

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY,
BANGALORE

Datathon 7

Authors:

Saikiran Reddy

IMT2017030

Sanjay

IMT2017037

Khalid

IMT2017028

Rohith

IMT2017044

Supervisor:

Prof. Jaya Sreevalsan Nair

November 26, 2020

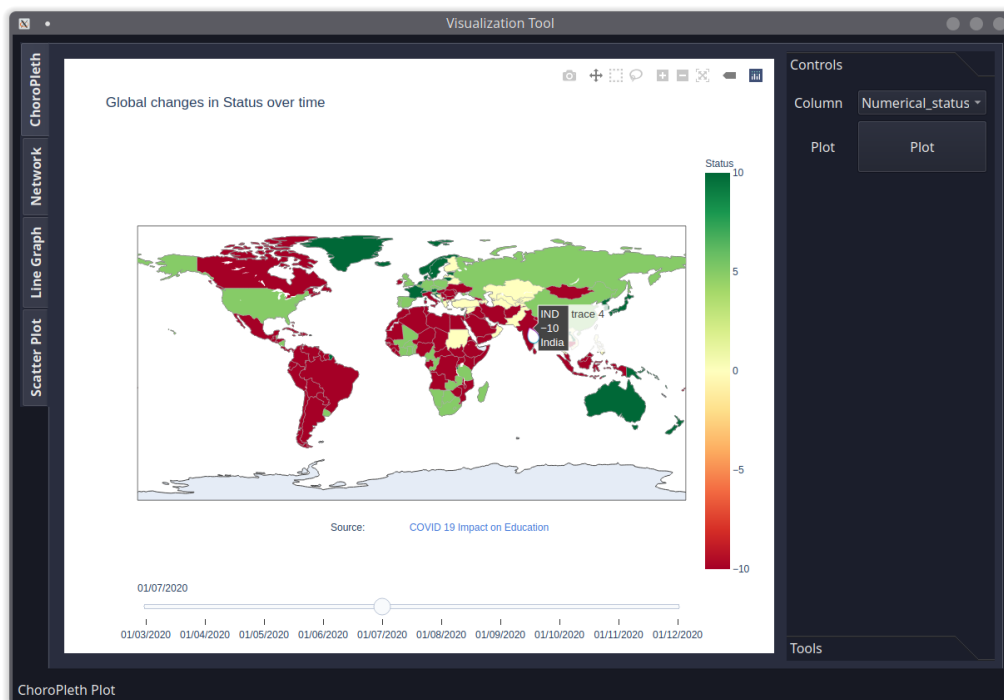


Figure 1: Choropleth Plot in the GUI Application.

Contents

1	Introduction	4
1.1	Problem Statement	4
2	Methods	5
2.1	Visualizations tools/API	5
3	Choropleth Maps	6
3.1	Preprocessing	6
3.2	Technology Used	6
4	Network Plots	7
4.1	Pre-Processing	7
4.2	Technology Used	7
4.3	Seriation	7
5	GUI	9
6	Questions	15
7	Conclusion	19
	References	19

List of Figures

1	Choropleth Plot in the GUI Application.	1
1.1	Splash Screen of our GUI Application.	4
3.1	Choropleth Plot	6
4.1	Network Plot with Negative Correlation	7
4.2	Network Plot with Cities as Nodes	8
4.3	Network Plot	8
5.1	Splash Screen of the Application.	9
5.2	Start Up Screen of the Application.	10
5.3	Open File Dialog of the Application.	10
5.4	Choropleth Visualisation of the Application.	11
5.5	Network Visualisation of the Application.	11
5.6	Exporting Visualisation in the Application.	12
5.7	Line Graph and Scatter Plot Visualisation of the Application.	12
5.8	Interactiveness of the Application.	13
5.9	Menu Bar of the Application.	13
5.10	Status Bar of the Application.	14
5.11	Line Graph and Scatter Plot Visualisation of the Application.	14
5.12	About Dialog of the Application.	14
6.1	Different Visualizations of the Application.	15
6.2	Choropleth Plot	16
6.3	Choropleth Plot	16
6.4	Choropleth Plot	16
6.5	Choropleth Plot	17
6.6	Network Plot	17
6.7	Network Plot	18
6.8	Network Plot with Negative Correlation	18

1 Introduction

1.1 Problem Statement

Dataset: Tabular datasets source [here](#) and [here](#). It contains 15,700 rows, 16 columns with much different information from various countries.

The task entails building a visual analytical tool which has a graphical user interface (GUI), where the user can manipulate the data as much as possible.

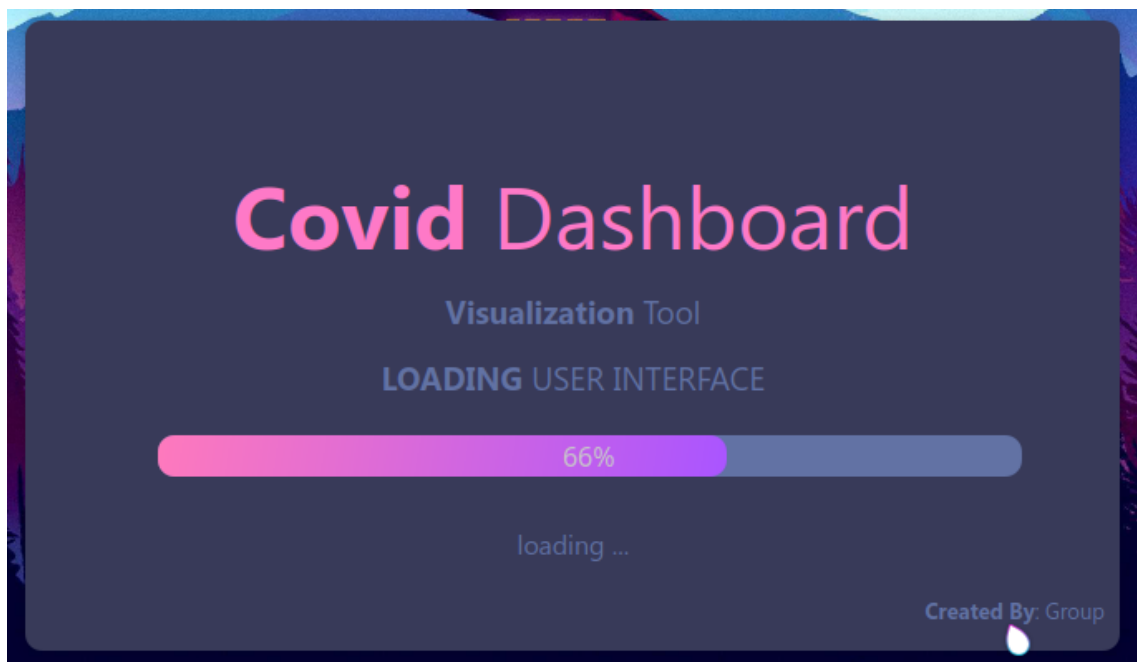


Figure 1.1: Splash Screen of our GUI Application.

2 Methods

Khalid implemented the choropleth maps for this dataset. Sanjay implemented the Network Plots. Rohith implemented Seriation, Line Graph and preprocessing of datasets and derived inferences. Saikiran implemented the GUI and interactivity of application with the dataset and Scatter Plot of this dataset.

2.1 Visualizations tools/API

The following visualizations tools/API were used to implement the seriation methods.

1. Plotly Choropleth Maps[\[1\]](#) (Khalid)
2. Networkx Plots (Sanjay)
3. Seriation (Rohith)
4. Pyside2 (Qt for Python)[\[2\]](#) (Saikiran)
5. Matplotlib (Saikiran, Rohith)

Details about the Map implementation on dataset is explained in later chapters.

3 Choropleth Maps

We use a choropleth plot to analyze the impact of covid-19 on schools.

3.1 Preprocessing

The dataset we used was a time series data set, this data set consisted of date, location and school statues which was one of Fully open(10), Partially Open(5), Closed due to break(0), Closed due to covid(-10), where the numbers in the brackets indicate the value we gave to each of the status. We used a color map on this value and displayed the changes over time. This tool allows interaction for users to slide through time through the slider provided.

3.2 Technology Used

For this visualization We have used plotly's choropleth map function. To make it interactive We have made a list of dictionaries for each time stamp that indicate the map features. In the end We have used plotly's slider argument in the plot function to visualize each of the maps in the list.

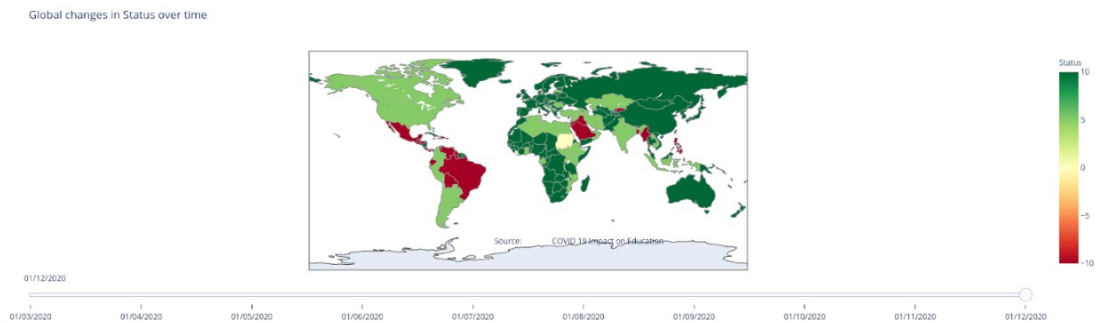


Figure 3.1: Choropleth Plot

4 Network Plots

4.1 Pre-Processing

There are some null values in the dataset and they are represented by ‘.’. So We replaced it with NaN and then dropped all NaN values. For visualizing between cities or dates, We have done some extra pre processing. Like there are separate columns for year, month, and date. So We combined all the three columns to get a new column ‘date’ and We dropped those three columns. And to convert city id to city names, We used the file ‘GeoIDs-City.csv’. And then We created a new data frame with cities as columns and dates as rows with one of the columns (from the original data frame) as data.

4.2 Technology Used

For plotting the node link, We used Networkx. Depending upon the type of visualization(like between cities, dates, all columns), We calculated the correlation network for that data using the function corr. Then We relabeled the nodes using the function relabel nodes in Networkx. Then We used the graph draw_networkx_edges and draw_networkx_nodes in networkx to draw edges and nodes respectively. This function takes inputs such as file name, column name(data for visualizing between dates or cities), correlation direction, minimum correlation value, and the size of the dataframe to consider. This also takes a variable for specifying the type of the plot(like between cities, dates,and all columns).

4.3 Seriation

Given a correlation matrix and and a seriate Boolean, the ‘seriate_on’ function performs seriation on the given correlation matrix and outputs an optimal ordering of columns. After rearranging the columns, the function outputs a correlation matrix along with a heatmap.

Below Figure shows the plot of negative correlation between all columns in the dataset.

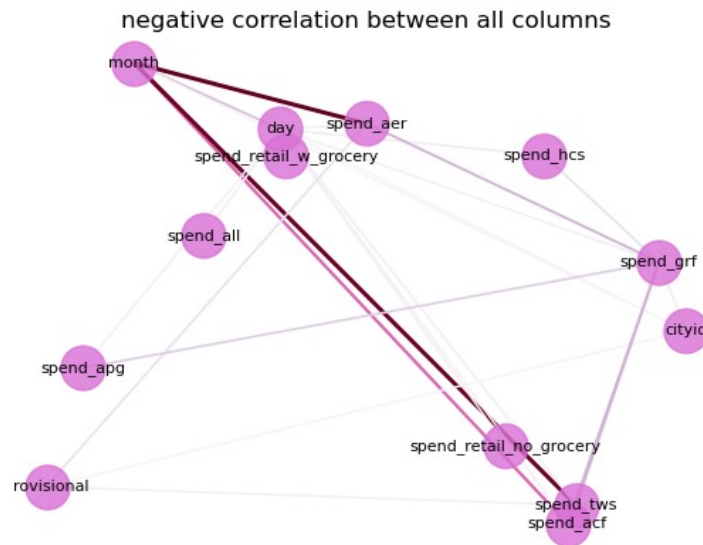


Figure 4.1: Network Plot with Negative Correlation

We also visualized correlation between different cities like in the below figure, We visualized the pay across the different cities.

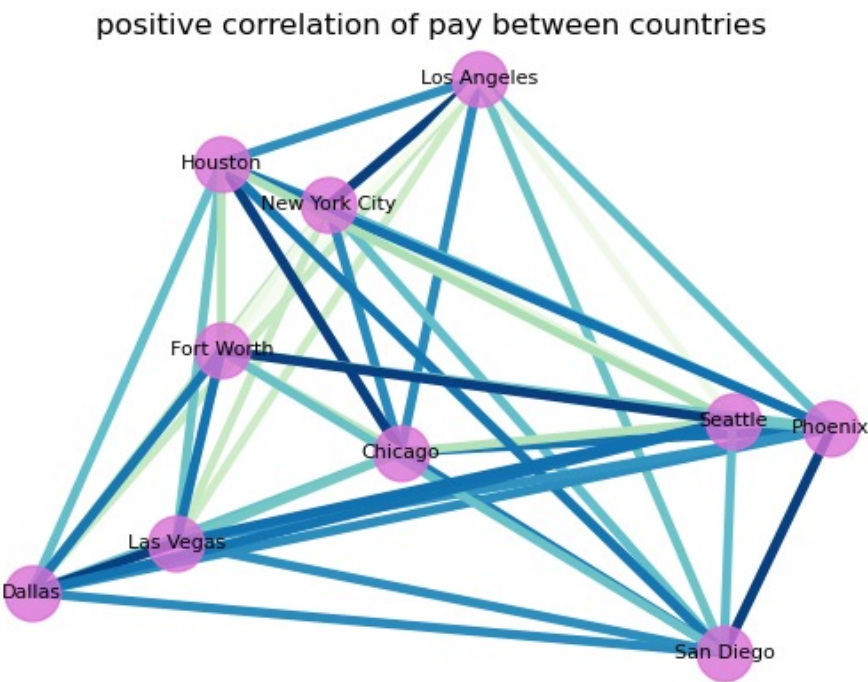


Figure 4.2: Network Plot with Cities as Nodes

Similarly in the below figure We visualized the amount spent in accommodation and food services across different dates.

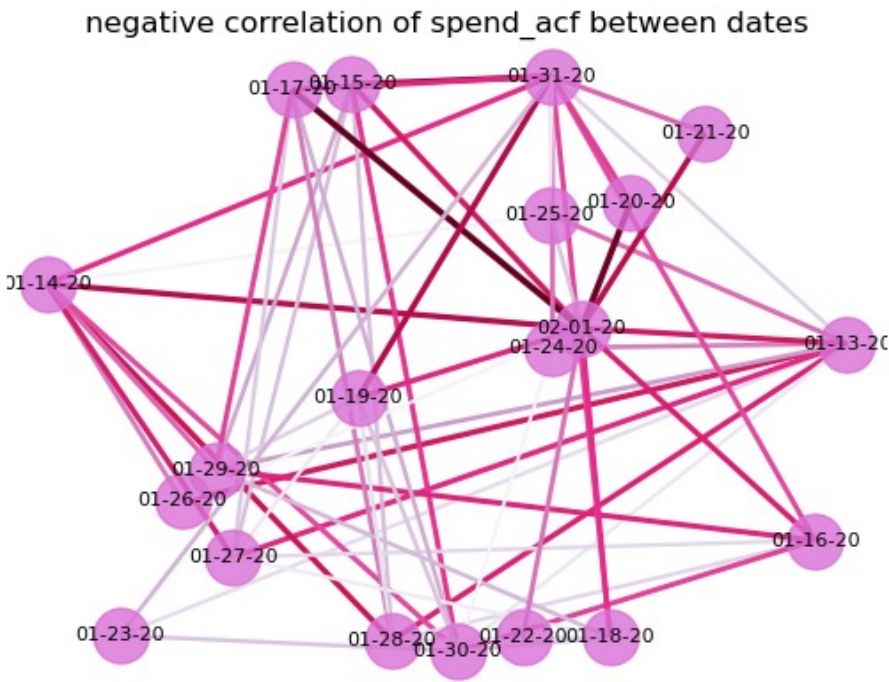


Figure 4.3: Network Plot

5 GUI

We have developed GUI with appropriate titles and aesthetic layouts with mind. We have Pyside2 (Qt for Python)[\[2\]](#) to make our GUI Application. It is completely interactive with native menu and status bar and a splash screen.

First the Application will display a Splash Screen showing app name and loading status with beautiful progress bar.

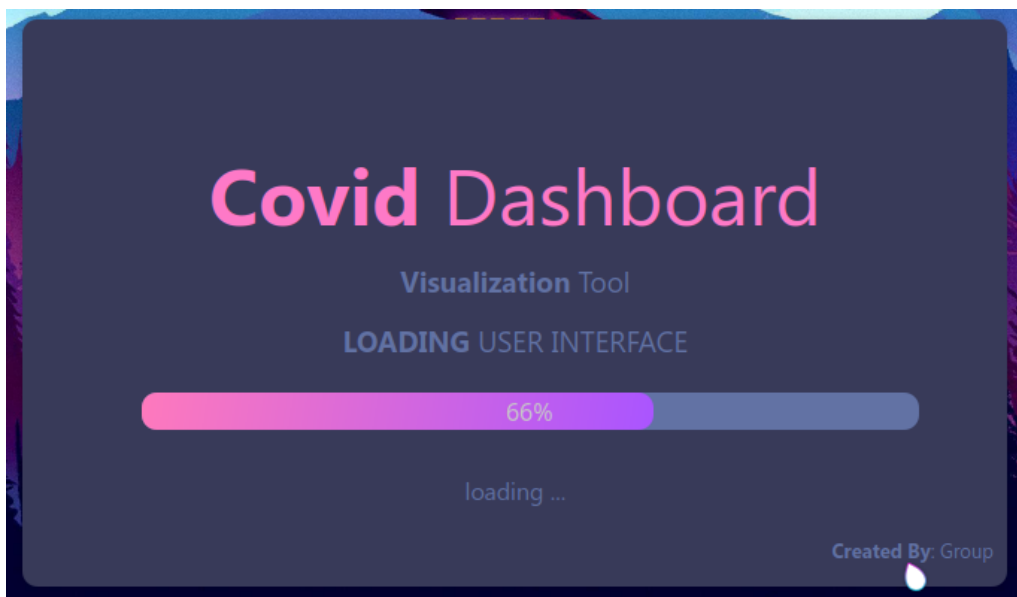


Figure 5.1: Splash Screen of the Application.

Then we can introduced to a startup dialog asking for dataset to visualize. We have provided a sample file which will used for demonstration purpose but any file with same generic format can be used to visualize with this tool.

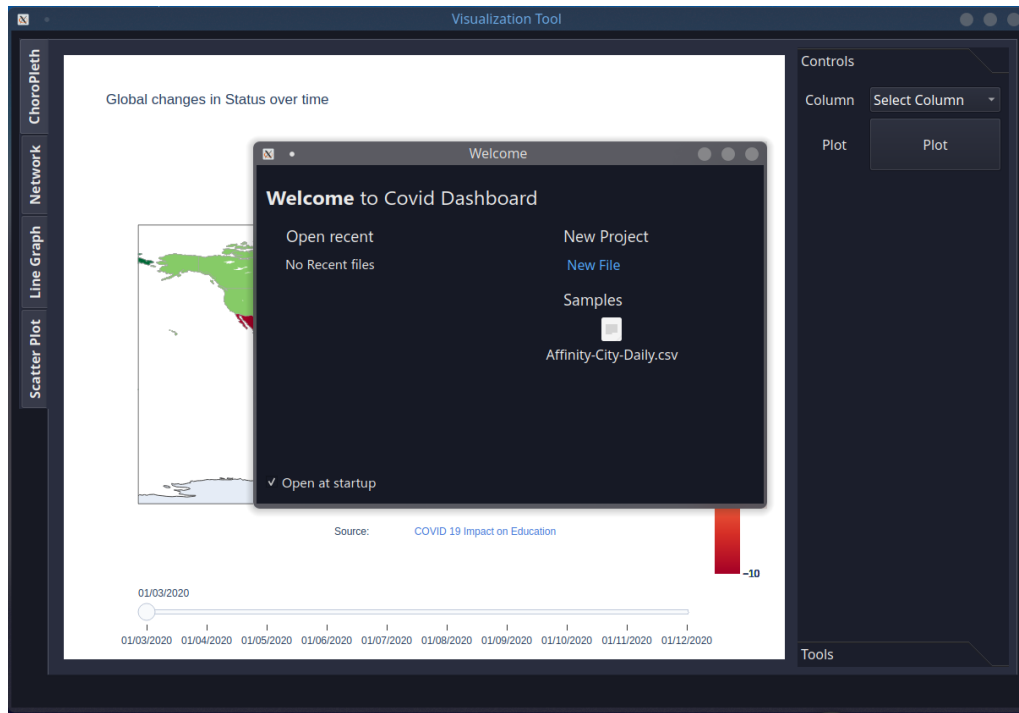


Figure 5.2: Start Up Screen of the Application.

When New file is pressed or Shortcut (Ctrl+O) is pressed or File-> New File File Dialog appears which accepts only CSV files which is then used for visualization.

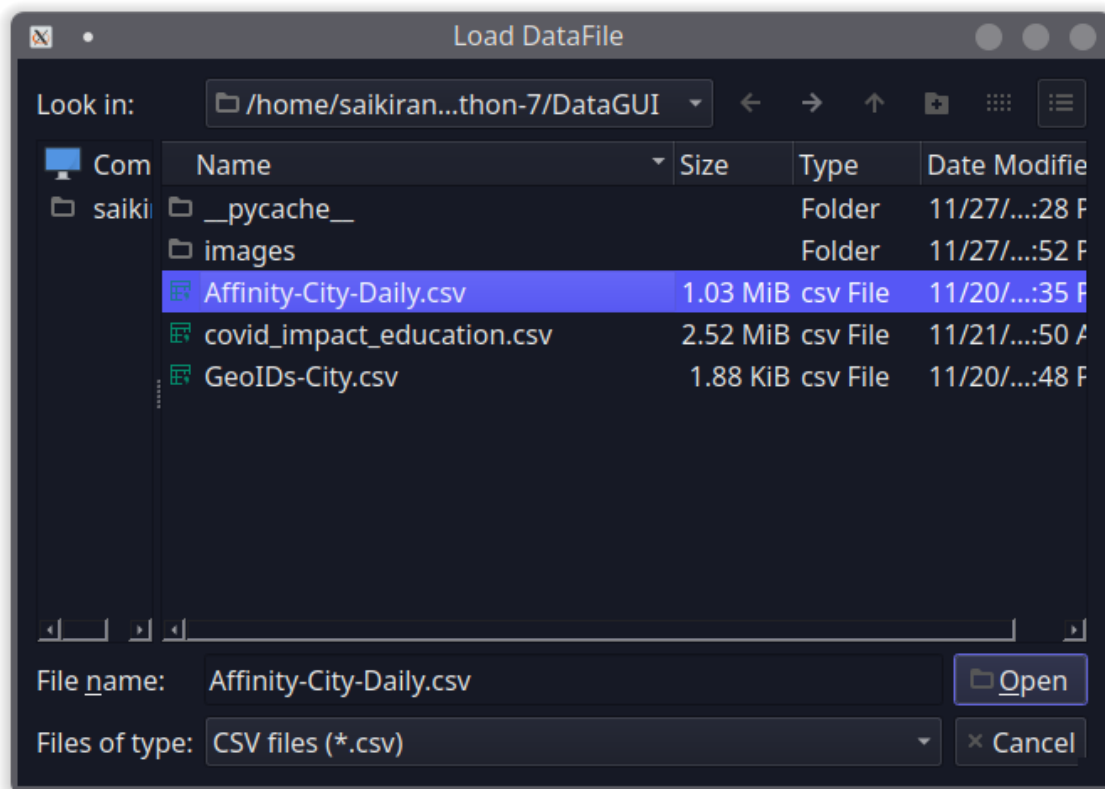


Figure 5.3: Open File Dialog of the Application.

After that we get Main Window with vertical tabs containing different Visualization with a toolbox at right with interactive controls. The ChoroPleth is displayed using QtWebEngine widget to display plotly html exported plot.

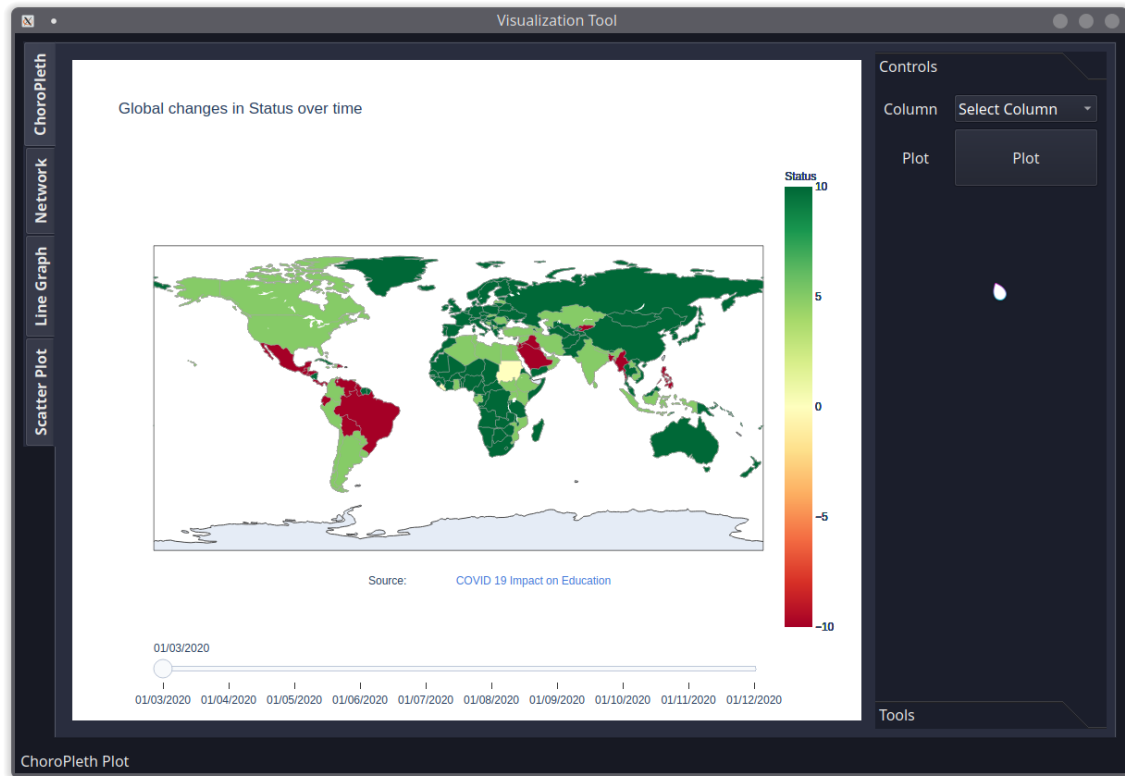


Figure 5.4: ChoroPleth Visualisation of the Application.

If we press Network tab we get Network Visualisation which as many interactive controls like we can choose correlation direction, its value and the attribute/Column we would like to use and Number of rows for clearer Plots. The Network and Other Plots are displayed using Matplotlib and Qt Canvas Widget.

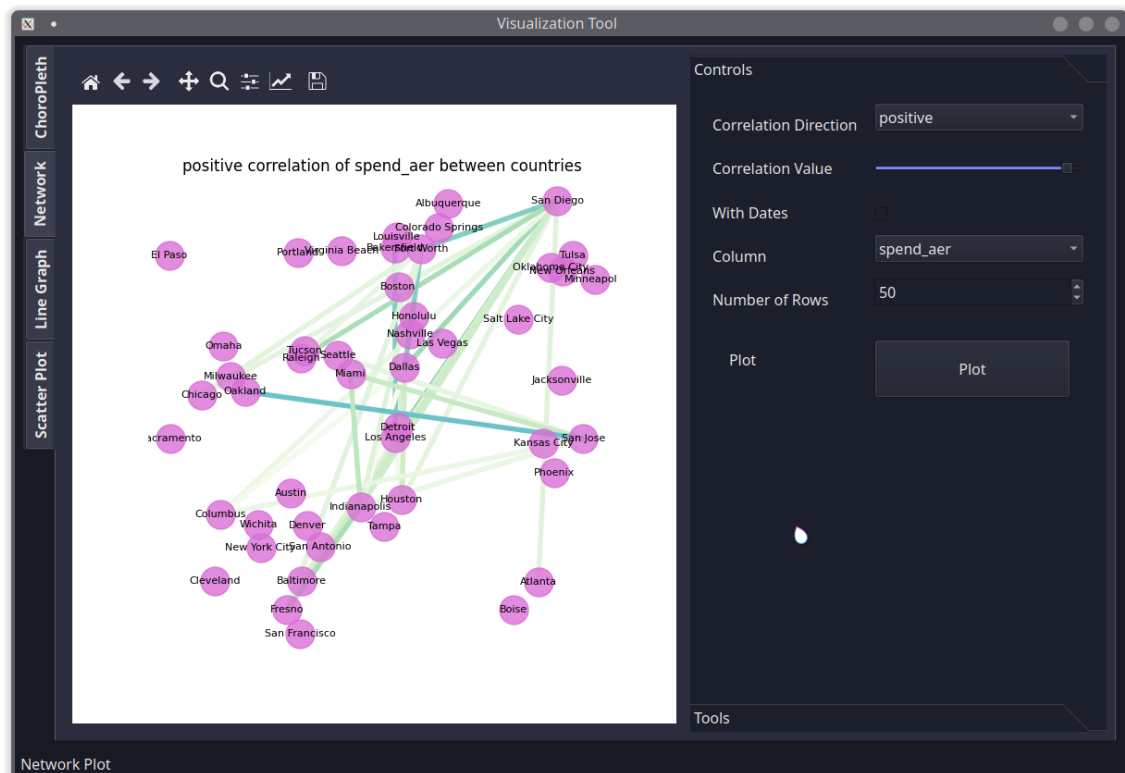


Figure 5.5: Network Visualisation of the Application.

We can also export our chosen Plot using export button by pressing Tools tab in toolbox.

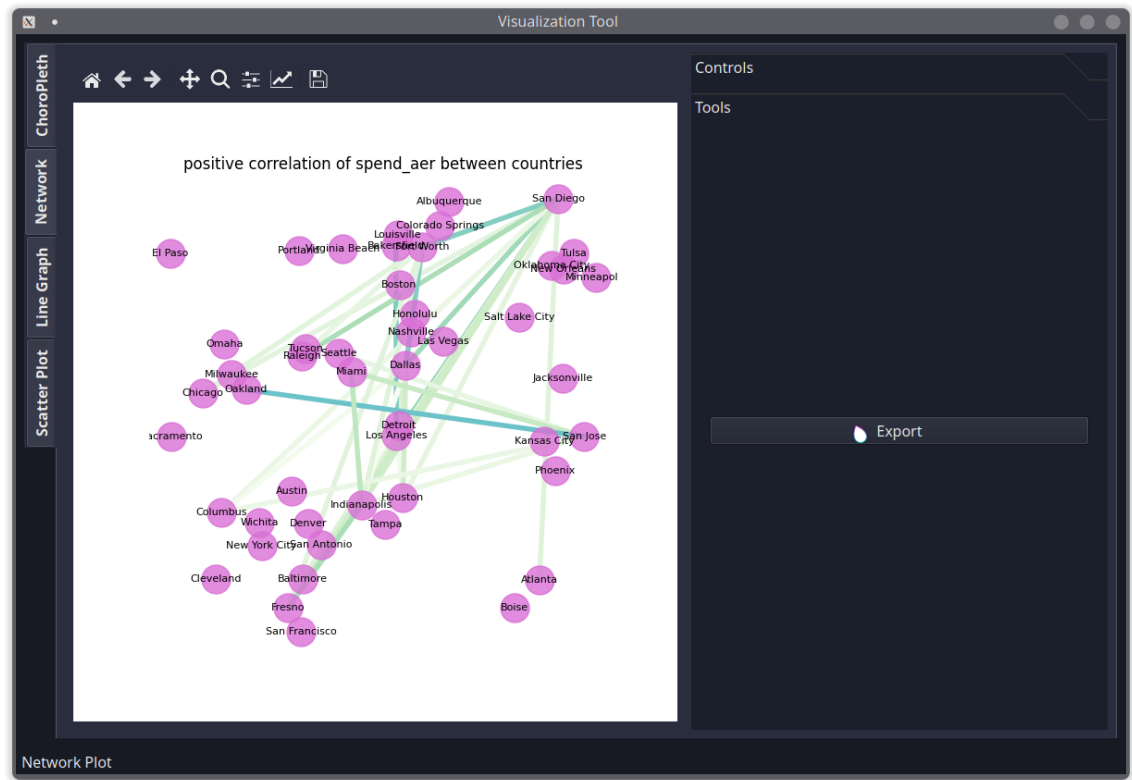
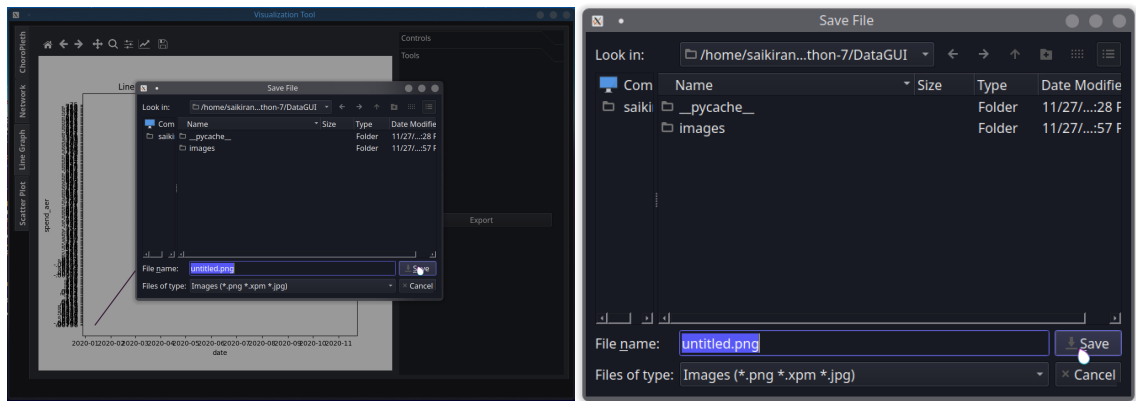


Figure 5.6: Exporting Visualisation in the Application.

When the export button is pressed we get a save file dialog which will allow us to save out plot in image format.



(a) Exporting Visualisation in the Application. (b) Save File Dialog in the Application.

Figure 5.7: Line Graph and Scatter Plot Visualisation of the Application.

The Application is interactive as in the graphs can be zoomed , modified and relevant information is obtained when hovered.

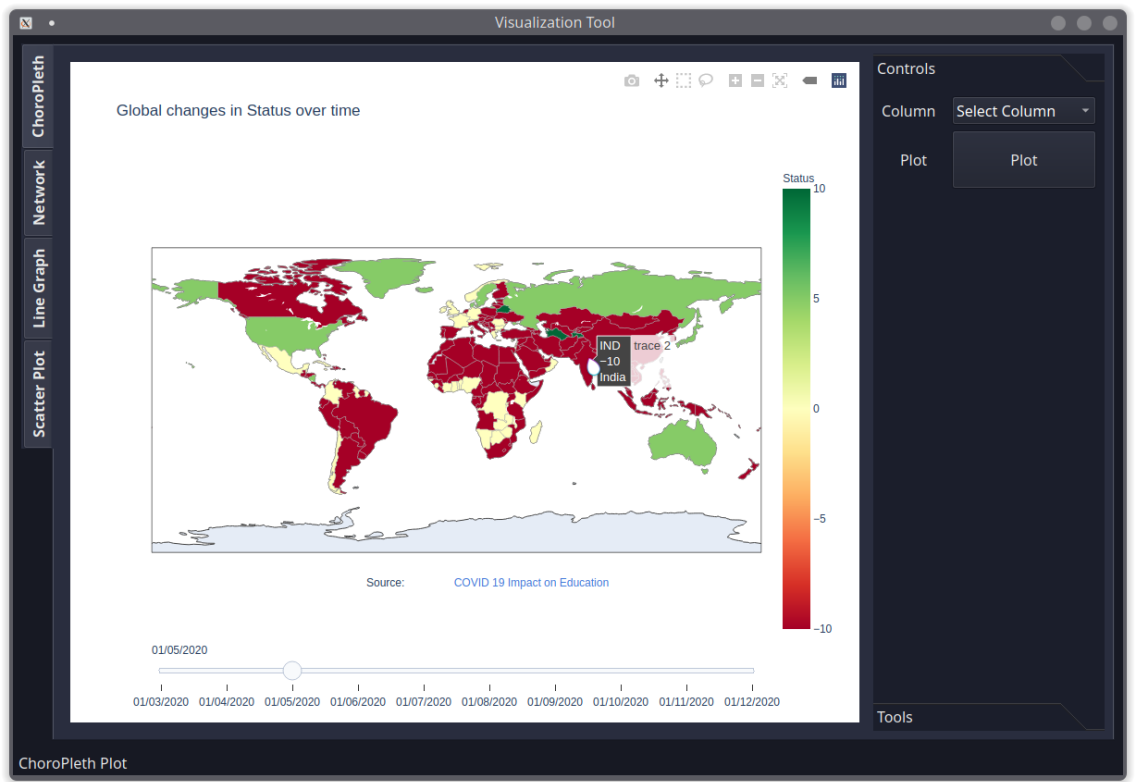


Figure 5.8: Interactiveness of the Application.

There is also Application Menu with standard options which cant be displayed as screenshots doesn't support system menus so only buttons are displayed which contains standard GUI actions like New File(Ctrl+O), Exit (Ctrl+Q) , show Startup Dialog and about.

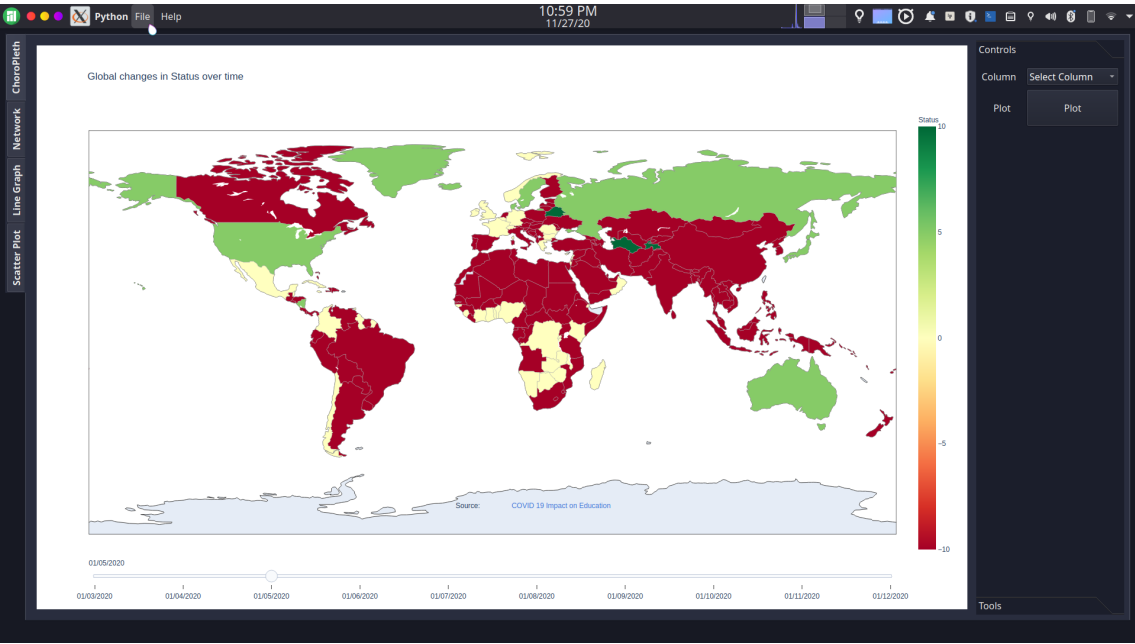


Figure 5.9: Menu Bar of the Application.

There is also a status bar at bottom of Application which displays useful info.

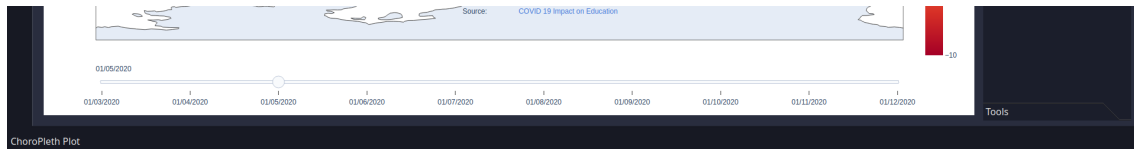
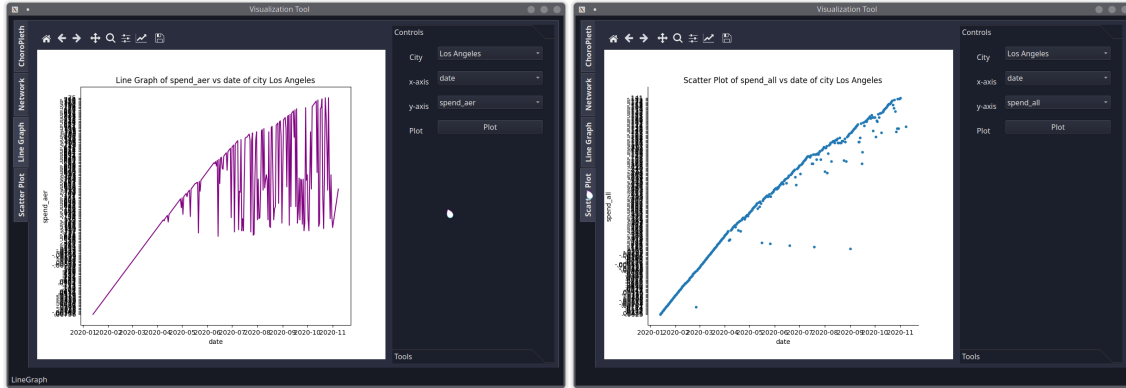


Figure 5.10: Status Bar of the Application.

The line and scatter plots are displayed similarly as network plot using matplotlib and Qt canvas.



(a) Line Graph Visualisation of the Application. (b) Scatter Plot Visualisation of the Application.

Figure 5.11: Line Graph and Scatter Plot Visualisation of the Application.

Finally a About Dialog is added which shows each Team Members individual Contributions.

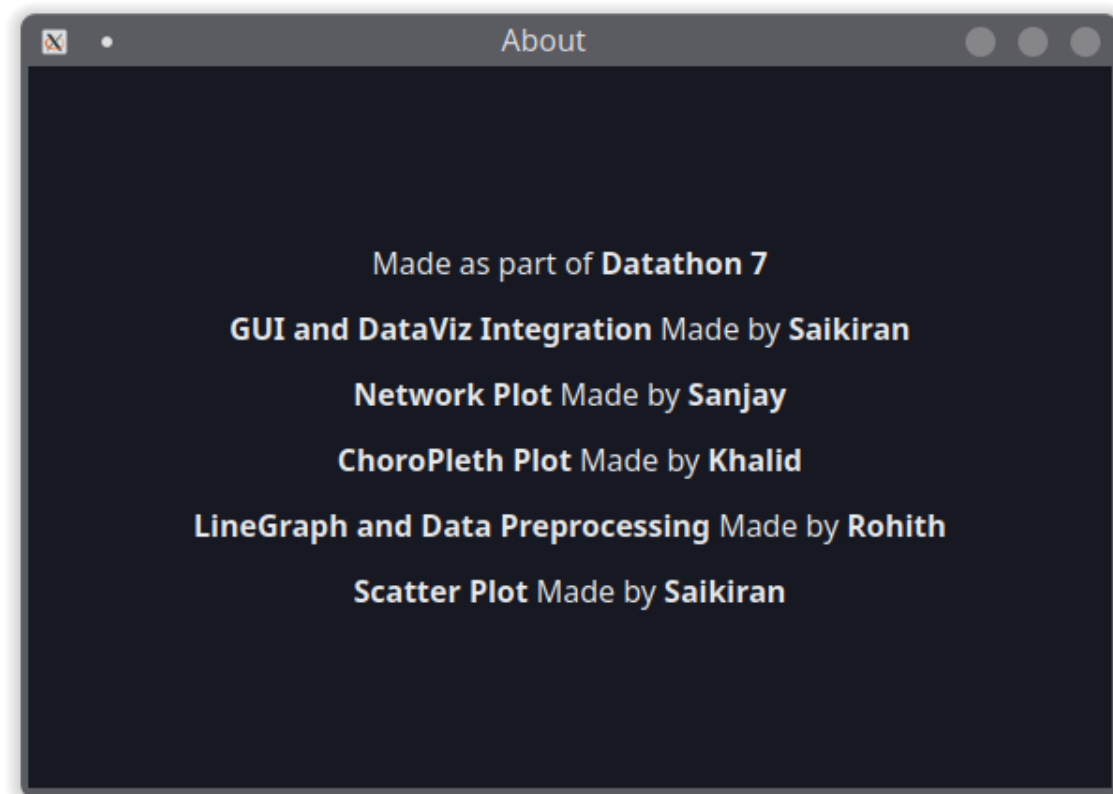
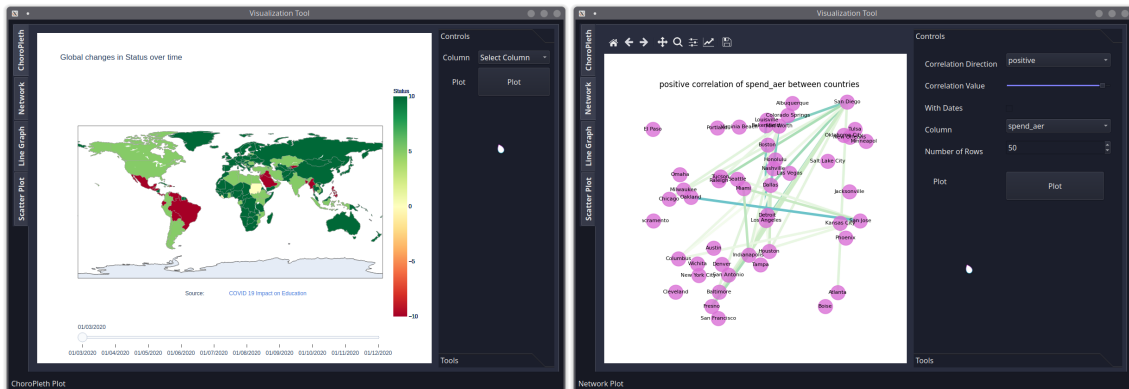


Figure 5.12: About Dialog of the Application.

6 Questions

1. which parts of the dataset were you able to use, and how have you been able to use ?

Ans: As discussed in Methods [chapter 2](#) we used entire two dataset and provided interactive controls to visualize the dataset.



(a) Choropleth Visualisation of the Application.

(b) Network Visualisation of the Application.

Figure 6.1: Different Visualizations of the Application.

2. . which methods did you choose, why, what technologies did you use for the visualizations?

Ans: This was explained in details in the [chapter 2](#) and visualization method was explained in details in their respective chapters which was Choropleth Map using plotly[1]library is explained in [chapter 3](#). The Network Plot uses Networkx library which is explained in [chapter 4](#) and seriation and GUI are explained in chapters [chapter 4](#) and [chapter 5](#).

3. What are my inferences?

Ans: Many interesting patterns were visible and interesting geospatial observations were made possible because of the choropleth graphs.

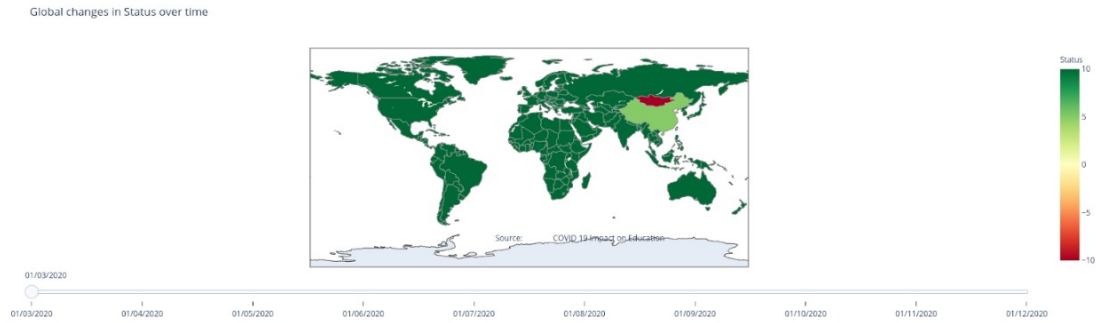


Figure 6.2: Choropleth Plot

The above plot is for the month of March 2020, We can see that most of the schools in the Western countries and middle Eastern countries are fully functional, whereas countries like china have their schools partially open which could be result of lifting the lockdown in certain areas of the country, which indicates their quick response to contain the spread of virus. But countries like Mongolia that are close to the origin of covid-19 virus, have completely shut down their schools due to covid-19.

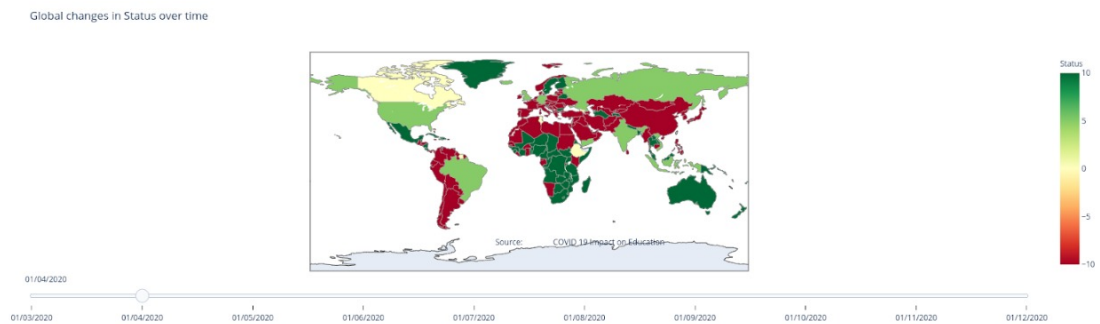


Figure 6.3: Choropleth Plot

The above plot shows the status of schools across country in the month of April 2020, It can be seen that there are more countries that have shutdown their schools, Including china this could be to contain the spread of virus as the number of cases in the month of April are much more than the number of cases in the month of March.

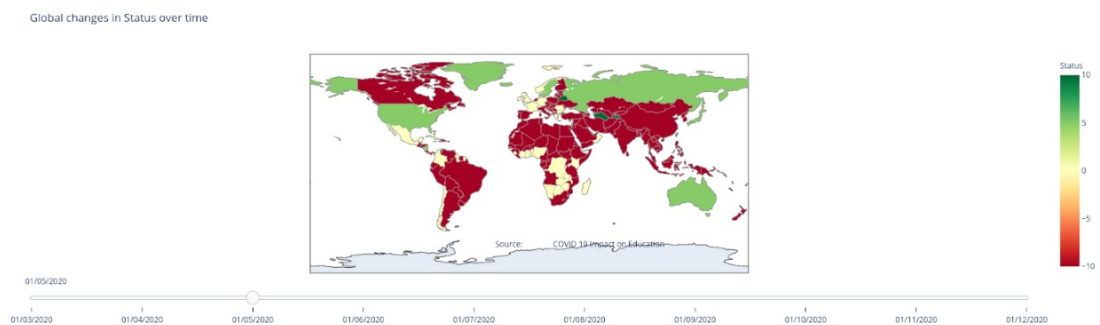


Figure 6.4: Choropleth Plot

The above plot shows the status of schools in the month of May 2020. Again the number of countries that are shutting down the schools to control the spread of the virus has increased. One interesting country is the US which has not shut down it's schools due to covid-19 till now, even

though their neighbours Canada have shut down their schools. This could either be due to the government not imposing lockdown.

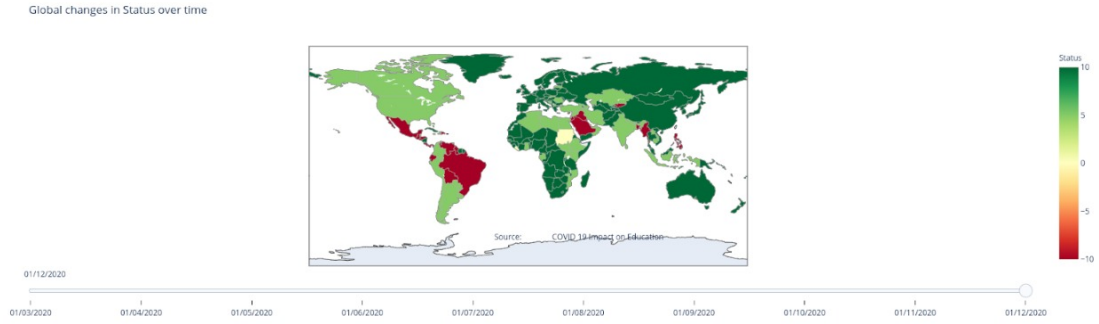


Figure 6.5: Choropleth Plot

The above plot indicates the status of schools in December 2020. This data shows that most of the countries have reopened their schools, the countries like Russia, China have fully reopened their schools as they are confident about their ability to contain the virus, this could be due to strict monitoring of people and isolating them in the past or In case of Russia the invention of Vaccine. Some countries have partially reopened, this could be explained by classes held online. Some countries like Brazil have still not reopened their schools, this could be due to lack of technical infrastructure to conduct classes online.

For the Network Graphs The following inferences were made.

For **Dataset : Low income employment level in all businesses, Daily**

Naturally, the total aggregate employee level in all businesses are positively correlated to all the individual sectors of the business. Also, the employment levels in all sectors are positively correlated to each other. This implies that the total employment level is affected by all the different sectors equally at least in the lower and middle quartile earning population in the country which is also true.

We can also infer that out of all the sectors, the transportation and warehousing and accommodation and food services correlated with total employment, according to edge weight. This is intuitive because of the lockdown; more and more people are confined to their homes and services like transportation were rarely used. Similarly, accommodations like motels and hotels and restaurants were also not used because people stayed at home.

When looking at the negative correlation between columns, we can see that the month is heavily negatively correlated with all the other employment columns because as the month passed the employment level kept decreasing. We can infer that as time passed from January to May, the employment level kept decreasing in all the cities.

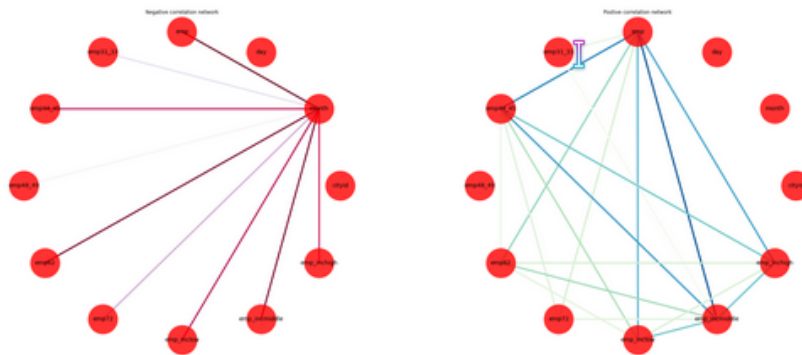


Figure 6.6: Network Plot

For **Dataset: Low income employee income in all businesses, Daily**

Naturally, the total aggregate pay of the low-income workers is simultaneously positively correlated with all the different columns. If we notice in the dataset, we will see that the pay of workers in sectors transportation and warehousing and retail trade are correlated heavily with the overall pay but also the columns throughout the dataset are very inconsistent i.e. the pay for the low income workers seems to vary both spatially and temporally. And as the employee level, all the columns show a negative correlation with money, implying that as the months passed, the amount paid to the lower income employees kept getting reduced.

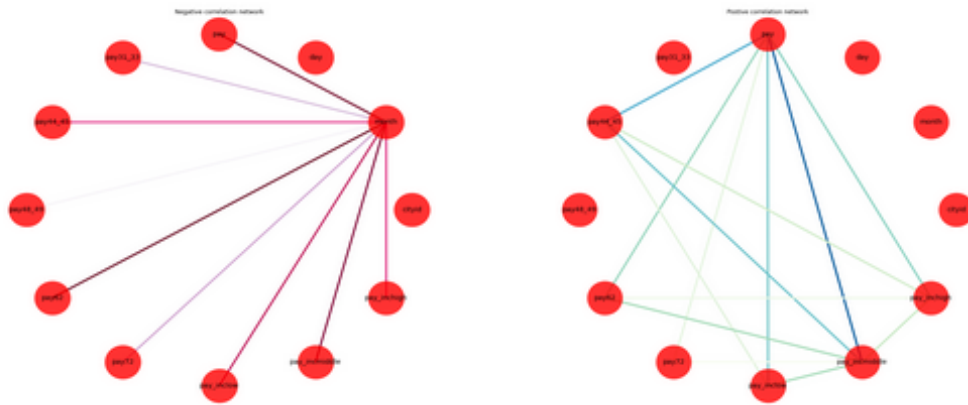


Figure 6.7: Network Plot

For Dataset: Affinity City, Daily

When looking at the negative correlation between columns (the below figure), we can see that the month is highly negatively correlated with entertainment and transport services and has small negative value with food and accommodation service. And this can be explained as food and accommodation are basic needs.

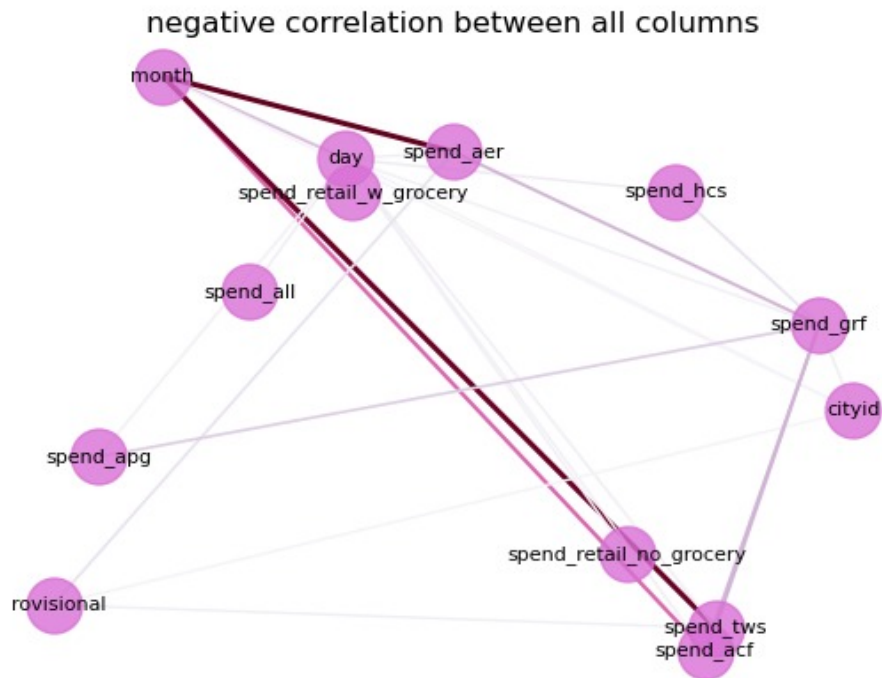


Figure 6.8: Network Plot with Negative Correlation

7 Conclusion

Our Group has successfully implemented GUI Application with Different Visualisations with mentioned datasets.

References

- [1] Plotly. <https://plotly.com/python/choropleth-maps/>. (Accessed on 26/11/2020).
- [2] Qt for python. <https://www.qt.io/qt-for-python>. (Accessed on 26/11/2020).