropleth map was used to present the confirmed cases in the globe, between a specified frame, obtained from John Hopkins University. The data in the choropleth Map were plotted based on the country code mentioned in the dataset. For making it interactive, several features were added, as the map is zoomed, when clicking on a nation and also on hovering on to a

country location, a pop-up raises to display the statistics. The shades indicate the number of cases in a country, and to visible eye, an individual can assess the global situation regarding the confirmed cases.

2] Bubble Map using D3.js

The Bubble Map was implemented using D3.js, the map for it was plotted using the same WorldGeo.json data. Although, unlike choropleth Map, the data was plotted using the coordinated of Longitude and Latitude specified in the Dataset. Th bubble Map was used to visualize the data of Vaccination records of all countries. For making it interactive, on hovering the mouse over, it displays the statistical data of the bubble.

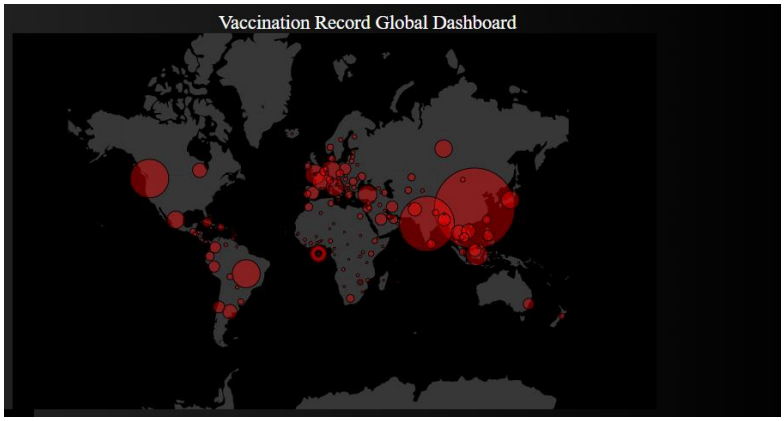


Figure 4.2 Bubble Map [2]

3] Horizontal Bar graph using D3.js

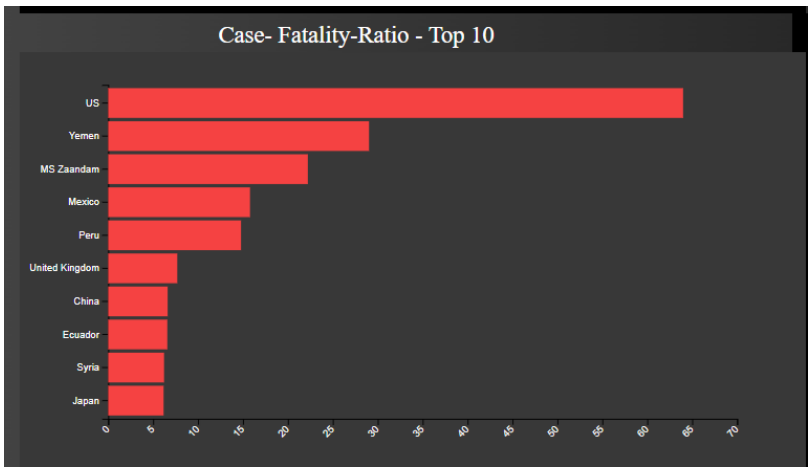


Figure 4.3 Bar chart [3]

The horizontal bar chart was used to display, the data of top 10 countries with highest case fatality ratio. The data from the CSV file is parsed using D3.js and the data gets updated based on update in CSV file.

4] Lollipop -Chart using D3.js

The Lollipop chart was implemented using D3.js, to visualize the data of top countries with Highest Deaths and Recovered Cases. Two buttons were merged with the dashboard, on clicking the button, the data is switched to the specified option. Also, the data can be modified in the CSV file, and the updates can be reflected in the dashboard as well.

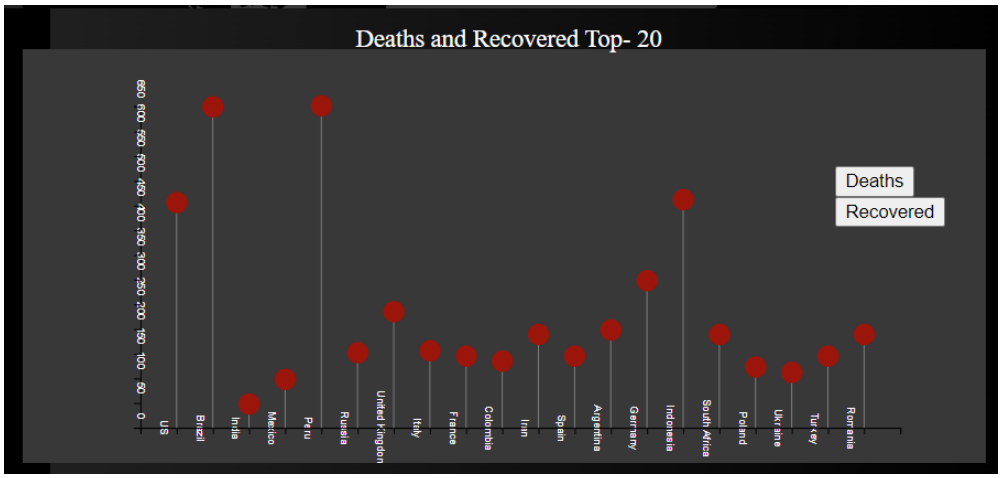


Figure 4.4 Lollipop Chart [4]

Future Work:

This project can further be expanded to a public dashboard, where not only general statistics are highlighted, but various combinations how particular situation in a country is impacting various sectors can be highlighted as well. Economic impacts of the country can be highlighted alongside the confirmed cases data dashboard, which will indicate key insights regarding the performance of various sectors in the country at the global level.

Furthermore, the data of stock market investments can be highlighted, where several industries helped curing the aid, by developing vaccinations, in that way, people will get the idea about the industries with current scope of development, with that, not only the companies will be benefitted but the people investing their money will also get few benefits.

Conclusion:

There have been efforts in the last few months to visualise and analyse various parts of big data connected to COVID-19. This project highlighted that it is possible to create COVID-19 scenario dashboards in low-resource metropoles. They provide real-time access to data that can aid in the fight against the disease. The outcomes of this idea suggest that national media-based prevention programmes should be backed up with fast community awareness actions. It is hoped that the positive findings will allow a small portion of the resources committed to combating the COVID-19 pandemic in a low-resource setting to be directed toward developing dashboards like this, particularly in highly populated areas.

References:

[1] D3.js Graph Gallery, "*Choropleth Map*", Available: https://d3-graph-gallery.com/choropleth.html [Online]. Accessed: [30th March 2022]

[2] D3.js Graph Gallery, "*Bubble Map*", Available: https://d3-graph-gallery.com/bubblemap.html [Online]. Accessed: [1st April 2022]

[3] D3.js Graph Gallery, "*Horizontal bar chart in d3.js*", Available: https://d3-graph-gallery.com/graph/barplot_horizontal.html [Online]. Accessed: [1st April 2022]

[4] D3.js Graph Gallery, "*Change Lollipop Chart input data with a button*", Available: https://d3-graph-gallery.com/graph/lollipop_button_data_csv.html [Online] Accessed: [2nd April 2022]

[5] Covid-19 Dataset, John Hopkins University, "*COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University",* Available: https://github.com/CSSEGISandData/COVID-19 [Online]. Accessed: [1st April 2022]

[6] Kaggle, "COVID-19 World Vaccination Progress", Available: https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress?resource=download [Online]. Accessed: [1st April 2022]

[7] 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. [Accessed: 1st February 2022].

[8] Fan Bao, and Jia Chen. "Visual Framework for Big Data in D3.js." 2014 IEEE Workshop on Electronics, Computer and Applications (2014): 47-50. Web. [Accessed: 28th March 2022]

[9] Schwab, Michail, David Saffo, Nicholas Bond, Shash Sinha, Cody Dunne, Jeff Huang, James Tompkin, and Michelle Borkin. "Scalable Scalable Vector Graphics: Automatic Translation of Interactive SVGs to a Multithread VDOM for Fast Rendering." IEEE Transactions on Visualization and Computer Graphics PP (2021): 1. Web. [Accessed: 28th March 2022]

[10] Amr, Tarek & Stamboliyska, Rayna. (2016). Practical D3.js. 10.1007/978-1-4842-1928-7. [Accessed: 28th March 2022]

[11] Shander, B. (Speaker). (2019). *Animation in D3* [Video]. SAGE Research Methods: Data Visualization https://dx.doi.org/10.4135/9781529773620 [Accessed: 28th March 2022]

[12] Michael Bostock, Vadim Ogievetsky and Jeffrey Heer, "Data-Driven Documents", Index Terms—Information visualization, user interfaces, toolkits, 2D graphics. Available: http://vis.stanford.edu/files/2011-D3-InfoVis.pd [Online]. Accessed: [29th March 2022].