FIT5147 Data Visualisation Assignment - 2

Jaimon Thyparambil Thomas
Student ID : 29566428
Email : jthy0001@monash.student.edu
Monash University
8 June, 2019

Contents

Introduction

Every year approximately 800000 people die due to suicide which is approximately 1 person per second. There is also sources which say that for every person who has committed suicide there is approximately 20 persons who attempted and failed. Suicides happen globally across all age groups, gender etc. at different rates. Suicide is a serious problem worldwide but it is one of the most preventable when comparing to other cause of death.

Message to be conveyed through visualization:

In this data exploration project, various patterns and facts in suicide rates based on country, age, gender etc. can be easily explored. Also relationship between the suicide rates and socio economic factors such as GDP, happiness Rank etc. can be viewed. So based on the patterns and relationship found from these webApp regarding suicide rates. Respective authorities can take necessary preventive measures to avoid or handle suicide case.

Intended Audience:

Intended audience is any individual or group of people trying to bring a change in society by doing good for others like NGO's, Politicians, Government officials. As through this webApp they can find various trends and take necessary preventive measures.

Degree of Difficulty:

I had used D3 for visualisation. I had also used a large dataset that is almost 30K rows on which various filters had to be applied dynamically.

Design

Design process began by finding all possible ways in which data could be made interactive. By finding how the data should be categorized. So that the message is properly conveyed without any redundancy. The main idea behind the selection of visualization was that it should be clear and precise so that the audience viewing this won't take much time figuring out what the visualizations actually was meant to convey.

Sheet1

Sheet 1 of the design sheet included all the possible ways in which data can be categorised and visualized. In current situation as we find more methods we increase the chance of getting a better option for visualization. Initially I had to decide on the parameters on which I was going to represent the information. The parameters that I considered important for visualization are:

- Gender Wise Comparison
- Age Wise comparison
- Country Wise comparison
- Socio-economic factor wise comparison
- Year wise comparison

After I had come up with the parameters I had to find ways in which I could visualize these parameters efficiently so that all the information can be incorporated. So I had to find all possible ways in which all this information can be visualised. Out of the ideas that I had found out I had to select the best ones. And filter the data by removing duplicates or overlaps. Then similar ideas were combined together like the ones considering only countries were combined together. The ones

considering gender, age etc. all these were grouped together respectively. Then out of these groups I had to check which all can be combined and the plots that conveyed same information was removed. Finally, I had to check if these plots that I had found out was able to answer my questions. If I am satisfied regarding this then I can move on to next sheet.

Sheet 2

In this sheet I thought of grouping all the data based on Year. So that I can know the respective statistics for that particular year

Bar chart for representing Age groups:

In this design I will plot bars for different Age groups. Where the height for the bar represents the no of suicide and the colour and x axis would represent different age groups. This was a viable option to convey the age groups respective suicide rates efficiently.

Donut Chart to represent Gender wise comparison:

In this design we can clearly see the comparison of different genders respective suicide rates.

Heat map of the world to represent suicide rates over the world:

In this design we can clearly see the how suicide rates are spread across the world as the colour of the country shows the no of suicides on each country

Sheet 3

In this sheet I thought of grouping all the data based on Country. So that I can know the respective statistics for a particular country.

Stacked Area chart for showing age wise info over the year:

In this design the x axis represents the year and y axis represent the no of suicides here each age groups information is stacked one over other and the colour of the area represents the respective age group. This is a viable option because it shows the trend of each age groups individually and as a whole clearly.

Donut chart for showing gender wise information:

In this design we can clearly see the comparison of different genders respective suicide rates.

Radial Bar chart for suicide rates based on socio economic condition like GDP:

In this design we can clearly see the trend in suicide rates for different socio economic conditions.

Sheet 4

In this sheet I thought of grouping all the data based on Socio economic condition like select only countries with high GDP values. So that I can know the respective statistics for a particular socio economic condition

Line chart for showing age wise info over the year:

In this design the x axis represents year and the y axis represents no of suicides and colour of the line represents different age groups. This is a viable option because we can clearly see the trend of suicide rates over the year for each individual age group.

Word cloud for showing the Top countries based on suicide rates:

In this plot all the top countries name is clearly visible as the size of the name is proportional to the no of suicides. It is a viable option as we can clearly identify the top countries names.

Donut chart for showing gender wise information:

In this design we can clearly see the comparison of different genders respective suicide rates.

Sheet 5

In this sheet I thought of grouping all the data based on Socio economic, country, year like select only countries with high GDP values. So that I can know the respective statistics for a particular socio economic condition or view the statistics for a particular year. I picked all these conditions as I feel all these conditions are relevant so that I can view for each one respectively also see the impact of one condition over the other more clearly. It also provides even drill down further by selecting multiple conditions.

Stacked Area chart for showing age wise info over the year:

In this design the x axis represents the year and y axis represent the no of suicides here each age groups information is stacked one over other and the colour of the area represents the respective age group.

I had picked this over lined graph because in lined graph even though we can know the individual age groups trend over the time period we can't know the overall trend. Similarly, for bar chart even though the individual value is shown clearly we can't see its impact over the year. But in stacked Area chart both information is being conveyed properly. Due to these factors I had picked stacked Area chart over line graph and bar chart.

Donut chart for showing gender wise information:

In this design we can clearly see the comparison of different genders respective suicide rates. Compared to that of pie chart and bar chart. So this was selected.

Radial Bar chart for suicide rates based on socio economic condition like GDP:

In this design we can clearly see the trend in suicide rates for different socio economic conditions. So this was selected.

Heat map of the world to represent suicide rates over the world:

In this design we can clearly see the how suicide rates are spread across the world as the colour of the country shows the no of suicides on each country. I had picked this over word cloud because in word cloud we can see only few of the countries whereas in world heat map we can see the information of all the countries

Implementation

For the implementation I have decided to use D3. As I feel that I will be able to represent the visualizations that I need in D3 in a better way compared to that of Shiny.

The Libraries I have used are as follow:

- https://d3js.org/d3.v4.min.js - 4th version of D3 library

- http://d3js.org/queue.v1.min.js - Used for loading multiple files
- https://maxcdn.bootstrapcdn.com/bootstrap/3.4.0/js/bootstrap.min.js - Used to get the predefined styles in CSS

I have kept the layout of each page as simple as possible so that any person who doesn't have any knowledge regarding the suicide rates can also try this and analyse the trends.
Initially I had given a Home page where user can know about the seriousness of the issue and other facts regarding the suicide rates. I Have also given one page each for
I have also given a page for each category wise analysis like gender-wise, age-wise, socio-economic factor and country wise. There in each respective page I had provided buttons for various filter options like gender, age groups, socio economic factors etc. on the left side of the page and its respective category-wise visualization on the right side of the page. So that the user can see the impact of each filter on the respective category through visualization.
I have also given an over view page where we can see all the visualisations simultaneously.
In the overview page the top portion shows different visualizations for each category. Whereas in the bottom portion I have provided buttons so that the user can select desired filters.
Initially the amount of rows for the main data set is 27820 rows. This main data set have been used on all category wise filters except for socio-economic factor. Then based on the desired filter the data is been dynamically filtered out and processed for visualization using JavaScript. In socio-economic factor wise page, the data set is a combination of suicide data plus the happiness rank and other datasets. So the information available for those data sets are only of the year 2015 and 2016. Therefore, the amount of rows in the combined dataset is just 820 rows. Here also rest of the process is same that is based on the filters selected data will be dynamically processed and visualised using D3.

In the all the graphs except for age wise category on hovering over the desired entity we can see the respective no of suicides for that entity after applying all the filters. Similarly, in the overview page if we click on a respective country. Then the information will be shown of the selected countries only. There is also an option provided in the filter section to reset the country based filter so that all user can see the suicide rates of all countries. These features added a much interactive feature to the visualization.

User Guide

All the information required to effectively utilize all the features of this webApp have been provided below.

Steps to start the WebApp:

In order to start the webApp unzip the files and host a server in the location where it was unzipped then click on index.html.

Home Page

Once you load the webApp the user will be directly taken to the Home Page which look like Figure 1

Figure 1

In the home page a brief introduction regarding the issue have been provided. Also some of the facts and trends regarding suicide rates which have been analysed is mentioned.

From the home if the user feels like to analyse the trends themselves then they can view the desired category wise information by selecting the respective tabs in the top navigation bar.

Gender

To go to this page, the user must click on the gender Tab on the top most navigation bar as highlighted with a red circle in the figure 2 given below

Figure 2

Then the user will be taken to the Gender page. In this page the user can see the gender wise summary and trends regarding the suicide rates. This page looks like figure 3 given below

Figure 3

Here as we can see the filters are provided as buttons on the left side whereas the visualization for the respective filter has been provided on the right side of the page. Various filter options provided for gender wise visualisation are

Suicide Rate Aggregate Filter:

Based on th value of this filter the user can choose how the suicide rates has been combined for displaying in the visualization like whether it is by average of all the suicide rates or just based on the sum of all values. By default, Average has been selected. If the user wishes to display the visualisation based on sum, then the user can click on the sum button. Then that button will be highlighted as in figure 4.

Figure 4

Age Groups Filter:

Based on the value of this filter the user can decide which all age groups information has to be considered for visualisation. By default, all the age groups will be selected. In this case the image looks like figure 5. If the user doesn't wish to select all the age groups information, instead of it they just wish to select just few of the age groups the user can select the desired age groups by click on it. Then the selected age groups will be highlighted as in figure 6. In order to reset the selection that is show information of all age groups the user just have to select All age groups tab.

Figure 5

Figure 6

GDP Range (in Billion $):

Based on the value of this filter the user can decide which all countries have to be considered based on their GDP. By default, all values will be selected. In this case the image looks like figure 7. If the user doesn't wish to select all the age groups information, instead of it they wish to select just few GDP range values then the user can do so by clicking on the desired GDP values. Then those values will be highlighted as in figure 8. In order to reset the selection that is show information of all GDP ranges then the user just have to select all values tab.

Figure 7

Figure 8

Suicide Rate Data Type:

Based on the value of this filter the user can decide which value should be considered for visualization. Like whether it is based on total no of suicides or based on the suicides No per 100K population. By default, suicide no is being selected. In this case the filter looks like figure 9. If the user wishes to select value based on suicide no per 100K population. The user can click on suicide no per 100K population button. The filter option that has been selected will be highlighted

Figure 9

Gender wise visualization:

The gender wise visualisation after all the filters looks like figure 10. Here the proportion of values based on gender is shown in percentage. On hovering over each section you can know the exact value as shown I figure 10 for male suicides.

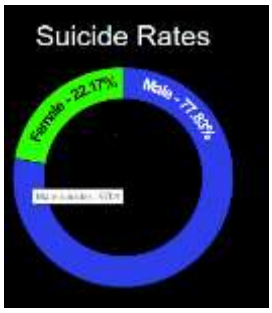


Figure 10

AGE Group

To go to this page, the user must click on the gender Tab on the top most navigation bar as highlighted with a red circle in the figure 11 given below

Figure 11

Then the user will be taken to the Age Group page. In this page the user can see the age wise summary and trends regarding the suicide rates. This page looks like figure 12 given below

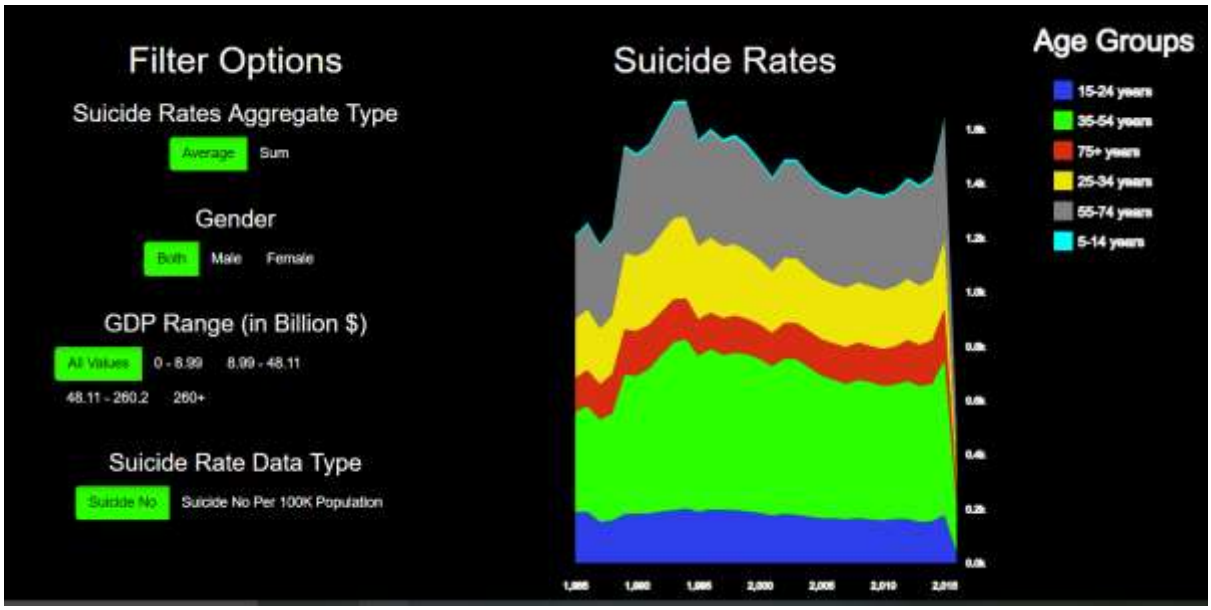


Figure 12

Here as we can see the filters are provided as buttons on the left side whereas the visualization for the respective filter has been provided on the right side of the page. Various filter options provided for age wise visualisation are:

Suicide rates Aggregate type: same as gender page case

Age Group: same as gender page case

Gender:

Based on the value of this filter the user can decide based on which gender data should be considered for visualisation. By default, both male and female will be selected as shown in the Figure 13. If the user wishes to select just one gender, then the user can click on the desired gender. Once the gender has been selected then that gender will be highlighted. In order to reset the value a show the information for all genders click on Both.

Figure 13

GDP Range (in billion $): same as gender page case

Suicide Rate Data Type: same as gender page case

Age Wise Visualization:

Here Figure 14 represents the Age wise information with x axis denoting the year and Y axis denoting the no of suicide rates. Here each colour represents each age group

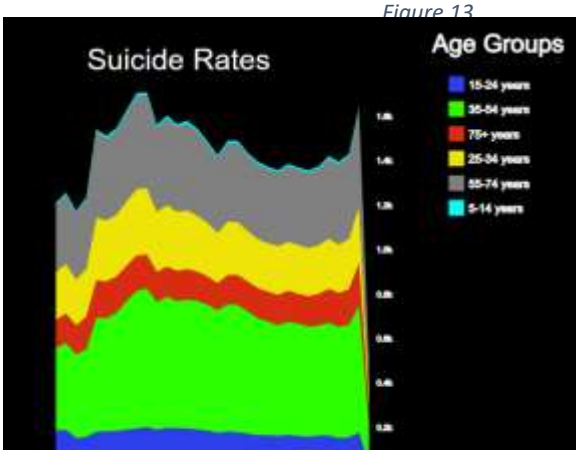


Figure 14

Socio-Economic Factor:

To go to this page, the user must click on the Socio Economic Factor Tab on the top most navigation bar as highlighted with a red circle in the figure 15 given below

Figure 15

Then the user will be taken to the Socio Economic Factor page. In this page the user can see the socio economic factor wise summary and trends regarding the suicide rates. This page looks like figure 16

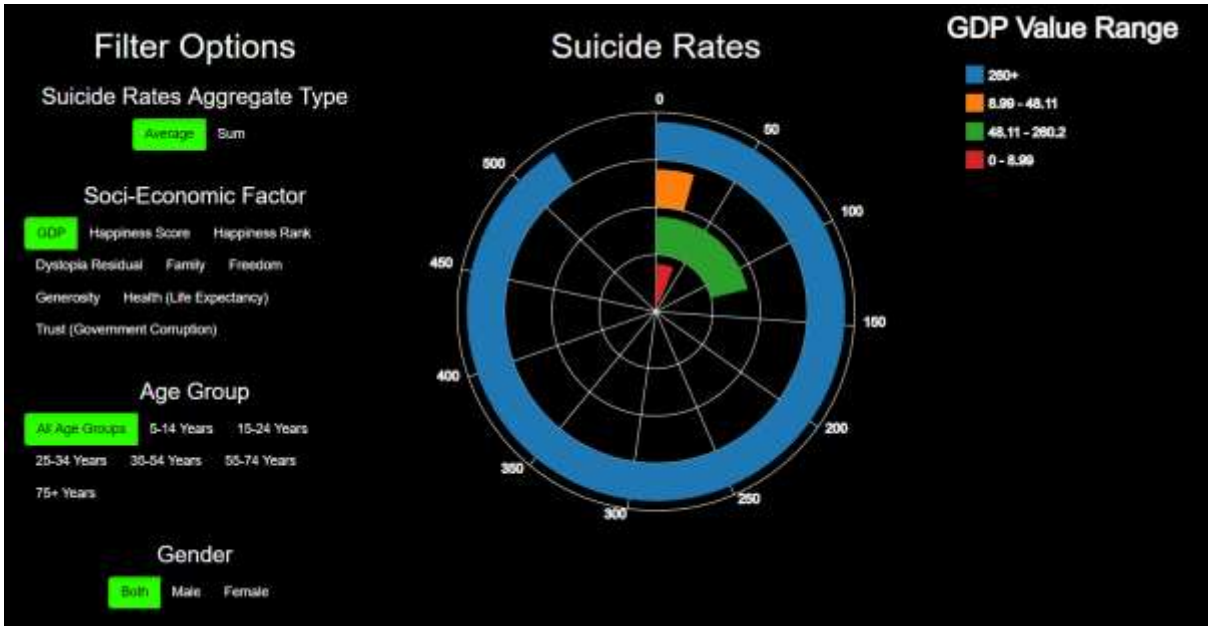


Figure 16

Here as we can see the filters are provided as buttons on the left side whereas the visualization for the respective filter has been provided on the right side of the page. Various filter options provided for socio economic factor wise visualisation are:

Suicide rates Aggregate type: same as gender page case

Socio-Economic Factor:

Figure 17

Based on the value of this filter the user can decide based on which Socio Economic factor data should be considered for visualisation. By default, GDP will be selected as shown in the Figure 17. If the user wishes to change it to some other factor the user can do so by clicking on desired factor once the factor have been selected it will be highlighted.

Gender: Same as Age Group page case

Age Group: same as gender page case

GDP Range (in billion $): same as gender page case

Suicide Rate Data Type: same as gender page case

Socio Economic Factor Wise Visualization:

Here Figure 18 represents the socio economic factor wise information regarding the suicide rates here each colour represents the respective value range. Also on hovering mouse over each of the radial bar exact value will be shown in the tooltip

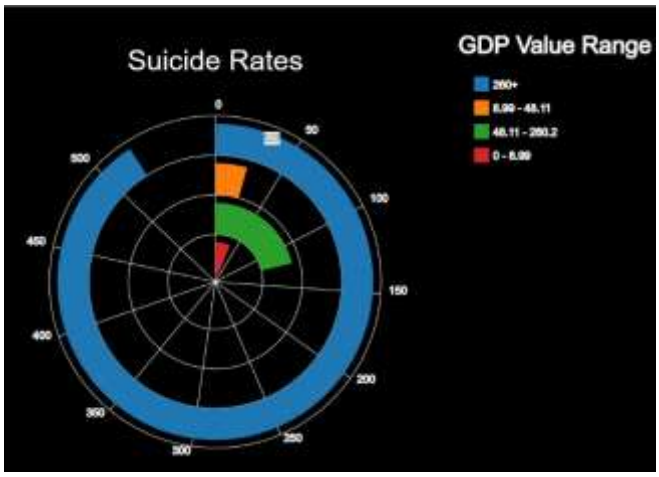


Figure 18

Note : in this case only the data of year 2015 and 2016 will be considered

Country Wise:

To go to this page, the user must click on the Socio Economic Factor Tab on the top most navigation bar as highlighted with a red circle in the figure 19 given below

Figure 19

Then the user will be taken to the Socio Economic Factor page. In this page the user can see the socio economic factor wise summary and trends regarding the suicide rates. This page looks like figure 20

Figure 20

Here as we can see the filters are provided as buttons on the left side whereas the visualization for the respective filter has been provided on the right side of the page. Various filter options provided for country wise visualisation are:

Suicide rates Aggregate type: same as gender page case

Gender: Same as Age Group page case

Age Group: same as gender page case

GDP Range (in billion $): same as gender page case

Suicide Rate Data Type: same as gender page case

Year:

Based on the value of this filter the user can decide which years information should be considered for visualisation. By default, All years will be selected as shown in the Figure 21. If the user wishes to change it and select just few years user can do so by clicking on desired years. All selected years will be Highlighted

Figure 21

Country wise Visualisation:

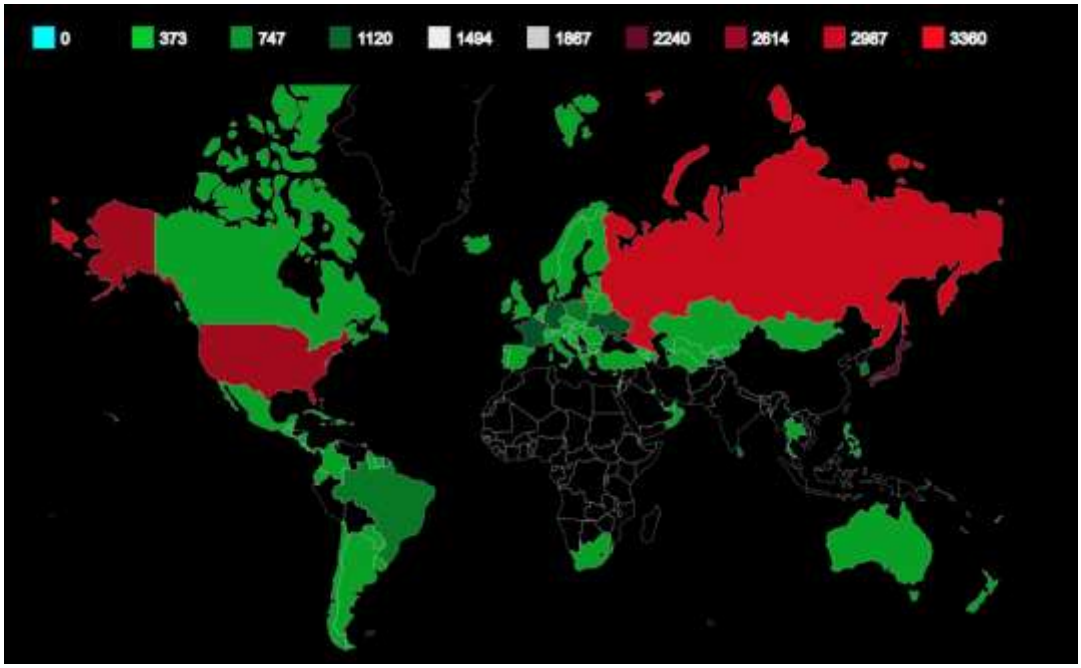


Figure 22

Here Figure 22 represents the country wise visualisation of the suicide rates. It is actually a heat map of countries with respect to their suicide rates with its colour indicating the no of suicides. Here the countries with light red colour has max no of suicides.

Over View:

To go to this page, the user must click on the Socio Economic Factor Tab on the top most navigation bar as highlighted with a red circle in the figure 23 given below

Figure 23

Then the user will be taken to generic overview page. In this page the user can see the visualization of all the categories that we have seen so far combined together. This page looks like figure 24

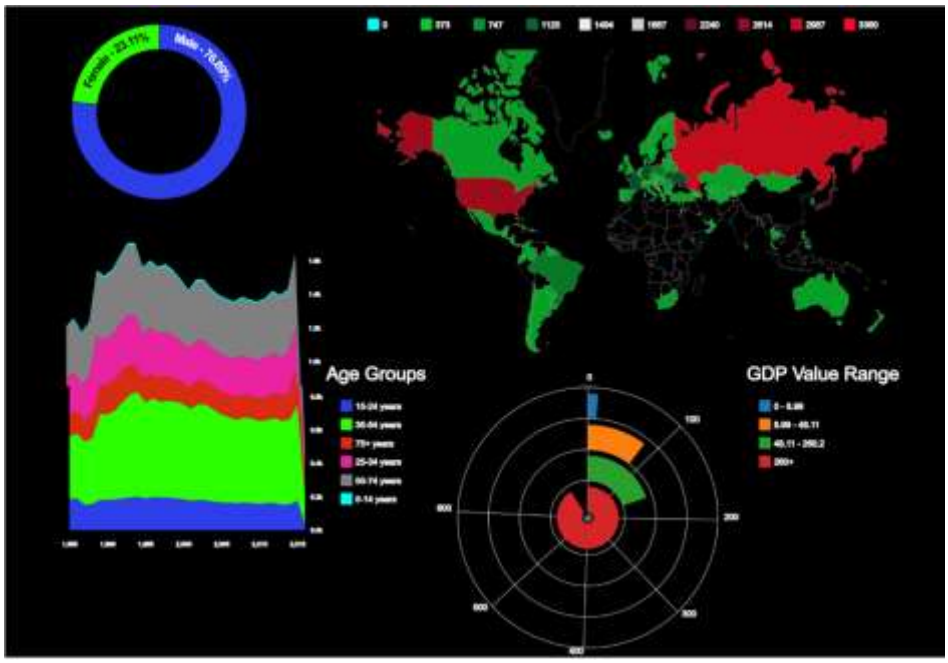


Figure 24

All the filter option which were previously available are available in this section as well. In this page if we click on a particular country

In this page if we click on a respective country. Then the information of that particular country only will be shown as in figure 25. You can select multiple countries by clicking on the desired countries.

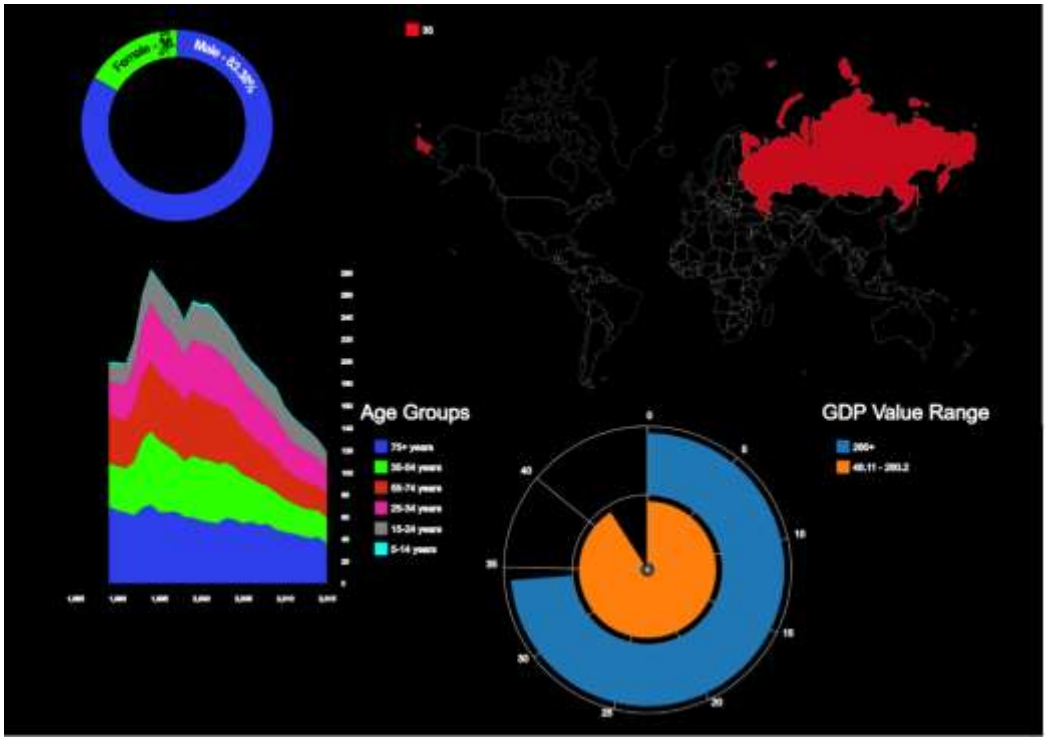


Figure 25

Also the name of the countries will be listed on the bottom of the page as in figure 26. There is also a button to 'show all countries' as in figure 26 which reset's the country based filter so that all user can see the suicide rates of all countries

Figure 26

Conclusion

From this project I was able to properly convey the various trends in Suicide rates based on various aspects. Out of which some of the conclusions are

- Countries with max average suicide no are Russia, US, Japan etc.
- Country that has max suicide no as of year 2016 is Thailand
- Country that has max suicide no per 100K population as of year 2016 is Lithuania
- Countries with max mean suicide no per 100K population are Lithuania, Russia, Srilanka etc
- 35 – 54 is the age group with max no of suicides for both male and female it also the age group with maximum population for both male and female.
- 75+ age group is the group with max no of suicides with respect to the population.
- In general, the no of males that suicide is more compared to that of the female.
- 1995 was the year with most no of suicides with respect to population.
- The general trend of no of suicides is decreasing over the period of time.
- Countries in the section of top GDP and happiness Rank has highest suicide rates. Which implies the countries with higher GDP has to take care of their citizens also.
- Maximum number of suicides per 100 K population is for countries with happiness rank less than 40. Which implies even if a country is considered to be a happy country still has lots of suicides

Reflection

From doing this project we will get a clear idea about how to choose an appropriate data visualization. We will also be able to implement interactive data visualizations using D3, html, CSS, JavaScript etc. In this project I could have considered various other factors like unemployment rate and its impact on suicide rates.

Reference

https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016
https://www.kaggle.com/unsdsn/world-happiness#2017.csv
https://www.who.int/mental_health/suicide-prevention/en/

Appendix

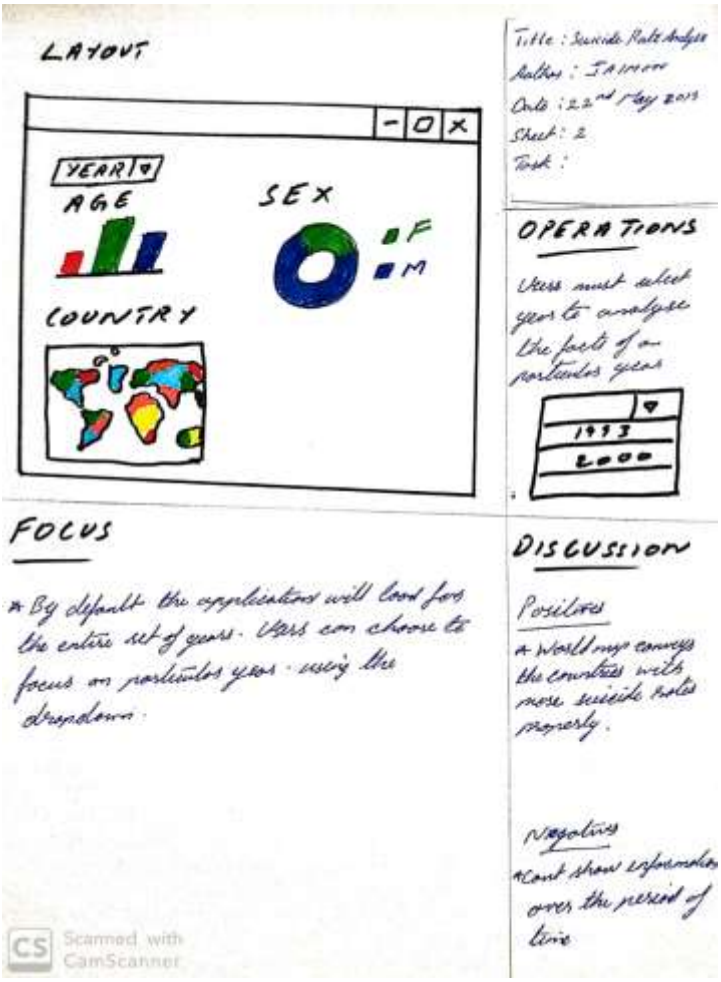
Data Source:

Suicide related data source for years 1985 - 2016
Type - tabular data (CSV)
URL - https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016
Spatial Info - 27.8k rows x 12 columns

Countrywise happiness rating data source for years in 2015 - 2017
Type - Tabular data (CSV)
URL - https://www.kaggle.com/unsdsn/world-happiness#2017.csv
Spatial Info - 2015 - 158 rows x 12 columns
2016 - 157 rows x 13 columns

Five Sheet Design:

Sheet 1



Sheet 2

Sheet 3

Sheet 4


LAYOUT
AGE
SEX
GDP
Suicide
Years
COUNTRY
SOCIO
YEARLY
COUNTRY
Title : Suicide Rate Analysis
Author : JAMSON
DATE : 22nd May 2019
Sheet : 5
TASK :
OPERATIONS
User can select desired years, country, socio economic factor
FOCUS
* By default all the statistics from all the countries are shown.
* The user can select a particular country and view its respective statistics like suicide rate for different age groups.
* The user can select a particular years and view its related statistics like suicide rate over the world.
* The user can select a particular GDP value and like high and view the suicide rates for that case over different age groups.
DISCUSSION
Positives
* Multiple country wise suicide rates conveyed beautifully.
* Trend of suicide rate for different age groups conveyed properly.
* Suicide rates for different GDP values & gender values conveyed properly.
Scanned with CamScanner

Sheet 5