

Visual Analysis of COVID-19

Group:13
Sai Vinay Nandigam
Tarun Upputuri

Introduction

In December 2019, an unknown infectious disease that seems like pneumonia was found in Wuhan, China. Now called as Corona Virus, which is a contagious disease. The people affected by this disease had gone through respiratory illness. Many people worldwide lost their lives from the start of the disease to date, and still, people are fighting with the disease. There are many ways out that can help stop the spread of the disease. One of the most effective ways is to share the daily cases and the trends from the start of the disease spread. Using the data visualization, we can make everyone from kids to older people easily understand.

Background and Motivation:

The COVID-19 has affected more than 200 countries. As of 16 October 2021, the confirmed cases are 241,033,168, and the confirmed deaths are 4,907,728. To control the spread of COVID-19 lot of research, data analysis, medical analysis, new medical treatments, etc., are taking place. There are many ways out that can help stop the spread of the disease.

As data is everywhere and evolving at rocket speed with the help of advanced technologies, a large volume of data can easily be collected, generated, and stored. By using the data, one can discover valuable information by processing the data. There are many methods to gain insights from the data. But "A Picture is worth a thousand words" is easy to understand and gain insights from the visualization. By using COVID-19 datasets, most of the available visuals and dashboards are being generated to understand the COVID-19 trends by the public quickly. Most of the visualizations are about confirmed cases and mortality. They reveal the spatial differences between the countries, states, and cities. And temporal differences between days and weeks. A lot of studies and visualizations are primarily focused on confirmed cases and deaths.

There is a need to know about the recovered patients and vaccinated people also. Because knowing the recovery and vaccination rate would make them mentally strong to fight against the pandemic. Just knowing the confirmed cases and deaths would lead them to mental depression. In this paper, compelling visualizations of COVID-19 are going to be displayed.

Using a Bubble Map, the severity of the covid-19 can be visualized by the sizes of the bubbles representing the severity. In Fig. 1, the severity of the cases is depicted, but many bubbles are overlapped. It is challenging to find out the seriousness of the virus in European countries, Asian countries, and North America. After gathering the related data from the source and updated it in an excel sheet, we have not observed any notable null or missing values in the data set of different attributes.

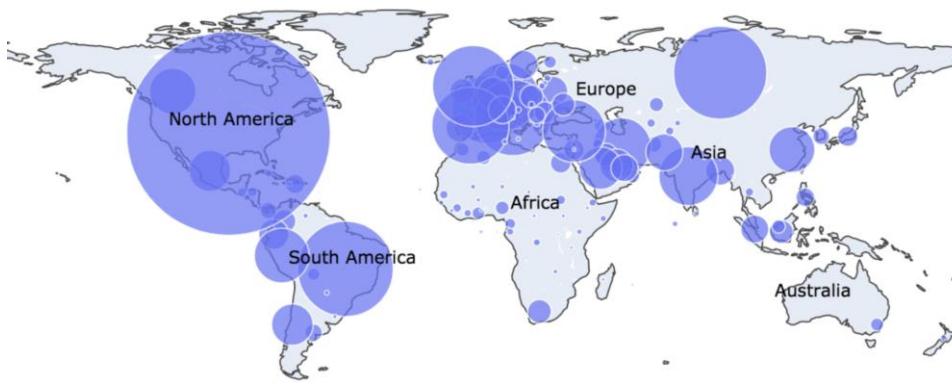


Fig. 1 Bubble Map

There is a need to know about the recovered patients and vaccinated people also. Because knowing the recovery and vaccination rate would make them mentally strong to fight against the pandemic. Just knowing the confirmed cases and deaths would lead them to mental depression. In this paper, compelling visualizations of COVID-19 are going to be displayed.

Project Objectives:

In the current period, a huge amount of data is being generated, which becomes difficult for one to understand and gain insight into the problem. A space filled with data visualizations has become the recent key to communicate vital information, further helping the respective users. Our primary goal is visually analysing the data, which makes it easier to comprehend and derive the knowledge related to the specific topic.

The primary questions which we are addressing in our project are

1. The number of COVID-19 cases in different countries and how it varies from region to region according to the time to understand at what rate the disease spreads and progresses, which would be the key to take countermeasures by the respective authorities. It is also helpful to use the data to impose travel restrictions to those respective regions and stopping the spread.
2. The most affected parts of the country or a region that would observe the positivity rates and the number of deaths would be helpful to the people to know about the condition and severity level of the disease. It is beneficial for the government to make modifications to the health policies and take rapid action according to the severity of the disease.

3. The five most affected countries according to the number of COVID-19 cases and the recovery rates will be identified to evaluate the measures taken and the results. Knowing this would be helpful to understand numerous factors that have led those respective countries to suffer more severely than others.
4. The number of COVID-19 cases registered in a specific region of the province at one time would be helpful for the local authorities in back tracing and movement of the disease spread.

Data

We took the reference data set from the "John Hopkins University" open-source repository resource centre. They collect data from different trusted sources and update it in an excel sheet. The data set contains the details of the confirmed cases, recovered rates, deaths Region/ province wise, world-wise from January 2020. We will be deriving the following tasks from the confirmed cases, death rates, and recovered cases.

- The daily number of COVID-19 cases.
- Fatality rates in different regions.
- Daily recovered cases.
- The most affected countries in the world.
- The most affected continent.

Data set link: <https://github.com/CSSEGISandData/COVID-19>

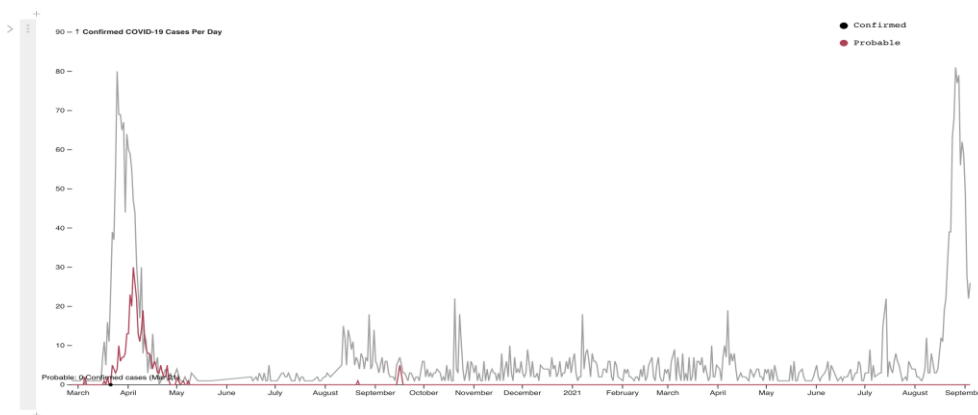
Data Processing

After gathering the related data from the relevant and trusted resources and updated it in an excel sheet, we have not observed any notable null or missing values in the data set of different attributes. As the daily data of the covid-19 is being updated by the Johns Hopkins University there are no null values in the dataset.

Visualization 1: A line graph for showing Confirmed cases:

We want to show the monthly confirmed COVID-19 cases around the world through a line graph. A line graph can show the data perfectly, the way the number of cases fluctuates.

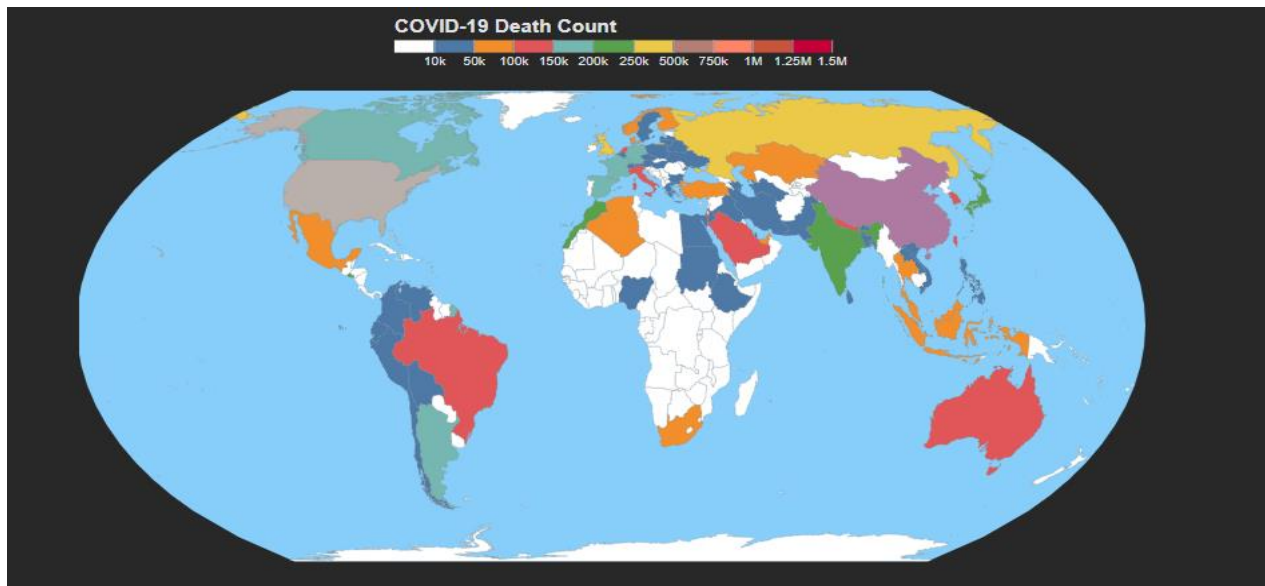
A Sample Line Graph



Visualization 2: A Colour map showing number of deaths in all countries

We want to show the death count of people in all the nations. Instead of a bubble map, we are using a Colour map. As there are over 207 countries affected by COVID-19 in the world and death counts in each nation, a bubble map would result in overlapping between the nations. To differentiate which countries have more death counts, we will use colour hue according to the severity. This visualization would result in a compelling visualization.

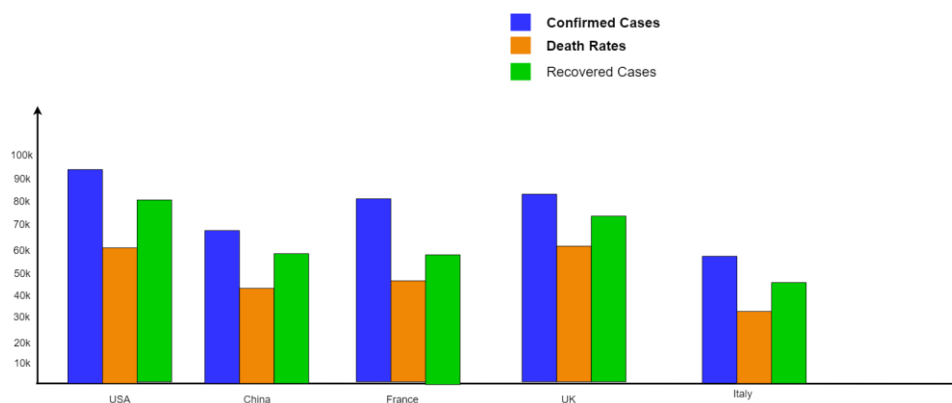
A sample colour map indicating deaths



Visualization 3: A Grouped Bar Chart comparing five most affected countries.

We want to compare the most severely affected five countries in a grouped bar chart considering three attributes: Confirmed cases, recovered count, and death count. We will use grouped bar charts instead of individual bar charts where grouped bar charts help to gain the knowledge in the single graph. Whereas using bar charts to get the same information, we need to draw multiple graphs. In a grouped bar chart, using three distinct colours can result in a compelling graph.

A Sample Grouped graph



References:

- [1] Big Data Visualization and Visual Analytics of COVID-19 Data, Carson K. Leung; Yubo Chen; Calvin S.H. Hoi; Siyuan Shang; Yan Wen; Alfredo Cuzzocrea, IEEE,
<https://ieeexplore.ieee.org/document/9373130>
- [2] Interactive Data Driven Visualization for COVID-19 with Trends, Analytics and Forecasting,
<https://ieeexplore.ieee.org/document/9373291>