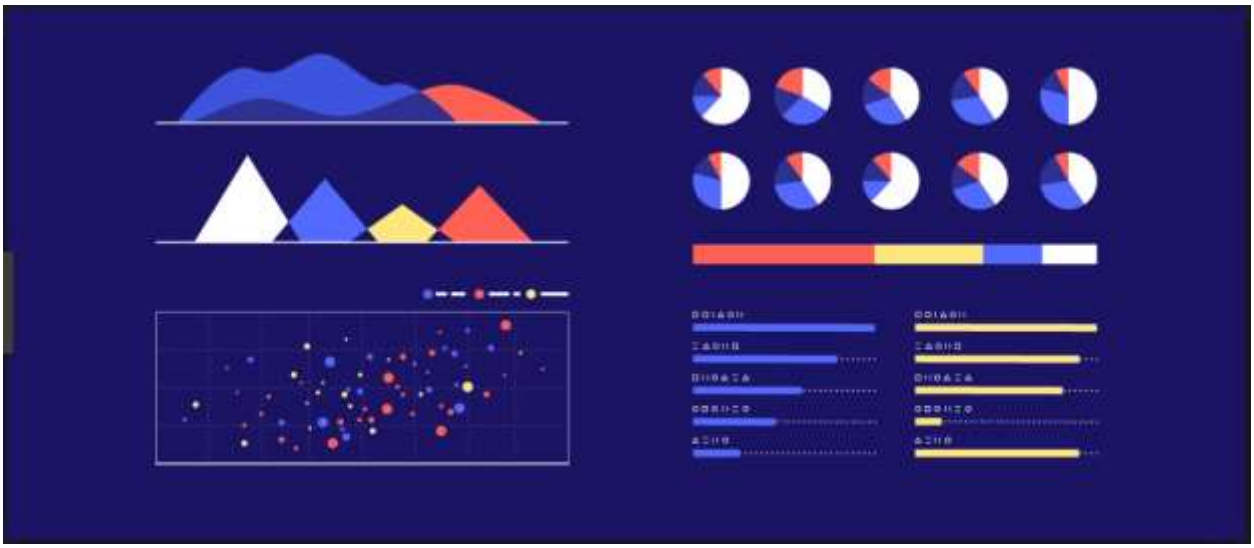
Project Report

Suicide Rates Overview

CSE 564-Visualization



Shourya Sharma: 111913983

Abhinav Srivastava: 112129749

INTRODUCTION

Suicides have been a part of society for a long time, no matter what the era or the geographical position. Apart from the personal reasons behind it, we would like to indulge and find out if there are other factors that somehow act as contributing attributes.

Through this project, we go through the data taken from multiple datasets and take a dive into it by processing it analytically and visually.

DATASET

The dataset chosen for this project is compiled from four other datasets

1. United Nations Development Program. (2018). Human development index (HDI). Retrieved from http://hdr.undp.org/en/indicators/137506
2. World Bank. (2018). World development indicators: GDP (current US$) by country:1985 to 2016. Retrieved from http://databank.worldbank.org/data/source/world-development-indicators#
3. [Szamil]. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook
4. World Health Organization. (2018). Suicide prevention. Retrieved from http://www.who.int/mental_health/suicide-prevention/en/

Above mentioned datasets are linked together by time and place and the master merged dataset is taken from Kaggle:

https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016

MISSION and HYPOTHESIS

By this project, we will try to cover below listed missions and will try to find if any of the hypothesis exists or not.

1. Data covers a span of 30 years, so one of the primary goals is to find out the suicide rate if it has decreased or increased globally.
2. Data covers for more than 50 countries, so one of another goal will be to find out the country-specific trends like for countries which have some increasing or decreasing trends of suicide rate.
3. If we can find out point #2, then using GDP and HDI values for those countries, can we somehow relate the suicide rates with the development of the country? For example in the countries where HDI and GDP values have reduced over time then if suicides rates have increased. This can bring another interesting relation between suicides and economic development of a country.
4. By categorizing data by age-groups and birth years, we will try to bring out a relationship between the suicide rate among different age groups of people. For example, the suicide rate among different age groups among different countries.
5. Since data is covered from various countries, we will try to bring out a relationship among suicide rates and geographical locations, hence using this we will try to find out a relationship that if geographic location contributes in any way to increase or decrease in the suicide rate.

Visualizing suicide rate by age-groups and by gender using Pie Charts, Line Charts, and Stacked-to-Grouped Bars.

The left Pie chart below is created using two attributes from the entire data set i.e. age groups and number of suicides. Then using python code at the backend data corresponding to each age group is summed up and represented as a Pie chart. So the angle of a sector is proportional to the frequency of suicides occurred in that age group. Using the same technique the right pie chart is created by counting the total number of males and females from the complete data set for an entire span of time.

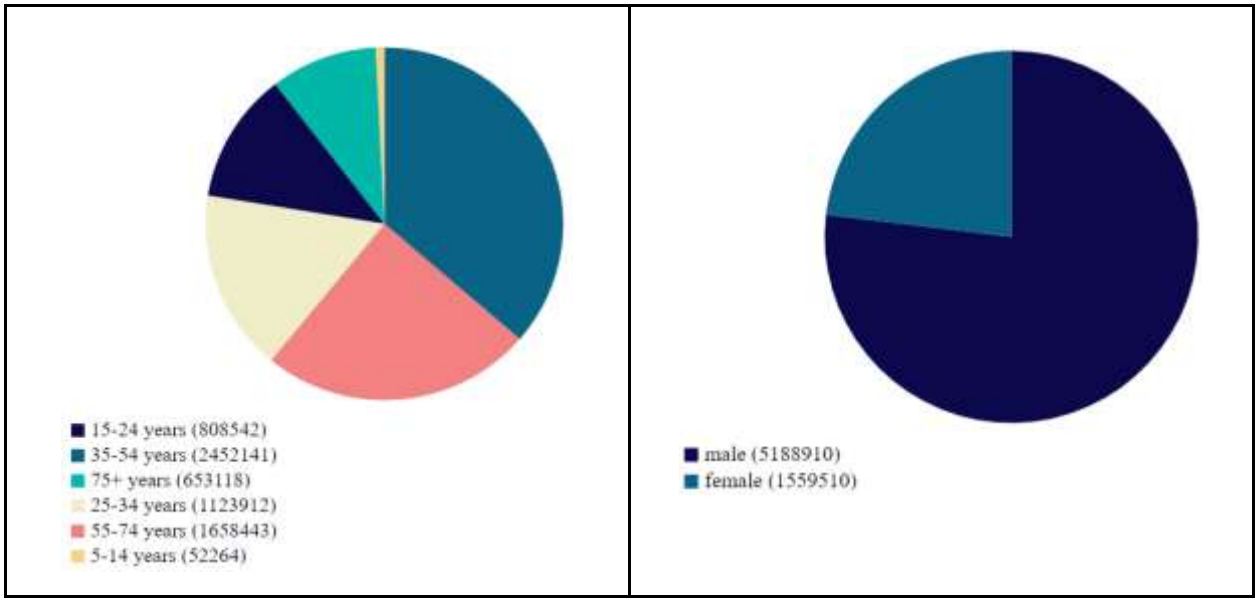


Figure 1: Left: Each sector is age group with number of suicides. Right: Sector is gender

Using Pie charts to segment data in forms of categories but it doesn't convey the complete or detailed information like how each year wise suicide rate was for gender wise or age group wise. So a better representation of the same data is done using Stacked to Group bar chart.

Stacked-to-Group Bars

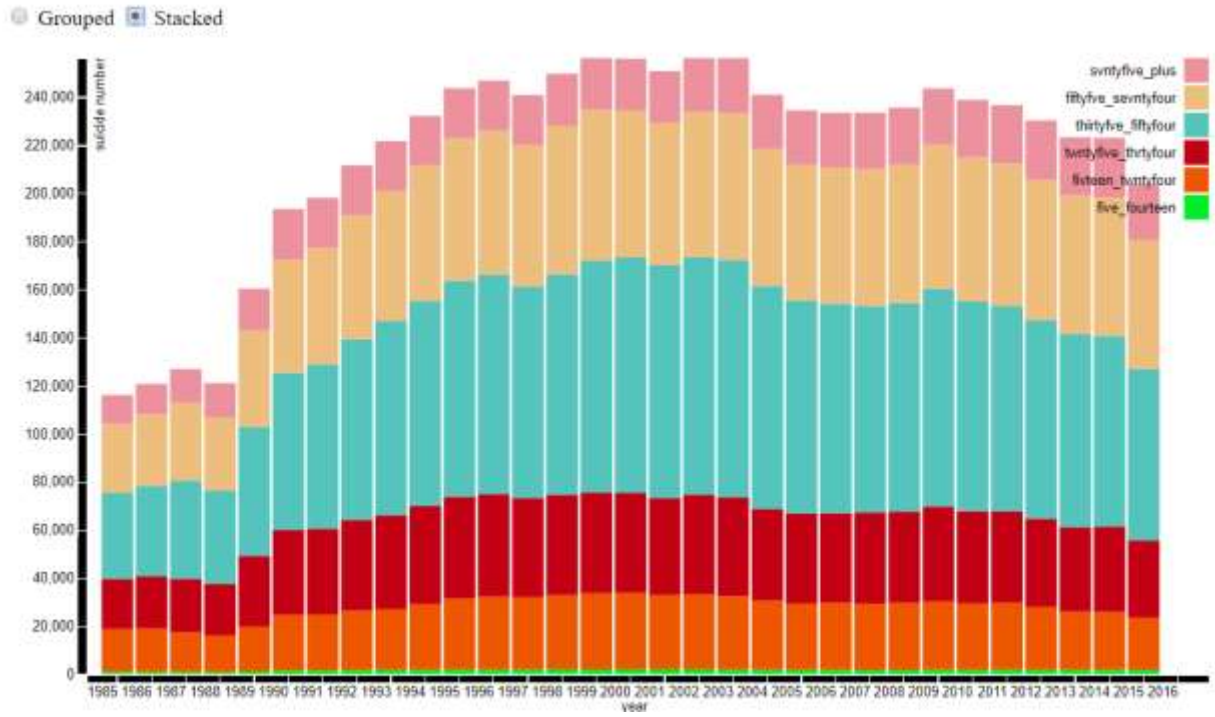


Figure 2: Stacked bar graph: x-axis: year vs y-axis: number of suicides

Figure 2 represents more clear and detailed information. Each stack/bar represents the year and each section of the stack represents different age groups. So a more detailed analysis can be done using the above chart that people belonging to age group 35-54 years have the highest number of suicides and people which age group 5-14 years have committed least suicides over each year. Figure 2 also shows the information that in which year, the highest number of suicides occurred across the globe.

A radio button is provided to change this graph to individual groups so as to have a much clear analysis of data. So Figure 3 is the follow-up of Figure 2. As it can be seen a trend is increasing for suicides among the age group 35-54 years from the year 1985 to 2001 but after that has plateaued. Apart from this, another analysis is that the maximum number of suicides are also contributed by the age group 35-54 years followed by 55-74 years, age group. The least number of suicides is from the age group 5-14 years.

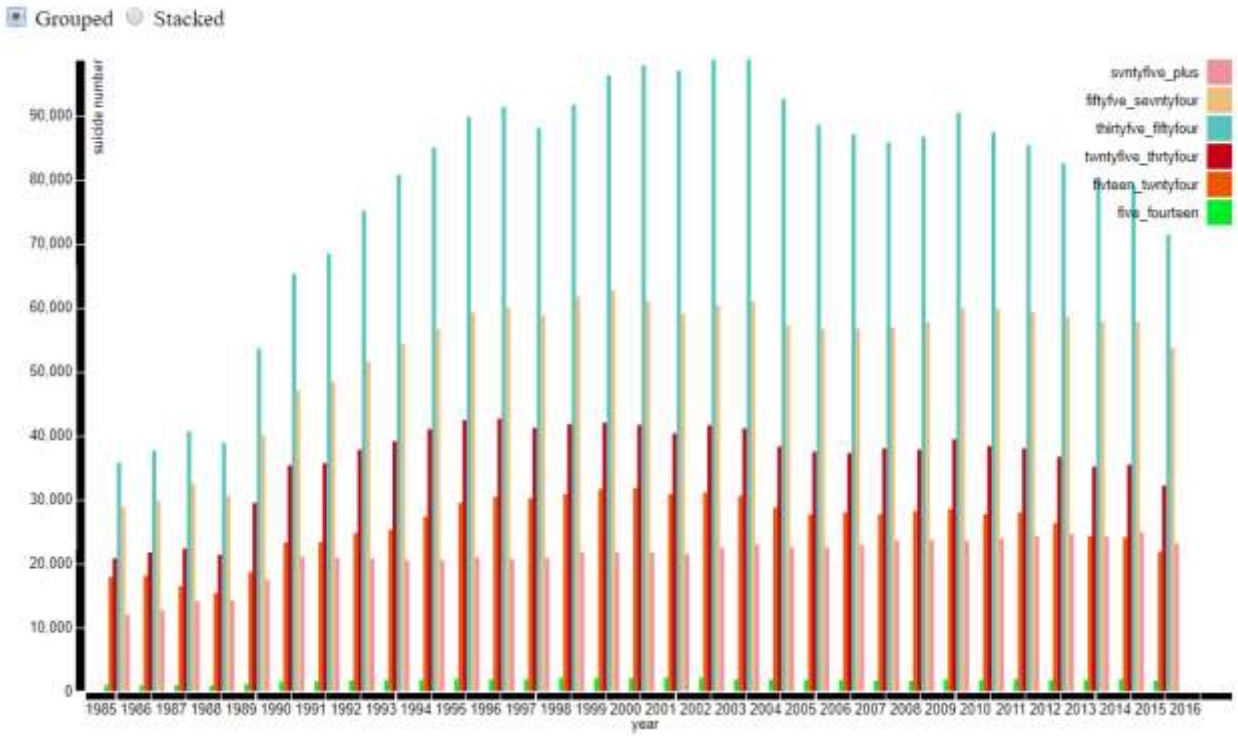


Figure 3: Group Bar graph, where groups are age groups. On x-axis with year and on y -axis is the number of suicides

The beauty of data visualization can be attributed to the fact that a story can be build using charts and plots and it is easy to convince and deceive the audience using this. That's why it is always advised to view data from more dimension to have a better understanding. Continuing the above principle, we plotted the same Stacked-to-Group chart with the year representing each stack and each section of the stack by 1. Age group and 2. Gender. But this time the height of each section stack is filled by suicides committed per 100K population. Suicides occurred per 100K population is much better and true data on which results should be analyzed.

Figure 4 and Figure 5 represents a better analysis as suicides are captured per 100K population. It clearly shows that people belonging to the age group 75+ years when calculated per 100K population have the highest height. A point to observe here is that suicides in all age group has increased each year and there is a sudden increase in number of suicides for year 1996.

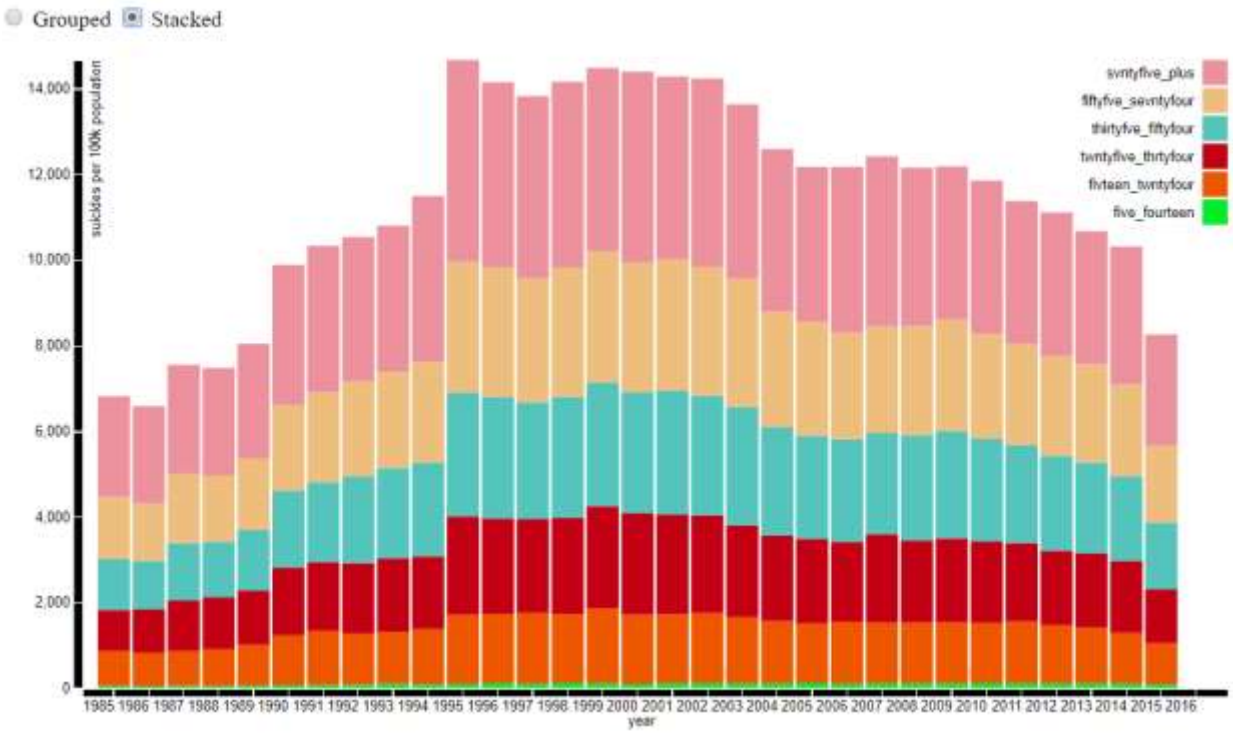


Figure 4: Stacked bar graph: x-axis: year vs y-axis: number of suicides per 100K of population

Here we have Stacked Bar Charts that have been grouped according to the age-groups. As expected we see that suicide rate is very low among children.Also the other groups approximately similar proportions, with the elderly dominating slightly here.

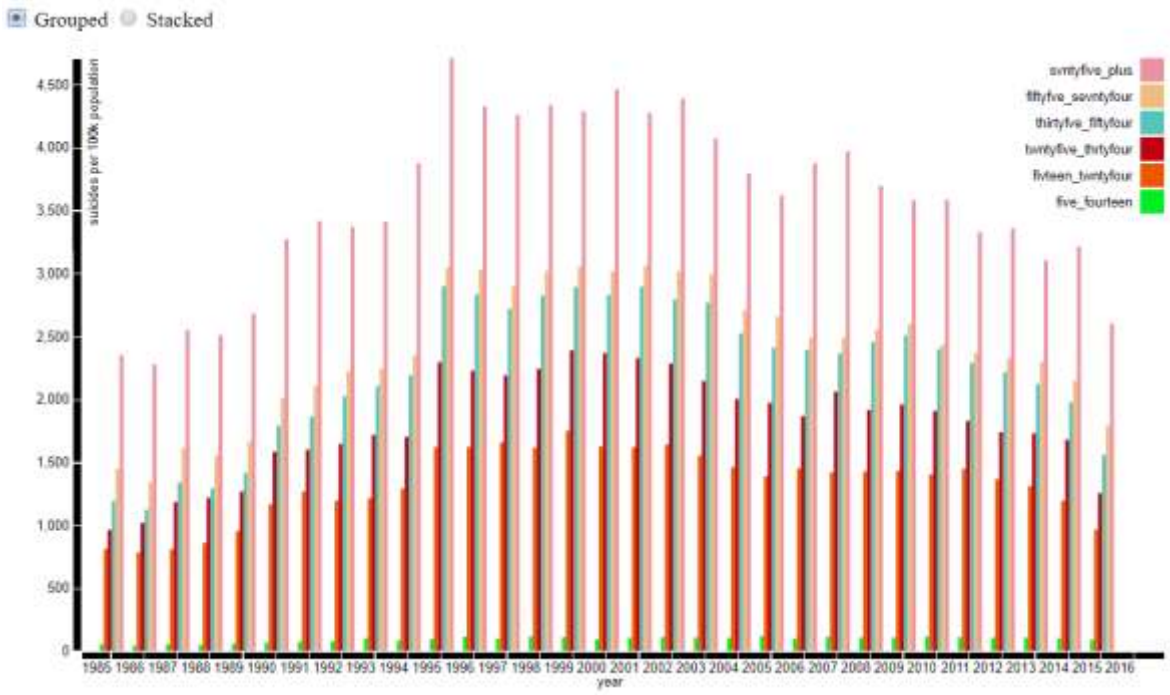


Figure 5: Group Bar graph, where groups are age groups. On x-axis with year and on y -axis is the number of suicides per 100k population

Here we have Bar Charts over the years for each of the genders. We see that the males always have a much higher suicide rate over the years.

Line Chart

Line Charts are a good way to represent data in the form of trends when data points are plotted correctly on the axis. Figure 6 and Figure 7 shows the trend of suicide per year wise across the globe. In Figure 6, the y-axis is plotted using suicide number only and years on the x-axis. So it clearly shows that people belonging to age group 35-54 years have the highest representation.

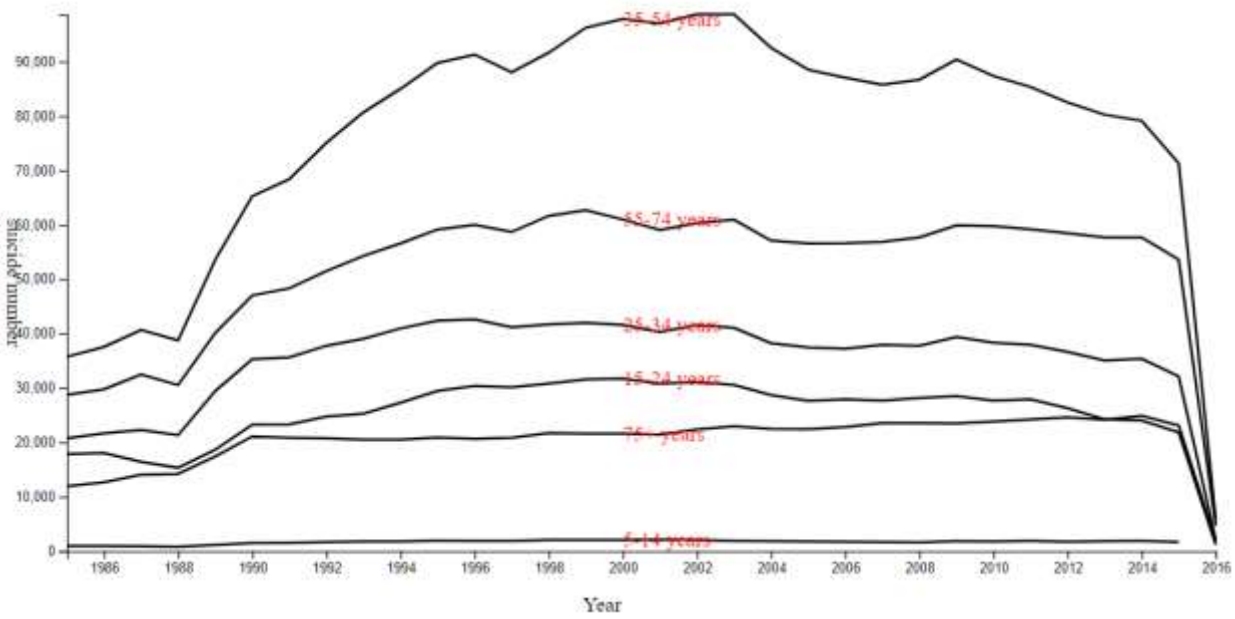


Figure 6: Line chart where each line is for different age groups: x-axis: year, y-axis: number of suicides

In Figure 7, the y-axis is plotted using suicide number per 100K population and years on the x-axis. So it clearly shows that people belonging to age group 75+ years have the highest representation.

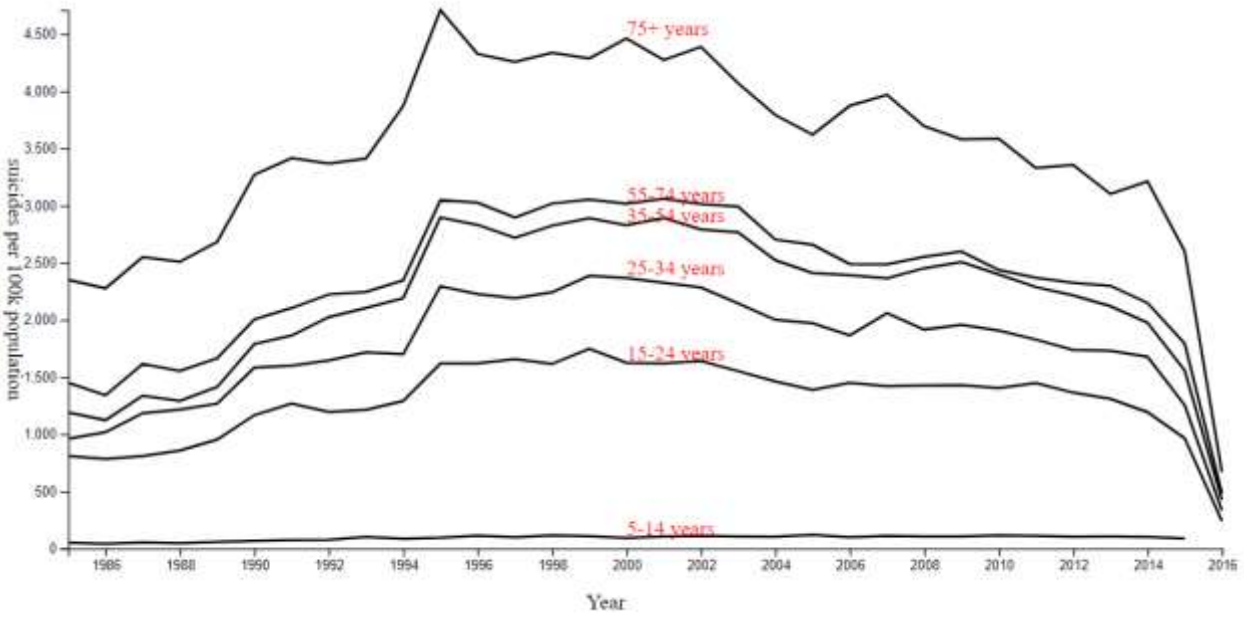


Figure 7:Line chart where each line is for different age groups: x-axis: year, y-axis: number of suicides per 100k population

A common trend which can be easily observed from both the graphs is that number of suicides have increased over the period of time but have plateaued from 2000 to 2010 and have been seen to reduce from then on. The lines converge over the recent years because we do not have data from all countries for the recent years

Visualizing suicide rate Country-wide.

1. To show the suicide rate over the period of time for countrywide and over the globe, we have used line charts to represent the data. The below line chart displays global data.

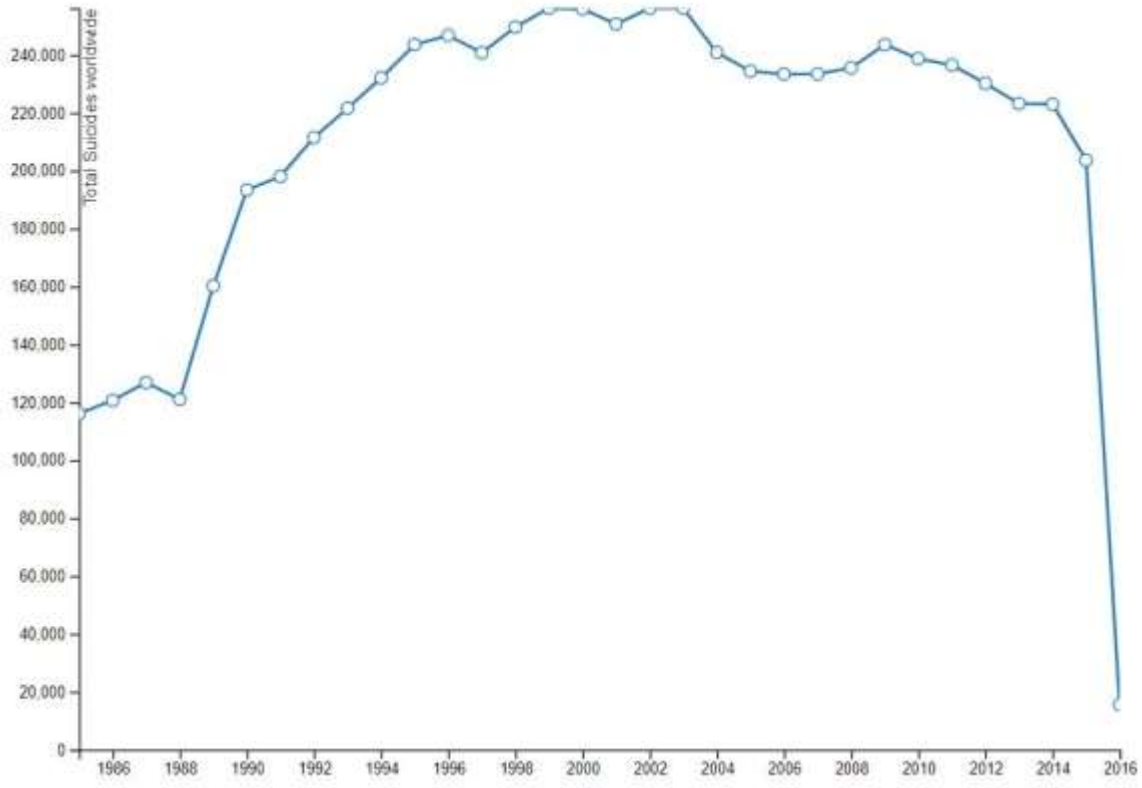


Figure 8: Total Suicides Worldwide

Although this chart gives us a holistic view, it would have certain limitations:

a. We cannot trust this representation as we don't know how the total population over the globe is changing. For example, if the total population is decreasing and the total suicides are increasing, then the situation is even worse.

b. Towards the year 2015, the total suicides drop drastically, this is because we don't have data for that year from all the countries. Thus, this could lead to misinterpretations.

2. To overcome the above limitations, we have used suicides rate, i.e. number of suicides for every

100,000 population. Such a representation provides us a better understanding. Also, we have added functionality to focus on a particular data point to view the year and the value of the data point.

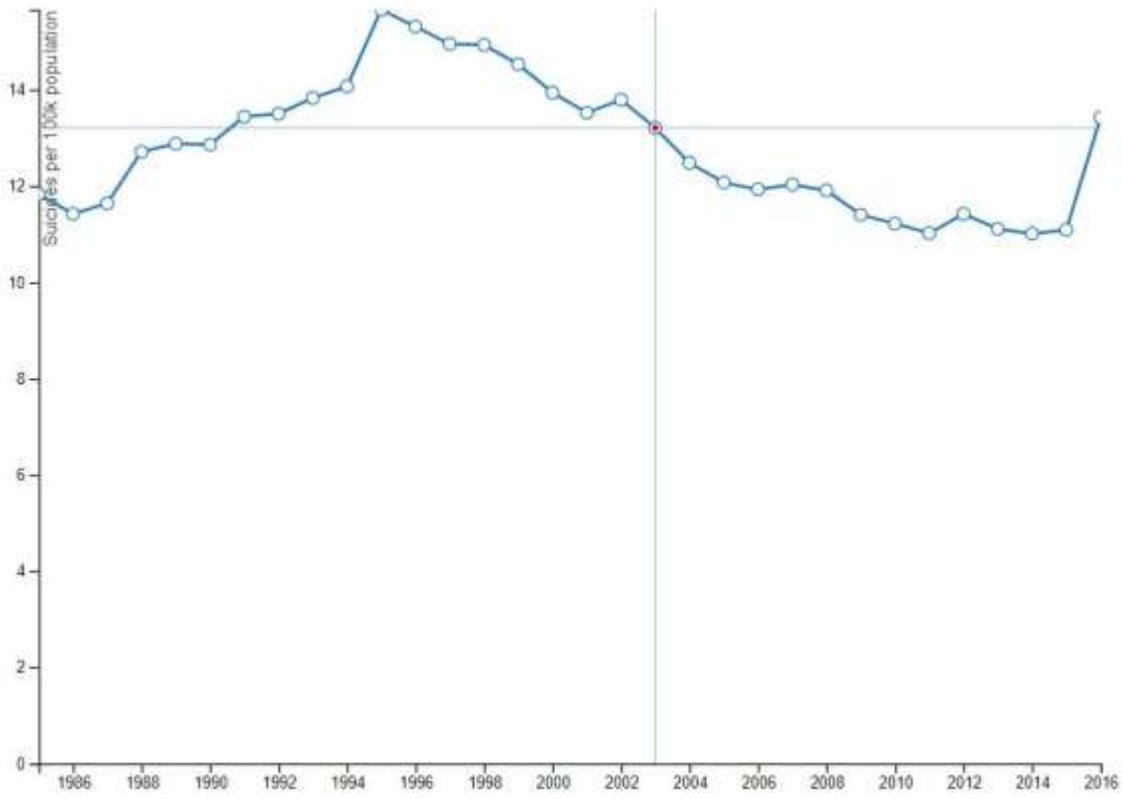


Figure 9: Worldwide suicides per 100k population

3. To represent how various countries compare with the global data, we added line chart for three other countries to the same global average chart.

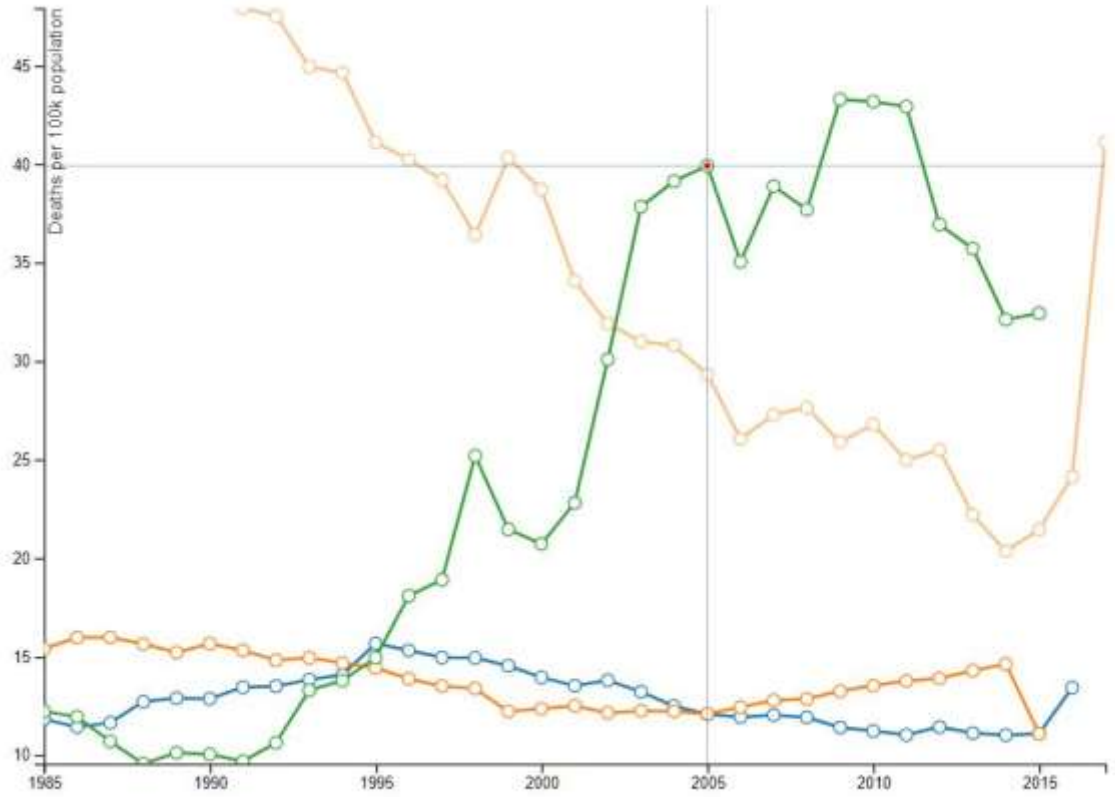


Figure 10: Global, USA, South Korea and Hungary suicides per 100k population

Here, blue represents the global data, orange represents USA, green represents South Korea and yellow represents Hungary. While presenting multiple countries together it helps us to compare them, but at the same time it might get a bit cluttered.

4. To overcome this congestion, the user can click and drag over an area to zoom into it.

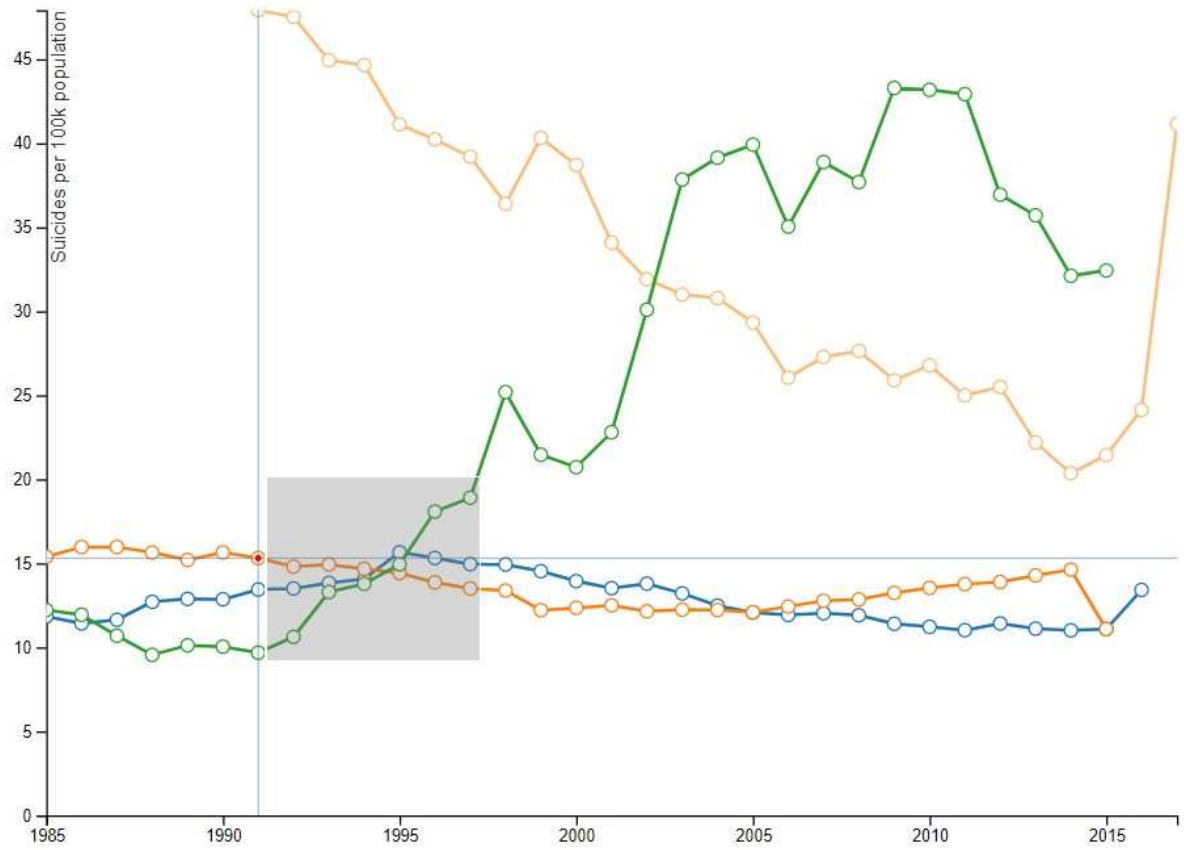


Figure 11: User can zoom over a particular area

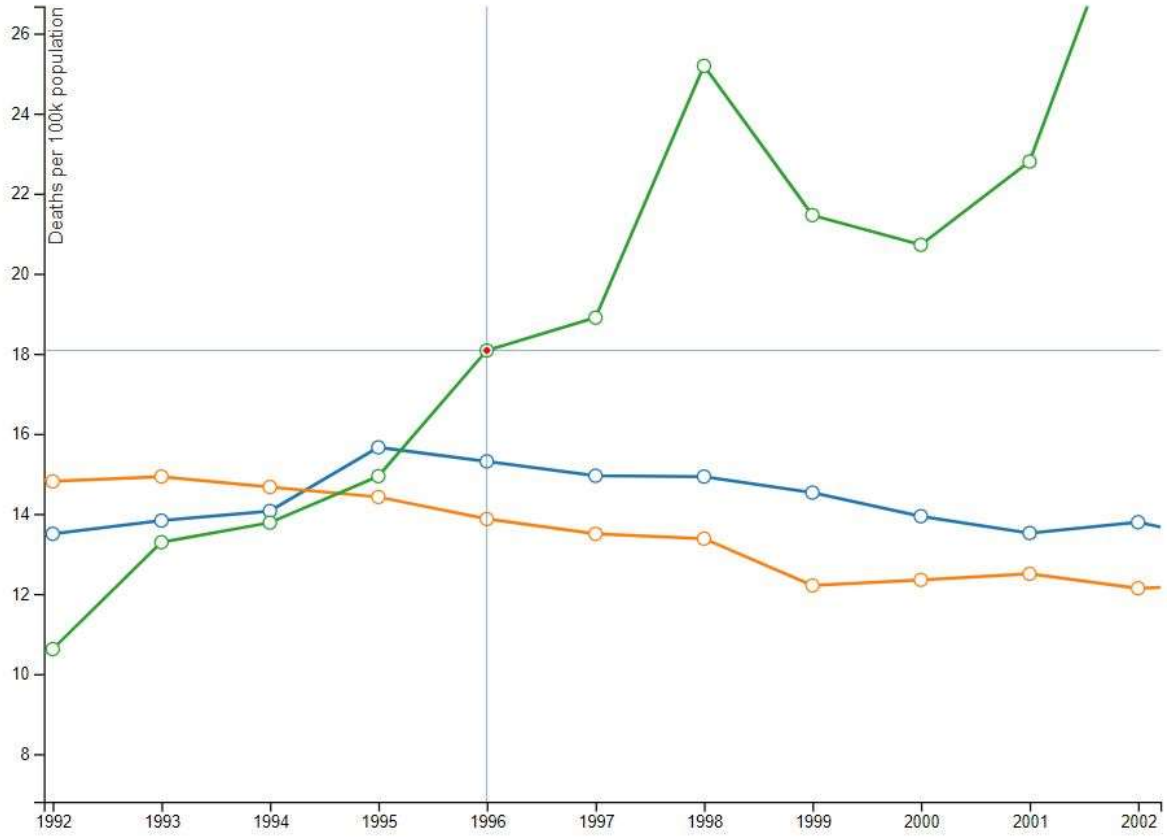


Figure 12: The line chart has been zoomed in

Now as the chart has been zoomed in into the area selected by the user, the values on the x and y axis change as well. To zoom out and return to the initial state, the user just needs to double-click on the chart.

Dashboard

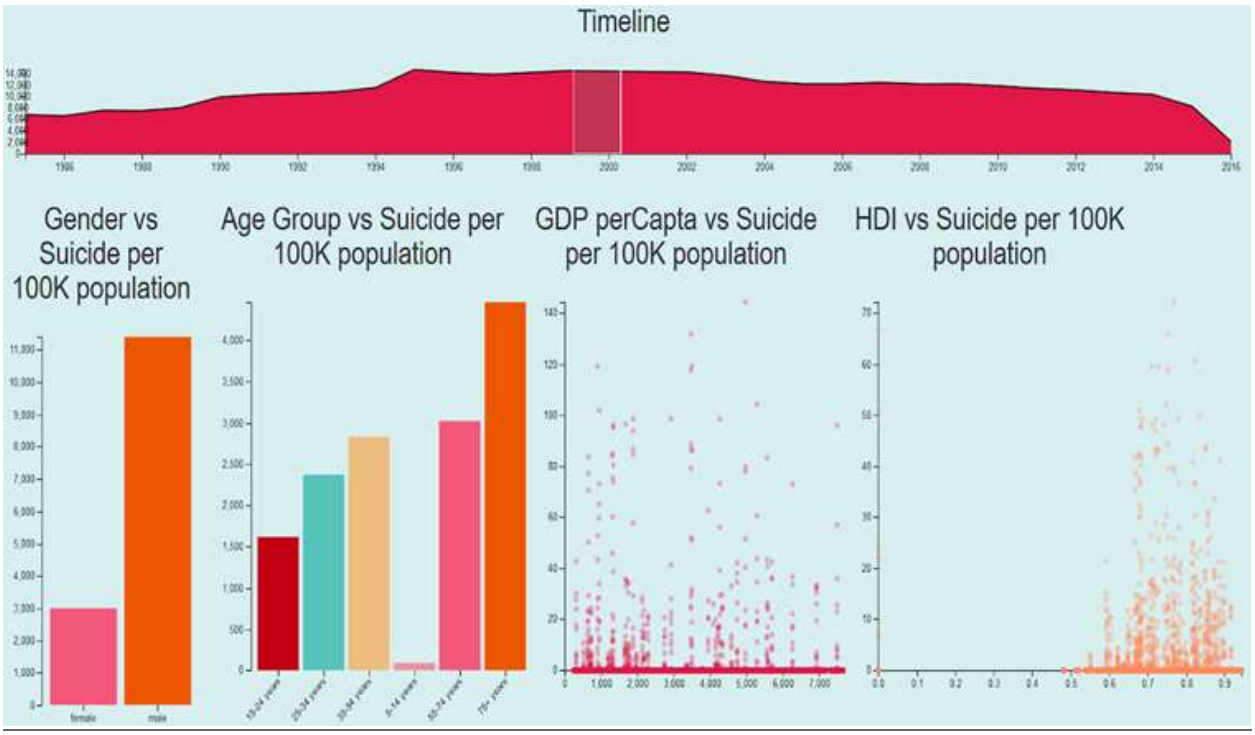


Figure 13: Dashboard that displays brushing and linking in multiple graphs

- As part of brushing the data, we have created a dashboard that visualizes the global data.
- On the top we have a timeline which also shows the suicides per 100k population.
- As we select a certain timeline the other graphs also change values accordingly.

Zoomable Tree Map

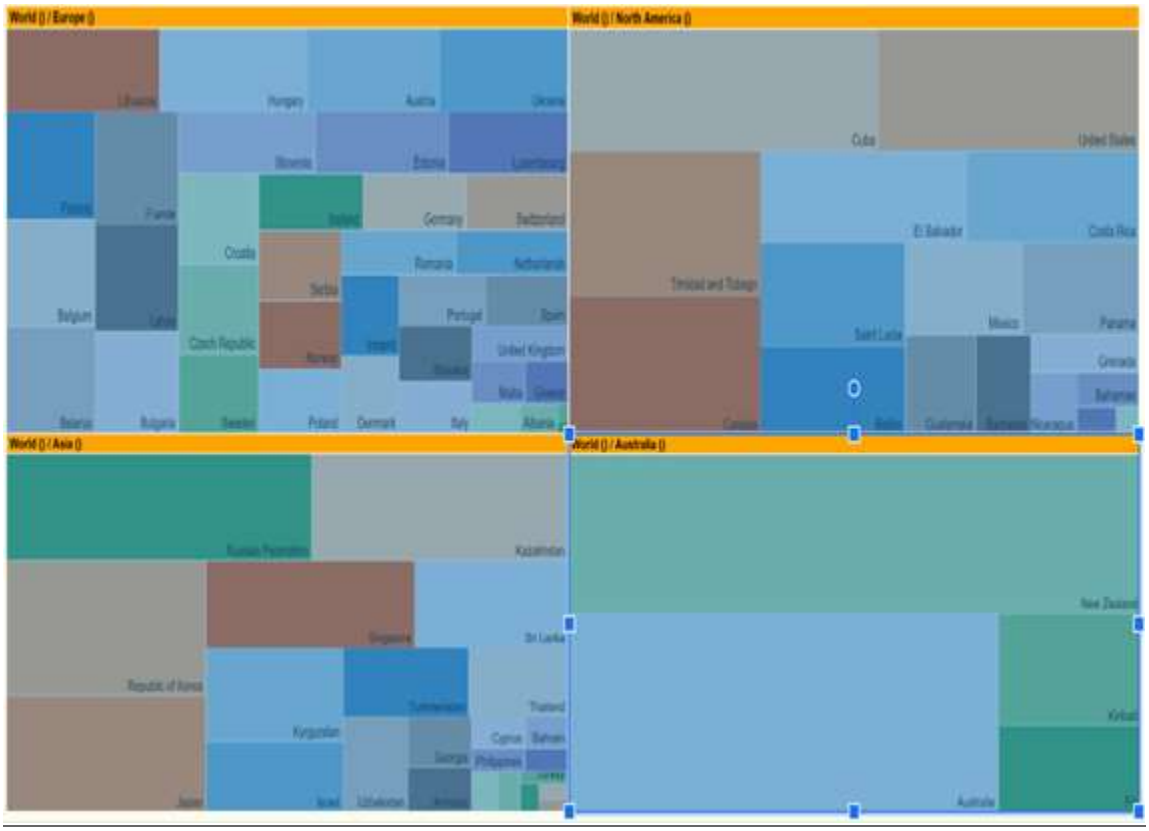


Figure 14: Zoomable Tree Map for various continents

From the Tree Map which is plotted for Continents→ Countries→ Suicides per 100K population, it shows that highest number of suicides have occurred in Europe followed by Asia and North America. The map can be further zoomed to get suicide contribution by each country in that continent.

Trends among GDP, Suicides, and HDI

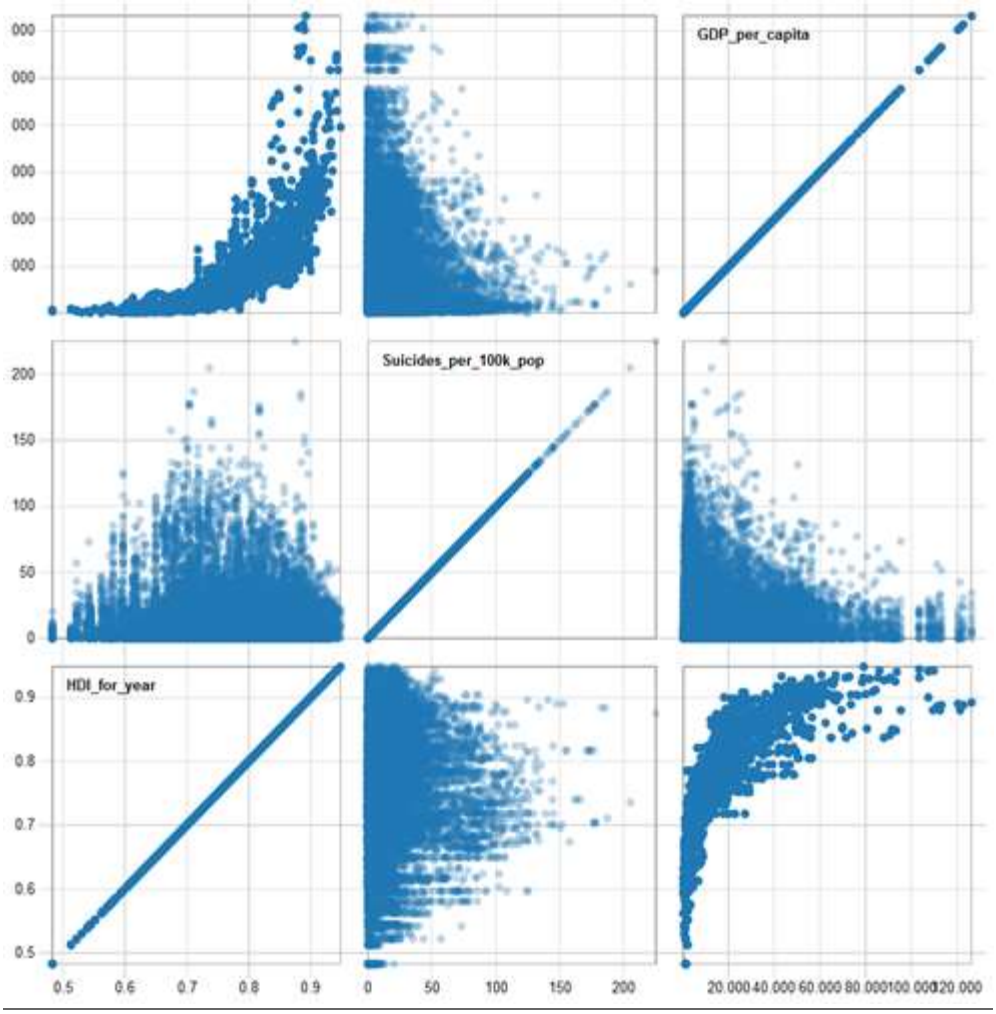
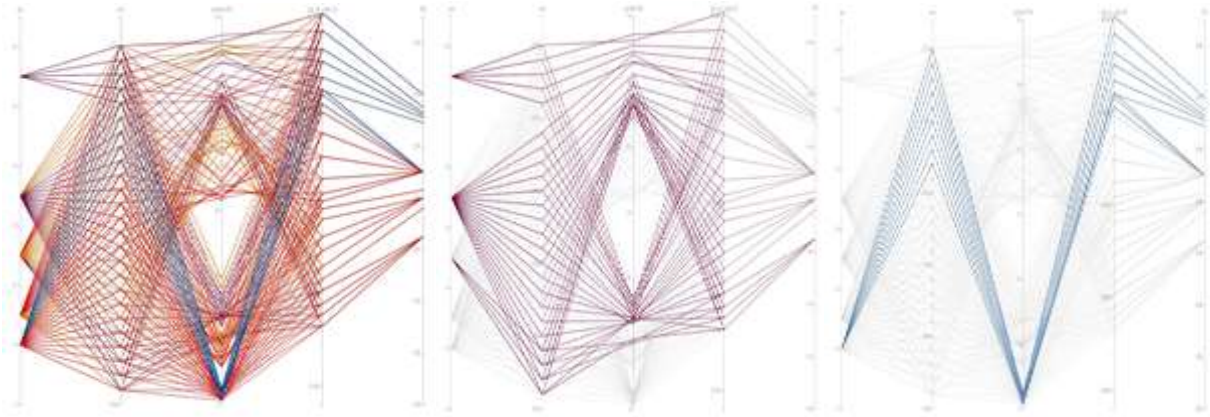


Figure 15 : Scatterplot Matrix for HDI, suicides, and GDP per capita

Through this scatterplot matrix, we try to find the relationship among GDP per capita, suicides per 100k population and HDI data globally. We can see that there is a slight positive correlation between GDP per capita and HDI. Also there is some negative correlation between GDP per capita and suicide rate. Moreover, it turns out that there is no correlation between HDI and suicide rate. This could also be due to inconsistent data and due to missing values in the dataset.

Parallel Coordinates



(a) **(b)** **©**

Figure 15 : Parallel Coordinates : a) USA, b) Boomer Generation c) Generation Z

The first graph shows the data for USA. Each color depicts a specific generation. In the (b) part we see the data of the Boomer generation. We can conclude how GDP-suicide rate and year-suicide are negatively correlated. In the (c) part we see the data of the Generation Z. Again we can easily visualize the relationship of how GDP-suicide rate and year-suicide are negatively correlated.

Note :

The coordinates from left to right in the graphs are : Age, Year, Suicides per 100k population, GDP per capita, and HDI

CONCLUSIONS AND INFERENCES

- From the line charts, a global trend can be observed that there was rise in suicide rate up to 1996 but after that it is gradually decreasing.

- From the global data plotted on line chart for two parameters i.e. a. Number of suicides and b. Suicides per 100K population tell different stories.
  - People belonging to 35-54 years age group have committed most suicides in every year.
  - When the same results computed over suicides per 100K population of that age group, it was found that people belonging to 75+ years age groups commits most suicides.

- From the scatter plot matrix graph of world data, it can be concluded that there is a positive correlation between GDP and HDI and a negative correlation between GDP and Suicide rate. Also there is no correlation between HDI and suicide rate.

- Parallel coordinates graph plotted for United States also supports above conclusion.

- From the Tree Map which is plotted for Continents→ Countries→ Suicides per 100K population, it shows that highest number of suicides have occurred in Europe followed by Asia and North America.