

### Assignment-3 : Interactive Visualization

Darshit Amit Pandya

#### Part-1 : Description of Data Domain and Visualization Design

- **Data Domain:**

Considering the recent spikes in the novel diseases across the world, it seemed to be a domain wherein interactive visualizations could play a critical role in the policy creation to mitigate and respond to such diseases. As the diseases relate to a human's life, it is critical to precisely infer from the data as it could be resulting in keeping a human's life at risk. Accordingly, in order to help the user precisely infer from a visualization, it is crucial to implement a visual representation of the data that meets both: Expressiveness and Effectiveness criterias, in accordance with the visualization laws.

In recent times, the Covid-19 disease demonstrated the severity with which an invisible disease can affect humankind. Governments across the globe rushed to respond to the crisis due to the disease. They implemented various strict restrictions and SOPs for people's safety. Although their SoPs might differ, the backbone/foundation/basis for the decisions was: Data. It reinforces the impact of data visualizations on humankind in such critical scenarios. Accordingly, I decided to attempt to utilize my comprehension of Data Visualization to create an Interactive Visualization using the Covid-19 dataset, pertaining to the availability and the quality of the data available in the public domain.

- **Dataset Description & Metrics:**

- Source: Kaggle Datasets  
(<https://www.kaggle.com/datasets/niketchauhan/covid-19-time-series-data/data?select=time-series-19-covid-combined.csv>)
- Dataset Metrics:
  - Attributes/Columns : 8 (+3 Implicit Attributes)
  - Entries/Rows : 72K

Attribute Name	Attribute Type	Attribute Description
Date	Ordinal	Date of the record
Country/Region	Nominal	Country
Province/State	Nominal	State
Lat	Quantitative	Latitude of the Country/State
Long	Quantitative	Longitude of the Country/State
Confirmed	Quantitative	Total Confirmed cases till this Date
Recovered	Quantitative	Total Recovered cases till this Date
Deaths	Quantitative	Total Deaths till this Date

- **Data Cleaning & Data Pre-processing:**

Although the dataset was mostly complete, following are the transformations that I implemented on the data before utilizing it for visualization:

- **Attribute : “Province”:** This attribute had the Province values for only the following countries: Canada, Australia, China, Denmark, France, The Netherlands, and the United Kingdom. For the rest of the countries, the data is only available for the country as a whole and Province-wise data is not available. Accordingly, instead of ignoring the Province data, I just visualized the data by mentioning the province name along with the country name wherever available, pertaining to the fact that this visualization might form a critical component for a user who may act accordingly, and hence, I tried to implement all the available data to maintaining Expressiveness.
- **Attribute: “Country/Region”:** As I will be utilizing the GeoJSON format to implement a Choropleth Map, I had to adjust the Country names to match the names in the GeoJSON file in order for it to work properly.
- **Attribute: “Confirmed”:** As per the description of the dataset by the author on the source website, this attribute represents: *Total Number of Confirmed Cases per day*. Although the description sounds like it represents the total number of cases confirmed on that particular day, it actually represents the Cumulative Cases or the total confirmed cases TILL that day from when the data has been recorded. I confirmed this fact by analyzing the actual data on the WHO website. Also, I calculated the new cases on a particular date (Daily Cases) by utilizing this attribute value.
- **Attribute: “Recovered”:** As per the description of the dataset by the author on the source website, this attribute represents: *Total Number of Recovery Cases per day*. Although the description sounds like it represents the total number of recoveries on that particular day, it actually represents the Cumulative Recoveries or the total recoveries TILL that day from when the data has been recorded. I confirmed this fact by analyzing the actual data on the WHO website. Also, I calculated the new recoveries on a particular date (Daily Recoveries) by utilizing this attribute value.
- **Attribute: “Deaths”:** As per the description of the dataset by the author on the source website, this attribute represents: *Total Number of Death Cases per day*. Although the description sounds like it represents the total number of deaths confirmed on that particular day, it actually represents the Cumulative Deaths or the total deaths TILL that day from when the data has been recorded. I confirmed this fact by analyzing the actual data on the WHO website. Also, I calculated the new deaths on a particular date (Daily Deaths) by utilizing this attribute value.

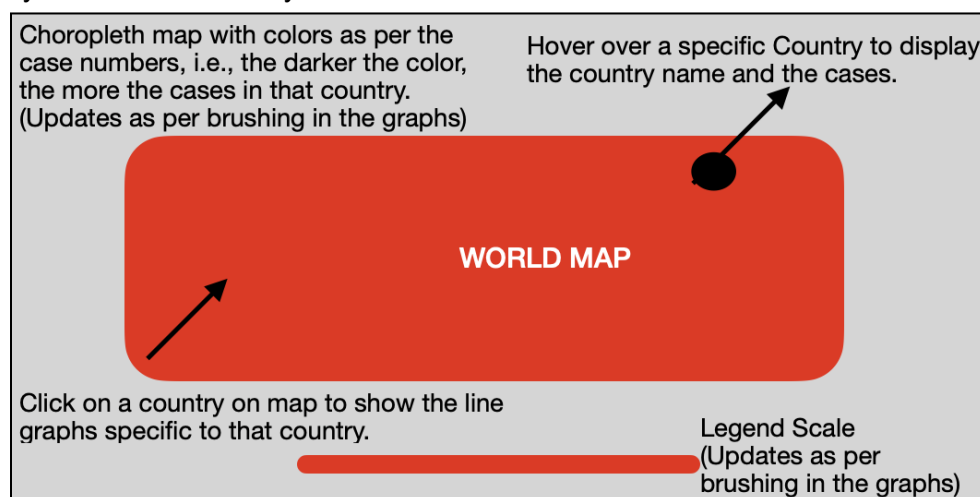
- I have also changed the Column names as follows to help with the code in d3:
  - Country/Region -> Country
  - Province/State -> Province

- **Interactive Visualization Design (Storyboard)**

I am planning to implement the following graphics in my interactive visualization:

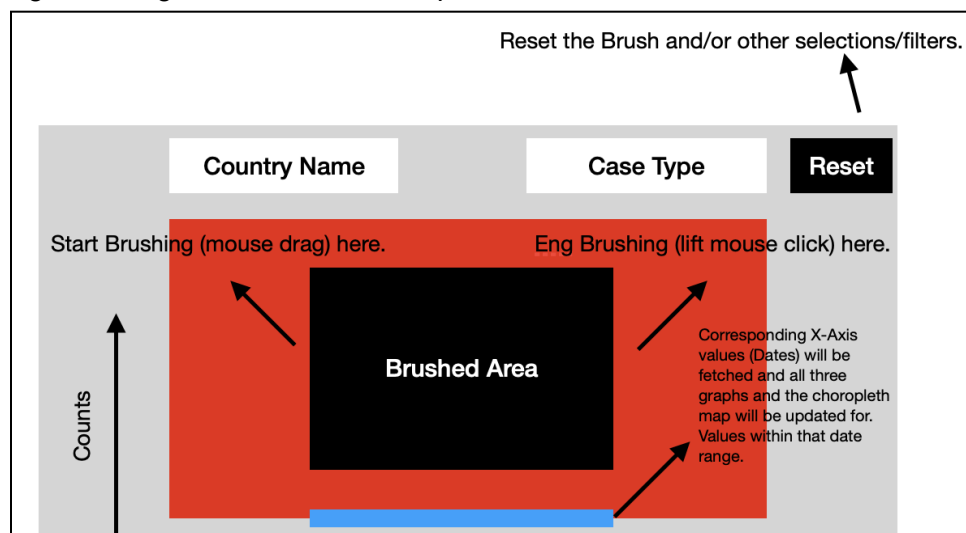
- **Choropleth Map:** I plan to use a Choropleth map that would be using a color scale indicating the total cumulative cases for the countries. The color scale would start from a white color to a dark red color, where darker color indicates a high number of cases. Also, a legend for the color scale would be included.
  - **User's Question Answered:** Generally, the sole purpose of visualization of a pandemic is to represent the spread of the disease. Accordingly, many users would be primarily interested in: How is the disease spread across the globe? OR Where is the spread of the disease more? To answer this question, the choropleth map would provide information regarding the spread of the design at a glance.
  - **Design Choice Rationale:** Although this user question can be answered using many other graphics (like a bar chart showing the total cases for every country), it would be difficult to compare different countries as they will have to remember certain information for comparison. However, a choropleth map provides the user with the necessary information at a glance and helps the user understand the spread of the disease in a pictorial format which improves perception.
  - **Interactions:**
    - **Tooltip:** I plan to display a tooltip when a user hovers over a country which displays the name of the country and the total cases for that country.
    - **Brushing and Linking:** I plan to update the choropleth map as per the date range selected or brushed by the user in the line graphs (next page). Also, on clicking a country on the map, the line graphs will be filtered to show the data corresponding to it.

The map layout and functionality would be as follows:

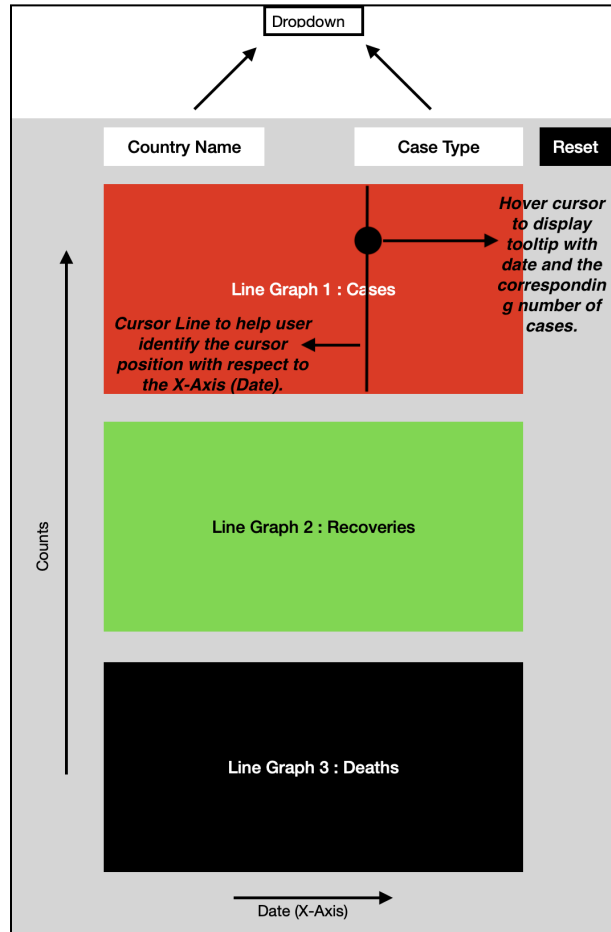


- **Line Graphs:** I plan to use line graphs to represent each of: Cases, Recoveries, and Deaths. The X-axis would be the date axis and the Y-axis will be the total case count.
  - **User's Question Answered:** One of the major questions that a user might have is : “What is the trend in the case number? Are they increasing?” OR “Where exactly are we in this Pandemic?” OR “How badly is one country affected compared to another country?”. Such questions can be answered using the Line Graphs that show trend lines.
  - **Design Choice Rationale:** As this is a time series data, a line graph would be a good choice for this visualization. It helps the user to decode the trends in the case numbers.
  - **Interactions:**
    - **Brushing and Linking:** I plan to use the brushing and linking feature to allow the user to select a date range from a graph and the graph updates accordingly.
    - **Dynamic Queries:**
      - **Country Dropdown:** I plan to include a dropdown with a list of countries so that the user can select single/multiple countries and filter the graphs accordingly for comparison.
      - **Type Dropdown:** I plan to include a dropdown with an option to select “Cumulative” of “Daily” case types as it would enable the user to select the type of trends they wish to analyze and the graphs would update accordingly.
      - **Tooltip:** I plan to implement a tooltip displaying the date and the number of cases for the user's pointer's x-position in the graph.
      - **Cursorline:** I plan to implement a cursor line to make it easy for the user to find the value for the number of cases for a particular date as this cursor line will be vertical and intersecting with the date axis. It will move as per the user's mouse position during mouse hover on the graph.

The Brushing & Linking feature would be implemented as follows:

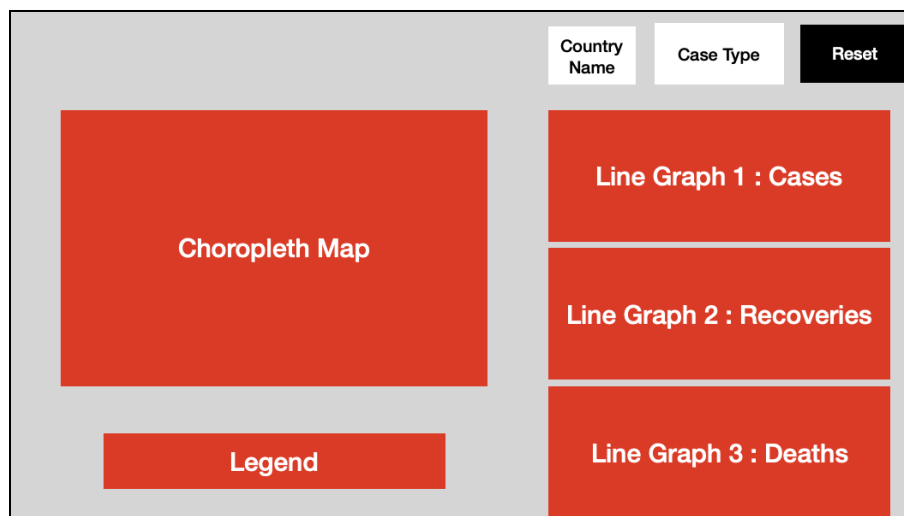


The layout of the graphs, dropdowns, buttons, and cursor line/tooltip would be as follows:



**Note:** I am using 3 graphs instead of one combined graph because of the significantly different Y-axis scales of the three graphs (deaths significantly lower than the cases) which does not make it effective. Also, as there are 3 graphs, there is space to compare data of different countries as compared to having just one graph.

The overall layout of the HTML Webpage would be as follows:



## **Part-2 Post-Implementation Summary**

**Summary:** I implemented the visualization using d3.js and HTML/CSS technologies. I designed the webpage to be divided into two halves with the left one having the choropleth map and the right one having the three graphs, dropdowns, and a button. There's a color scale legend below the choropleth map. I gave the visualization a title : "Covid-19 Tracker (World)". The visualization supports the following interaction functionalities:

- **Interactive Visualization Functionality:**
  - **Choropleth Map:**
    - Displays color intensity based on the case count
    - Hover over any particular country to highlight its border and display a tooltip
    - Brush over the line graphs to select a date range and the map updates accordingly
    - Click over any country of the map to display line graphs pertaining to the selected country on the right
  - **Line Graphs:**
    - Hover over the graphs to display a cursorline
    - Hover over the graphs to display a tooltip for a particular date
    - Brush over the graph to select a date range and filter all the three graphs (and the choropleth map) accordingly for the selected date range
  - **Country Name Dropdown:**
    - Select a Country Name from the dropdown to update all the three graphs accordingly
    - Select multiple country names by pressing and holding Command/CTRL key and selecting the countries from the dropdown to update all three graphs with one line for each selected country
  - **Type Dropdown:**
    - Select a Type: "Cumulative" or "Daily" to filter the graph to show the lines as per the selection
  - **Reset Button:**
    - Press this button to reset the brushing/filtering done in the line graphs & the choropleth map
- **Modifications:** There were no significant modifications performed compared to the original specifications.
- **Development Process & Timeline:**
  - It took about 38-40 hours to complete, starting from dataset selection to the final implementation and testing.

- A considerable portion of the timeline went to the pre-processing and calculation of implied attributes (daily values) tasks.
- The major portion of the timeline went into the implementation of the graphs; more specifically, the brushing over the graphs to select the date range and linking it appropriately to the complete visualization was the most time consuming and difficult part to implement
- Also, the functionality to compare trends for multiple countries in the line graphs accounted for a significant portion of the timeframe
- **Process:**
  1. Selected a dataset
  2. Completed the Part-1 of this documentation
  3. Data Cleaning & Pre-processing
  4. Implemented the Choropleth map & tooltip
  5. Implemented the Legend Scale
  6. Implemented the three graphs & tooltip
  7. Added filter dropdowns (Country Name, Type) & Linked them to graphs
  8. Implemented the brushing functionality over the graphs
  9. Implemented linking functionality to update all three graphs and the map
  10. Implemented reset brush button to reset the visualization to initial state
  11. Implemented the webpage design using HTML/CSS and organized the visualization components
  12. Completed the documentation (Part-2)
- **Procedure to Run the Code (usual d3.js code procedure):**
  1. Turn on the localhost python server via Terminal
  2. Open Browser (Google Chrome preferred for smooth operations) and enter the localhost URL
  3. Navigate to the code folder, and select index.html file
  4. It could take a couple of seconds for the Map to load
- **Troubleshooting:**
  - If the map doesn't load or if there is just one single color appearing across the whole choropleth map, RELOAD the page. Sometimes, based on the browser, the rendering of this large dataset takes a couple of seconds or a RELOAD.
  - Due to the large dataset size, the performance of the code may vary based on the machine. For M2 Pro Mac (my machine), it works smoothly.
- **As a second option**, I have also hosted this Assignment-3 on the web, so you may directly access it online at the following URL:  
<https://cs333-a3.glitch.me/>  
*(It could take a couple of seconds for the map to load)*
- **As a third option**, I have also attached a GIF link in the submission which demonstrates the complete functionality of the assignment. Kindly find the link to the GIF below:  
