CSE 578 Data Visualization

VAST Challenge 2010 MC2 - Characterization of Pandemic Spread

I. INTRODUCTION

This is the report for the Data Visualization project, CSE-578. We collaborated on the VAST Challenge 2010 Mini Challenge-2 which is Characterization of Pandemic Spread, as a team. we are provided with hospital admittance and death records data for cities involved in the epidemic. Initially, we have done some data preprocessing to the given data since it was huge and accordingly obtained the required attributes to build different visualizations. As per the aim, we have analyzed and designed the visualization which are easily understood by the user and can be utilized in future when there might be a next epidemic outbreak. We worked on the development of five different visualizations and each had a great significance.

II. PROJECT IMPLEMENTATION

To answer the given mini challenge MC1 and MC2 we came up with five distinct visualizations. They are

1. Multi-Line Chart
2. Stacked Bar Chart
3. Density Estimation Chart
4. TreeMap
5. Bubble Chart

A. Multi-Line Chart

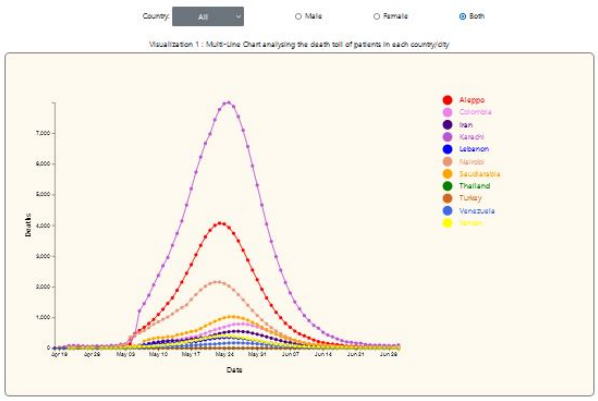

Fig. 1. Vis-1 Multi-Line Chart

In this visualization, we calculated the number of patients that died on a particular day and represented it using a line chart. Initially we have done some data processing to calculate the count of patients who died on each day from the provided dataset. Above method is applied to all countries/cities to analyse the issue better. This is a multi-line chart that is used to identify global trends and anomalies. Color hue is used to represent the countries/cities in the given dataset.

Interactions:

When user hovers over a particular chart, the death count on a single day in that particular country/city is available. We also provided a drop down in the navigation bar where user can select individual country/city or all and an on-click Gender selection option on navigation bar.

B. Stacked bar Chart

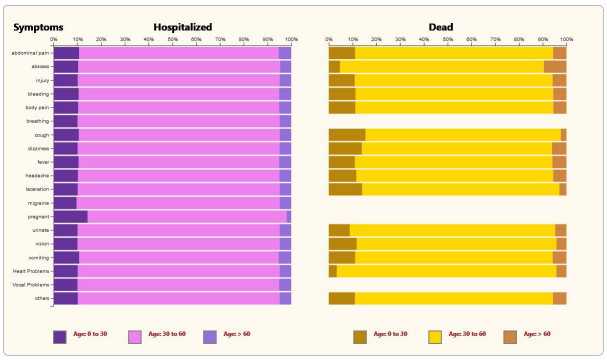

Fig. 2. Vis-2 Stacked bar graph

This visualization is developed so that it has stacks of bars which analyzes the percentage of people who has been hospitalized and dead with different symptoms. For more detailed analyses we have divided the data into three age groups. We did the data pre-processing to obtain the count and percentage for male, female and both.
There are three age categories such as 0 - 30, 30-60 and above 60 years. The count in each category is calculated and plotted. The left chart represents the data of hospitalized people and the right shows the details of dead.

Interactions:

When user hovers across different age groups the percentage and count of patients in that particular category that fall under a particular symptom is available. We also provided a drop down in the navigation bar where user can select individual country/city or all and an on-click Gender selection option on navigation bar.

C. Density Estimation Chart

This visualization consists of bars which represents the number of people hospitalized each day in particular country.

The data is collected by pre-processing and also obtained the mean of the count of the number of people who died and were hospitalized in the previous 6 days for each country/city and each gender.
In the visualization the green line shows the mean of hospitalized patients and the red line represents the mean of the patients who died after being hospitalized over the previous 6 days.

Fig. 3. Vis-3 Density Estimation Chart

Interactions:

When user hovers over the green line it displays the count of patients hospitalized and when user hover over the red line it displays the count of patients who died after being hospitalized. We also provided a drop down in the navigation bar where user can select individual country/city or all and an on-click Gender selection option on navigation bar.

D. Treemap

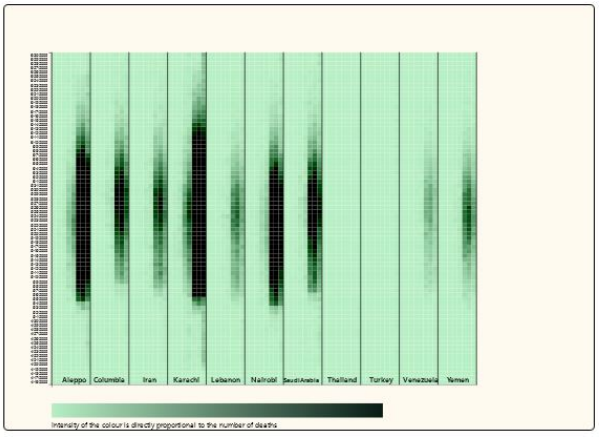

Fig. 4. Vis-4 TreeMap (Death)

In this visualization, we represented the number of patients affected by the symptoms in each country/city in format of treemap. We initially pre-processed the given data to get the number of people who died and the number of people hospitalized on each date for each country/city.

This chart is divided into 11 sections which represents each country/city. The top eight symptoms are picked to represented in each country/city. Each pixel in a section represents each symptom and the colour gradient is for the count of the dead or hospitalized patients. It is darker for the count which is high and vice-versa.

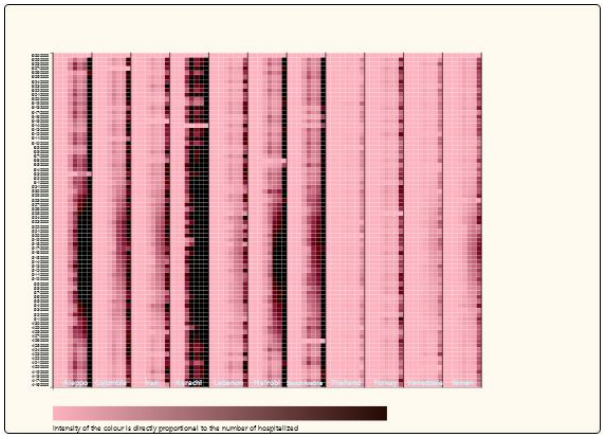

Fig. 5. Vis-4 TreeMap (Hospitalized)

Interactions:

When you hover the mouse on a pixel it displays the death or hospitalized count, the country/city selected, symptom selected, and the date. There is a drop-down to select either dead or hospitalized patients.

E. Bubble Chart - Innovative

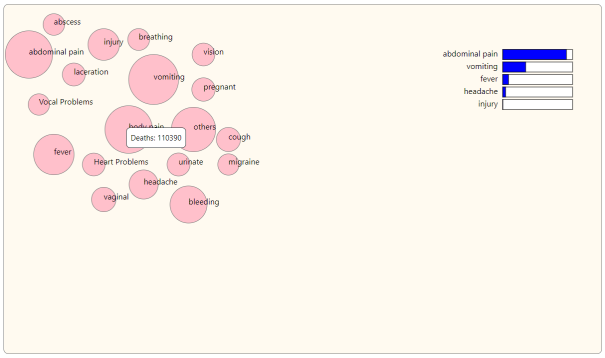

Fig. 6. Vis-5 Bubble chart

This visualization consists of bubbles of different symptoms and the size of each bubble depends on the number of patients died due to that particular symptom, the bigger size bubble says that the count of patients with that symptoms is the most among all other symptoms and vice-versa.

when user clicks on each bubble, there is bar chart analysis available on the right side of the page which shows the count of patients died with all other symptoms along with that particular symptom

Interactions:

When user hovers over the bubble it shows the count of patients who died and when user clicks on that bubble, a bar chart showing the count of dead patients with the common symptoms is displayed along with the count on hover.

III. Results of the Visualizations

The aim of our project is to answer all the MC questions using the visualizations. The results of all the visualizations are provided below.

A. Multi-Line Chart

Through Multiline chart we can analyse which country/city among the all has most deaths during the pandemic and we found that Karachi had the most deaths.
Observed Anomalies: We have found some anomalies with the country As the death toll for Thailand and Turkey is too low, the graph for these two countries is almost a horizontal line which is zero.

B. Stacked bar Chart

Through the stacked bar chart, we are trying to analyze the symptoms among the dead and hospitalized among different age groups.
Observed Anomalies: After plotting the chart we found that the most of the countries had the spread out with same percentage ratio in different symptoms among the age groups. This says that the age groups are equally effected with different symptoms.

C. Density Estimation Chart

Through this chart we can analyze how the death rate is being distributed among the hospitalized people.
Observed Anomalies: For countries like Thailand and Turkey the death toll is too low every day, so taking the 6 days mean for that low values is resulting in much smaller values, hence the flat red line.

D. Treemap

This chart analyses the symptoms in hospitalized and dead in different countries
Observed Anomalies: For countries like Thailand and Turkey, the death toll is too low every day. Hence, the low luminance for these two countries.

E. Bubble Chart - Innovative

This visualization show how each symptom has commonalities.
Observed Anomalies: For some symptoms like vision, cough, pregnancy, etc., the common symptom count shows zero. This is due to the very less count in the dead patients with that particular symptom.

IV. Contributions to the project

1) In this project, I initially worked on finalizing the visualizations based on the provided data and determined the attributes required for different visualizations
2) Scheduled the group meetings to discuss the visualizations, design of the application and worked on division of the work among the teammates.
3) Worked on the idea for the innovative chart, data processing for the chart, design and development of visualization-5 using D3.js.
4) Worked on merging the visualizations with the application using HTML,CSS and writing the detail explanation of project in the report

V. Lessons Learnt

1) I understood the structuring on D3.js and learnt the scope of D3.js
2) Learned how to utilize feature extraction to extract interesting and desired trends from a dataset and use them in data visualization.
3) We spent hours extracting the pre-processed data file from the code since the given dataset was too huge. For a quicker python implementation, we leveraged Google Colab and many cores in the local environment.
4) Learned how to combine the code that we each wrote separately into a single, fully functional project.
5) Learned how to choose the best up-on charts for the supplied data in order to derive useful conclusions.

VI. Team Members

1) Vineethkrishna Vemireddy
2) Aishwarya Reddy Dwaram
3) Avinash Senthil Kumaran
4) Madhura Ganga
5) Sindhu Sree Aita
6) Sucharith Reddy Desireddy

References

[1] Whiting, M., Haack, J., and Varley, C. Creating realistic, scenariobased synthetic data for test and evaluation of information analytics software. Proc. of BELIV'08, ACM (2008)

[2] G. Grinstein, S. Konecni, C. Plaisant, J. Scholtz and M. Whiting, "VAST 2010 Challenge: Arms dealings and pandemics,"2010 IEEE Symposium on Visual Analytics Science and Technology, Salt Lake City, UT, 2010, pp. 263-264, doi: 10.1109/VAST.2010.5649054.Conference Location:El Paso, Texas USA

[3] Fernanda, B., Martin, W., Jesse, K.,"ManyEyes: a Site for Visualization at Internet Scale", IEEE Transactions on Visualization and Computer Graphics, Irvine, 10.1109/TVCG.2007.70577, Nov. 2007.

[4] Enamul, H., Maneesh, A., "Visual Style and Structure of D3 Visualizations," IEEE Trans. Vis. Comput. Graphics, vol. 26, no. 1, pp. 1236–1245, Jan. 2020.

[5] https://observablehq.com/@d3/bubble-chart