

Does Country's Happiness Index (HI)/GDP affect Suicidal Deaths?

DATA EXPLORATION AND VISUALISATION

ARJUN BASU | 30387574 | FIT5147 Exploration Project

Table of Contents

1. Introduction.....	3
2. Data Wrangling.....	3
3. Data Exploration.....	6
4. Conclusion	14
5. Reflection	14
6. References	14

1. Introduction

Deaths by suicide has become the leading reason globally among various age groups and various ethnic and diverse communities (or, demography). The main motivation behind proceeding with this project is to find any sort of relationship between suicidal tendencies and happiness index (HI) or GDP of a country. Out of more curiosity, I would like to investigate further into the below questionable areas which I would like to figure out from tools like R and Tableau moving forward:

- Is there any effect on the rate of deaths by suicide by the components like GDPI and the Mental Health Disorder Index of a country?
- Does economy of a country depend on the happiness index of a country?
- Is there any link between how happy a citizen think themselves to be and the rate of suicides? Or in simpler words, the relation between Suicidal Death rates and the Happiness Index of a country.

2. Data Wrangling

Datasets chosen are:

- World Happiness Report of 2015
- Mental Health vs Suicide
- Suicide Rates (1985 – 2016)

On the basis of the below 7 factors, the happiness scores of 155 countries were given in the above datasets:

- Family
- Life expectancy
- Economy
- Generosity
- Trust in government
- Freedom and,
- Dystopia residual

These 7 factors when added up gave us the Happiness Score. The higher the value of the happiness score, better will be the happiness rank. The attribute, Dystopia, depicts the poor status/condition of a country in happiness level.

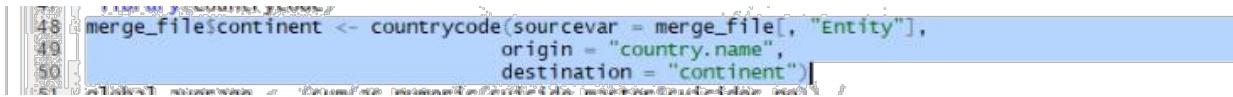
Below are the steps that has been done for data wrangling.

- Suicide Rates (1985 – 2016) data has 27820 x 12 records. Further cleaning/wrangling is done as per our need.

Following wrangling steps are done on the dataset:

1. The column, HDI, has very few values (lots of NULL) and as it can't be used for providing insights (unless consulted with domain experts), hence the column was removed.

2. Another attribute, Continent was added to analyse the data by continent using the library 'countrycode'.



```

48 merge_file$continent <- countrycode(sourcevar = merge_file[, "Entity"],
49                               origin = "country.name",
50                               destination = "continent")

```

3. The country Africa was removed, as its continent showed very less data.

4. Also, countries with data of very few years (i.e. below 3 years) have been removed.

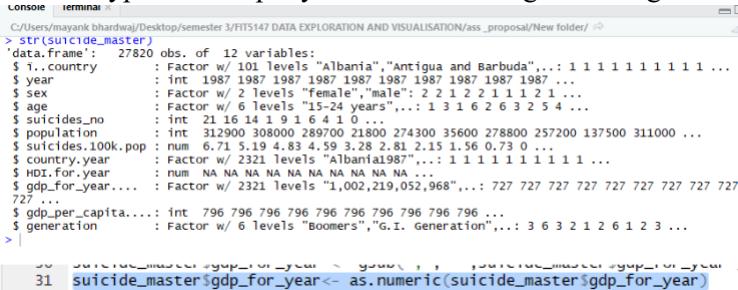


```

62 minimum_years <- suicide_master %>%
63   group_by(country) %>%
64   summarise(rows = n(),
65             years = rows / 12) %>%
66   arrange(years)
67
68 suicide_master <- suicide_master %>%
69   filter(!country %in% head(minimum_years$country, 7)))
70
71

```

5. The data type of GDP per year has been changed to integer from factor.



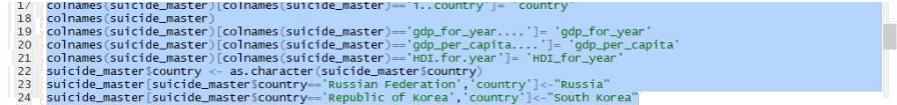
```

> str(suicide_master)
'data.frame': 27820 obs. of 12 variables:
 $ 1..country : Factor w/ 101 levels "Albania","Antigua and Barbuda"...: 1 1 1 1 1 1 1 1 1 1 ...
 $ year       : int  1987 1987 1987 1987 1987 1987 1987 1987 ...
 $ sex        : Factor w/ 2 levels "female","male": 2 2 1 2 2 1 1 2 1 ...
 $ age         : Factor w/ 6 levels "15-24 years"...: 1 3 1 6 2 6 3 2 5 4 ...
 $ suicides_no: int  21 16 14 1 9 1 6 4 1 0 ...
 $ population : num  312900 308000 289700 21800 274300 35600 278800 257200 137500 311000 ...
 $ suicides.100k.pop: num  6.71 5.19 4.83 4.59 3.28 2.81 2.15 1.56 0.73 0 ...
 $ country.year: Factor w/ 2321 Levels "Albania1987"...: 1 1 1 1 1 1 1 1 1 1 ...
 $ HDI.for.year: num  NA NA NA NA NA NA NA NA ...
 $ gdp_for.year...: Factor w/ 2321 Levels "1,002,219,052,968"...: 727 727 727 727 727 727 727 727 727 727 ...
 $ gdp_per_capita....: int  796 796 796 796 796 796 796 796 ...
 $ generation    : Factor w/ 6 Levels "Boomers","G.I. Generation"...: 3 6 3 2 1 2 6 1 2 3 ...

```

30 suicide_master\$gdp_for_year <- suicide_master\$gdp_for_year
31 suicide_master\$gdp_for_year <- as.numeric(suicide_master\$gdp_for_year)

6. Few column names like 1..country, gdp_for_year..., gdp_per_capita.... has been changed appropriately.

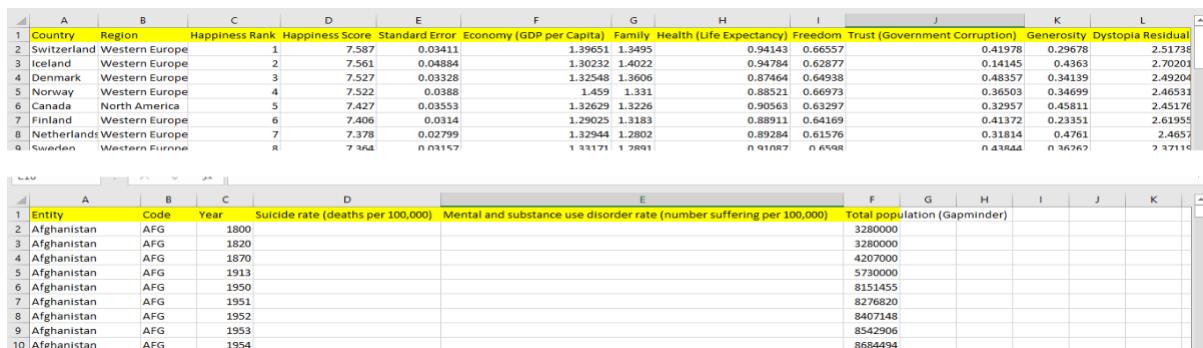


```

1/ colnames(suicide_master)[colnames(suicide_master) == 1..country] <- "country"
18 colnames(suicide_master)
19 colnames(suicide_master)[colnames(suicide_master) == "gdp_for_year...."] <- "gdp_for_year"
20 colnames(suicide_master)[colnames(suicide_master) == "gdp_per_capita...."] <- "gdp_per_capita"
21 colnames(suicide_master)[colnames(suicide_master) == "HDI_for.year"] <- "HDI_for_year"
22 suicide_master$country <- as.character(suicide_master$country)
23 suicide_master[suicide_master$country == "Russian Federation", "country"] <- "Russia"
24 suicide_master[suicide_master$country == "Republic of Korea", "country"] <- "South Korea"

```

- Secondly, the dataset (mental health vs suicide) has 21677 x 6 records and the world happiness report dataset has 158 x 12 entries.



Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity	Dystopia Residual
Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.3495	0.94143	0.66557	0.41978	0.29678	2.51738
Iceland	Western Europe	2	7.561	0.04884	1.30232	1.4022	0.94784	0.62877	0.14145	0.4363	2.70201
Denmark	Western Europe	3	7.527	0.03328	1.32548	1.3606	0.87464	0.64938	0.48357	0.34139	2.49204
Norway	Western Europe	4	7.522	0.0388	1.459	1.331	0.88521	0.66973	0.36503	0.34699	2.46531
Canada	North America	5	7.427	0.03553	1.32629	1.3226	0.90563	0.63297	0.32957	0.45811	2.45176
Finland	Western Europe	6	7.406	0.0314	1.29025	1.3183	0.88911	0.64169	0.41372	0.23351	2.61955
Netherlands	Western Europe	7	7.378	0.02799	1.32944	1.2802	0.89284	0.61576	0.31814	0.4761	2.4657
Sweden	Western Europe	8	7.364	0.03157	1.33171	1.2891	0.91087	0.6598	0.43844	0.34262	2.37116

Entity	Code	Year	Suicide rate (deaths per 100,000)	Mental and substance use disorder rate (number suffering per 100,000)	Total population (Gapminder)
Afghanistan	AFG	1800			3280000
Afghanistan	AFG	1820			3280000
Afghanistan	AFG	1870			4207000
Afghanistan	AFG	1913			5730000
Afghanistan	AFG	1950			8151455
Afghanistan	AFG	1951			8276820
Afghanistan	AFG	1952			8407148
Afghanistan	AFG	1953			8542906
Afghanistan	AFG	1954			8684494

Below steps are taken for wrangling the data:

- Records with null values in the mental_disorder_rate_100k column was removed.

```
123: happiness_temp<-happiness_temp[-3]
124: happiness_temp<-na.omit(happiness_temp)
```

- The Total_population attribute has too many null values, so it was replaced with suicide data set total population.

```
53: for(i in 1:nrow(final_file)){
54:   if(is.na(final_file[i,"Total_population"])){
55:     final_file[i,"Total_population"] = final_file[i,"x"]
56:   }
57: }
```

- Column names like ‘Suicide.rate..deaths.per.100.000.’, ‘Total.population..Gap minder.’, ‘Mental.and.substance.use.disorder.rate..number.suffering.per.100.000.’ were changed appropriately in the suicide rate dataset.

```
27: colnames(suicide_rate)
28: colnames(suicide_rate)[colnames(suicide_rate)=="suicide.rate..deaths.per.100.000."]= "suicide_rate_100k"
29: colnames(suicide_rate)[colnames(suicide_rate)=="Mental.and.substance.use.disorder.rate..number.suffering.per.100.000."]
30: colnames(suicide_rate)[colnames(suicide_rate)=="Total.population..Gapminder."]= "Total_population"
```

- In the happiness dataset, columns like ‘Trust..Government.Corruption.’, ‘Economy..GDP.per.Capita’. ‘Health.. Life.Expectancy’ were changed appropriately.

```
27: colnames(happiness)
28: colnames(happiness)[colnames(happiness)=="Economy..GDP.per.Capita."]= "gdp_per_capita"
29: colnames(happiness)[colnames(happiness)=="Health..Life.Expectancy."]= "Life_Expectancy"
30: colnames(happiness)[colnames(happiness)=="Trust..Government.Corruption."]= "Trust_government"
```

3. Data Exploration

Various data exploration techniques were used to gain valuable insights from the three datasets.

To begin with, I checked the correlation among the different numerical variables. From the correlation, I found the below insights depending on which the happiness of a country is influenced:

- We found that Happiness Rank is inversely proportional to all the 7 factors.
- Of all the 7 factors, GDP of a country, life expectancy and family strongly influences the happiness rank of a country, while factors like generosity and trust have very little impact.

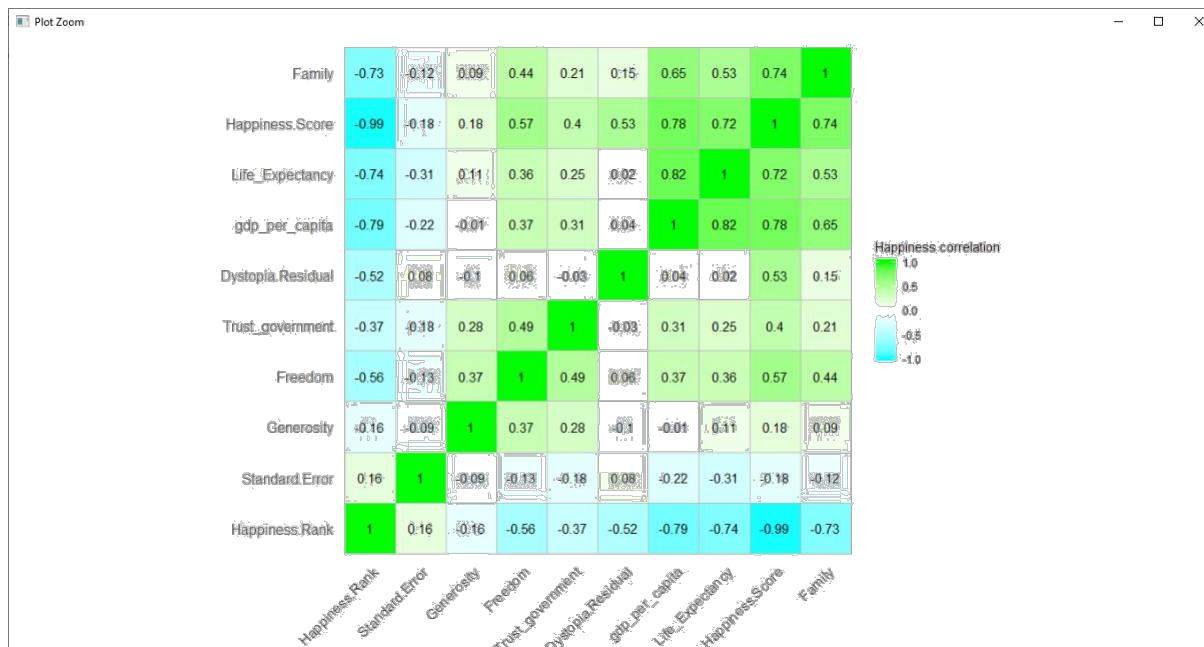


FIGURE:1 Correlation matrix between different numeric columns of happiness dataset.

Data from the continent Oceania were not included in the below plot as it consisted of only 2 countries. The below plot in figure 2 is the scatter plot (with smoother line) between GDP_per_capita and happiness_score_in_2015. We can see that as the GDP increases, the happiness also increases.

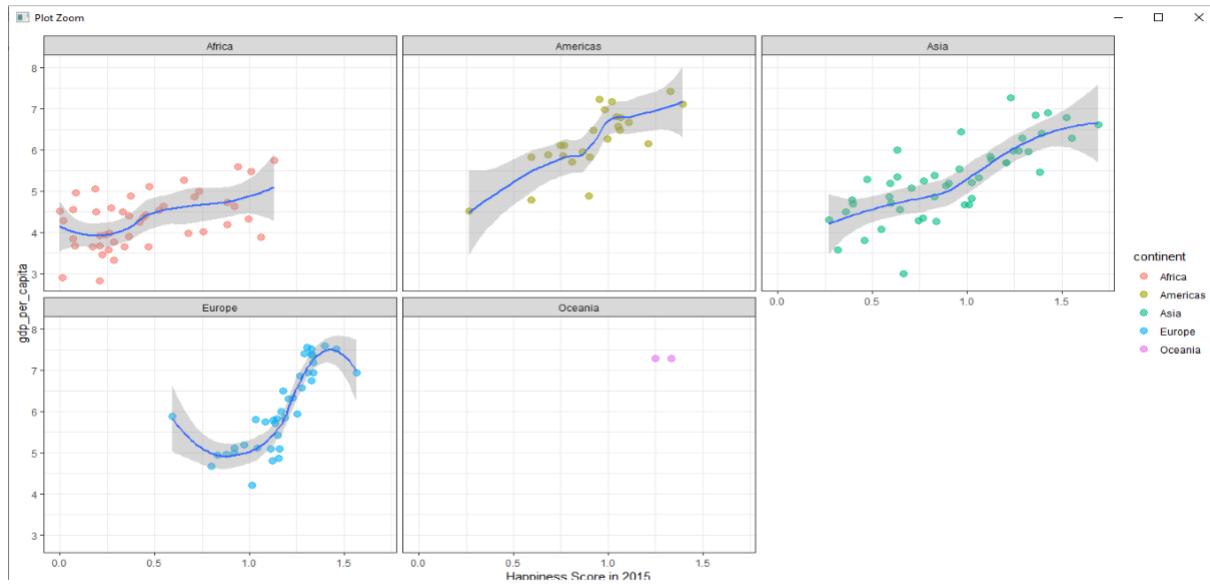


Figure:2 Scatter plot with smoother line with respect to GDP and Happiness score.

The below figure shows the correlation between life expectancy and happiness score among the different continents. The above said correlation is more significant in Europe, America, Asia than in any other continents. The attribute Freedom is strongly correlated to happiness in Europe, Asia, America than any other continents. Africa shows a weak correlation between trust and happiness score. When Smoother line for generosity is applied, it shows a positive direction only in Europe, horizontal for Asia and downward direction for Africa.

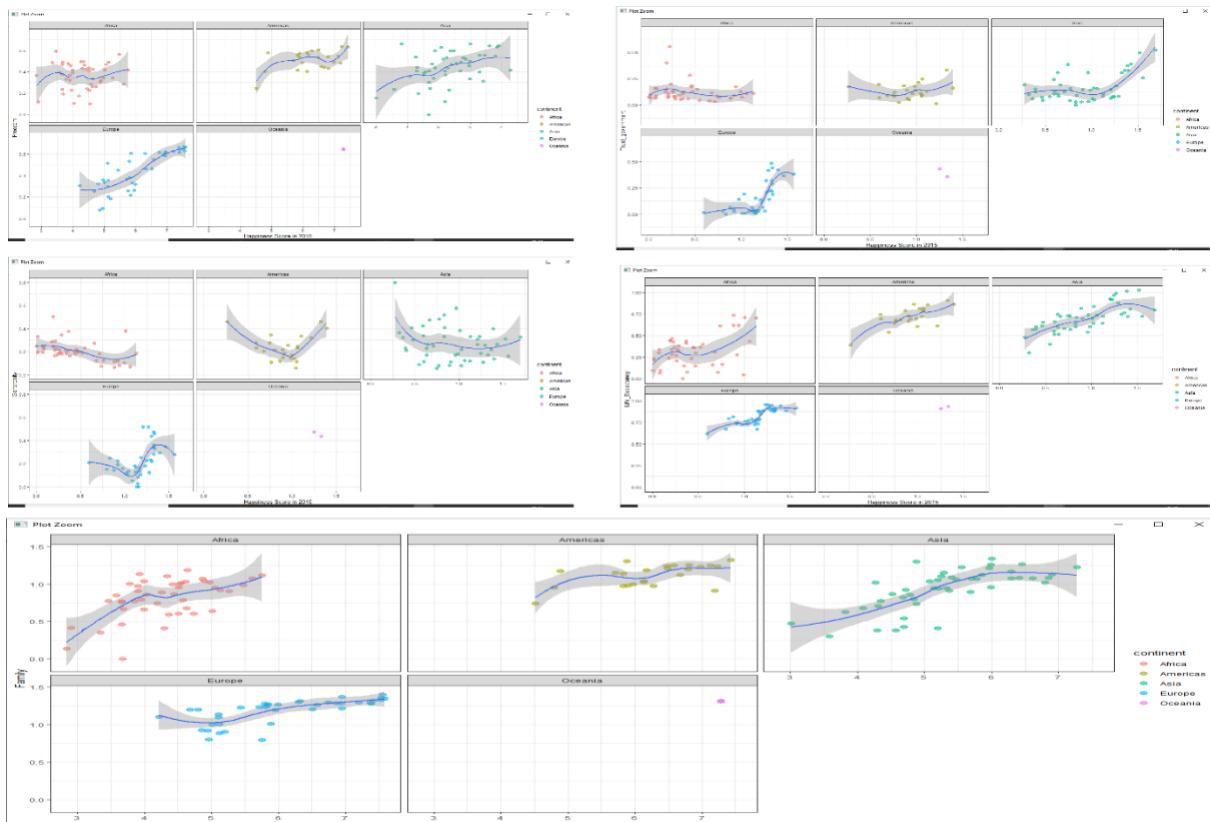


Figure 3: Scatter plot with smoother line for happiness score vs Freedom, Family, Trust_government, Life_Expectancy, generosity.

The below figures 4 and 5 (termed as SHAPE OF HAPPINESS) visualises the degree of happiness of a country through the polar plot of top 10 and bottom 10 countries.

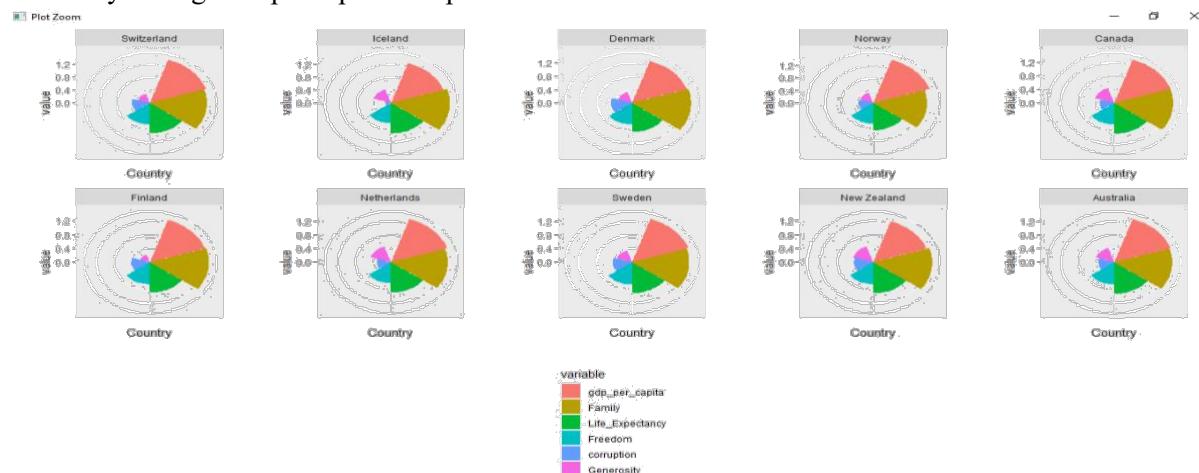


Figure 4: Polar representation of 6 attributes vs happiness for Top 10 countries

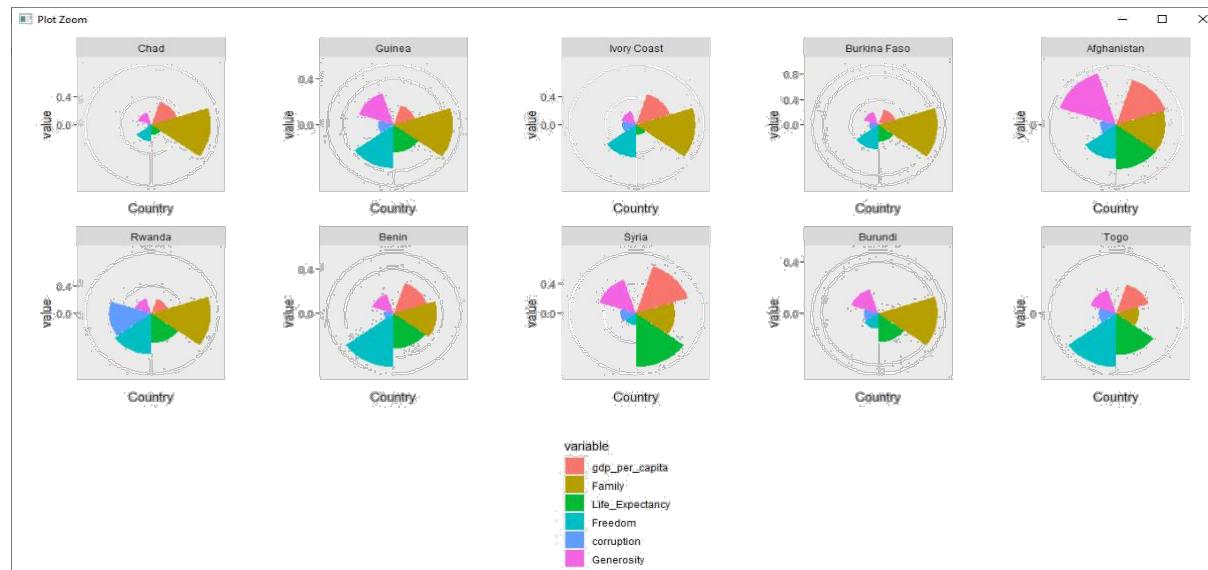


Figure 5: Polar representation of 6 attributes vs happiness for Bottom 10 countries

However, there was strange scenario captured for Singapore and Hong Kong. Despite having similar characteristics as other top 10 countries, still they show low happiness and ranked 24 and 72 respectively. There might be other factors which are affecting the happiness score in Singapore and Hong Kong (Figure:6).

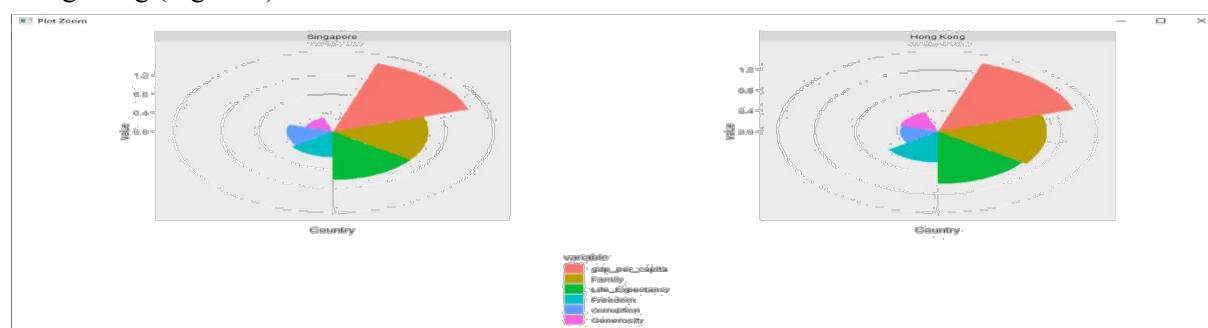


Figure 6: Polar representation of 6 attributes vs happiness for Singapore and Hong Kong country

From general perspective, there should be a relation between suicide and mental/substance abuse. Lets see through various plots through data collected from WHO site. Below are the two plots showing the variation of suicides over the world map (figure 7) and mental disorder present over the world (Figure 8).

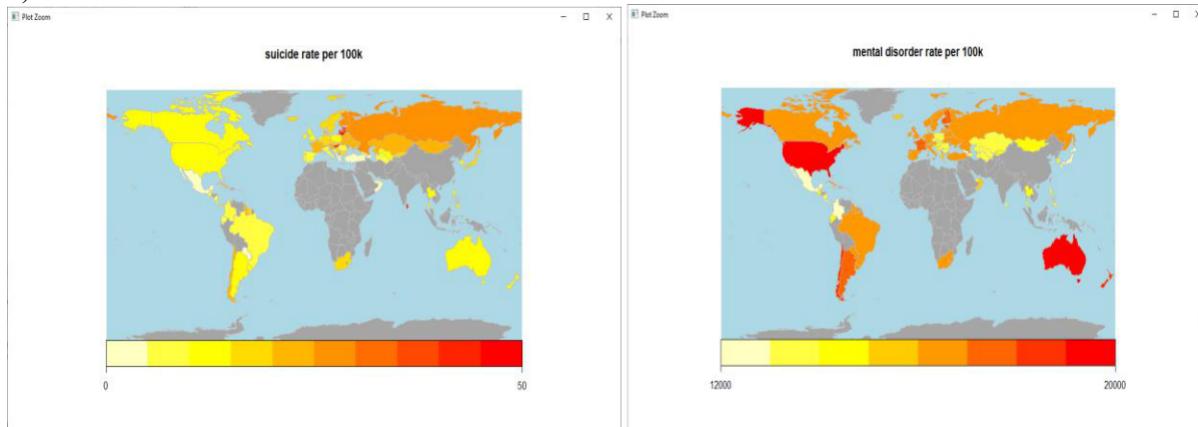


Figure 7: Heat map of suicide rate per 100k

Figure 8: Heat map of world for mental illness.

From the heat map above, we can see that Sri Lanka and few of the European countries have the highest suicide rates (suicides per 100k) followed by the Russia, Kazakhstan. Australia and America have the least suicide rates. The areas in grey in the above heat maps suggests that there was no data available from those countries/continents. Countries like Australia, America and Alaska have the highest mental disorder (from figure 8 above) followed by Argentina, Russia, South Africa, Kazakhstan and Mongolia. As the heat map is providing the overall overall status of the data over the world, it is difficult to find useful insights from it. Hence, I plotted the graph between suicides and mental disorder over the year for top 10 suicidal rate countries and immediately we are getting a useful insight from it. We can see that `suicide_rate_100k` is changing while `mental_disorder_rate_100k` is static which depicts that there is no clear evidence in supporting a relationship between suicide rates and mental disorder rates.

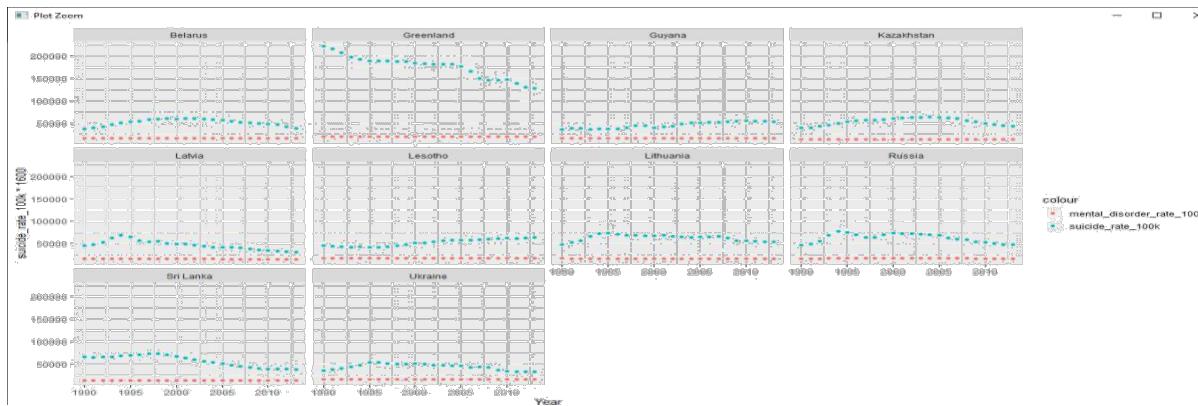


Figure 9: Relation plot between Suicides, mental disorder vs year.

To further analyse on the relationship between suicide rates and mental disorder rates, I plotted a correlation graph between the two and it shows a weak correlation among them again suggesting that there is no clear evidence in supporting a relationship between suicide rates and mental disorder rates.

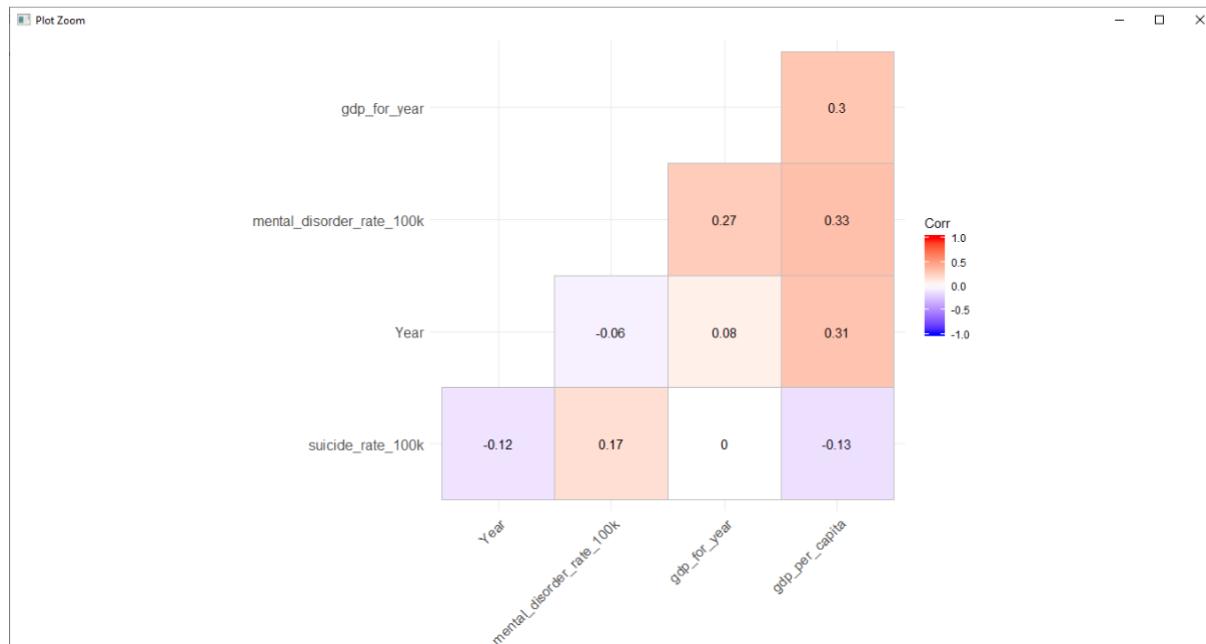


Figure 10: Correlation in suicides vs mental disorder.

Few interesting insights were found from the plot of suicide data from round the world between 1985 – 2015.

In the below figure 11, it can be seen that the overall number of suicides have decreased over the years. The year 1995 showed the peak value of suicide number and the a gradual decrease and reaching below the average line from 2006.

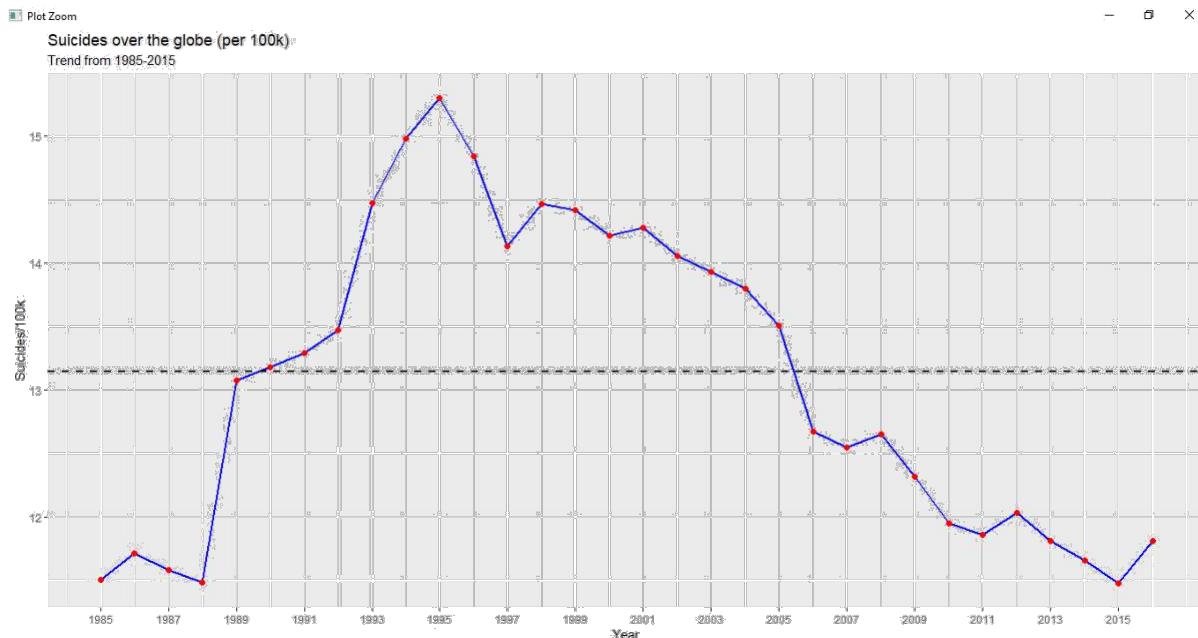


Figure 11: Suicides over the globe from 1985 till 2016.

The below figure shows the point plot (with smoother line) of suicide and happiness score for a country. It can be observed from the plot that increase in happiness also increase the suicide rates. Strange, right? The suicide rate was increasing gradually till happiness score reached 6 and then decreased from then onwards.

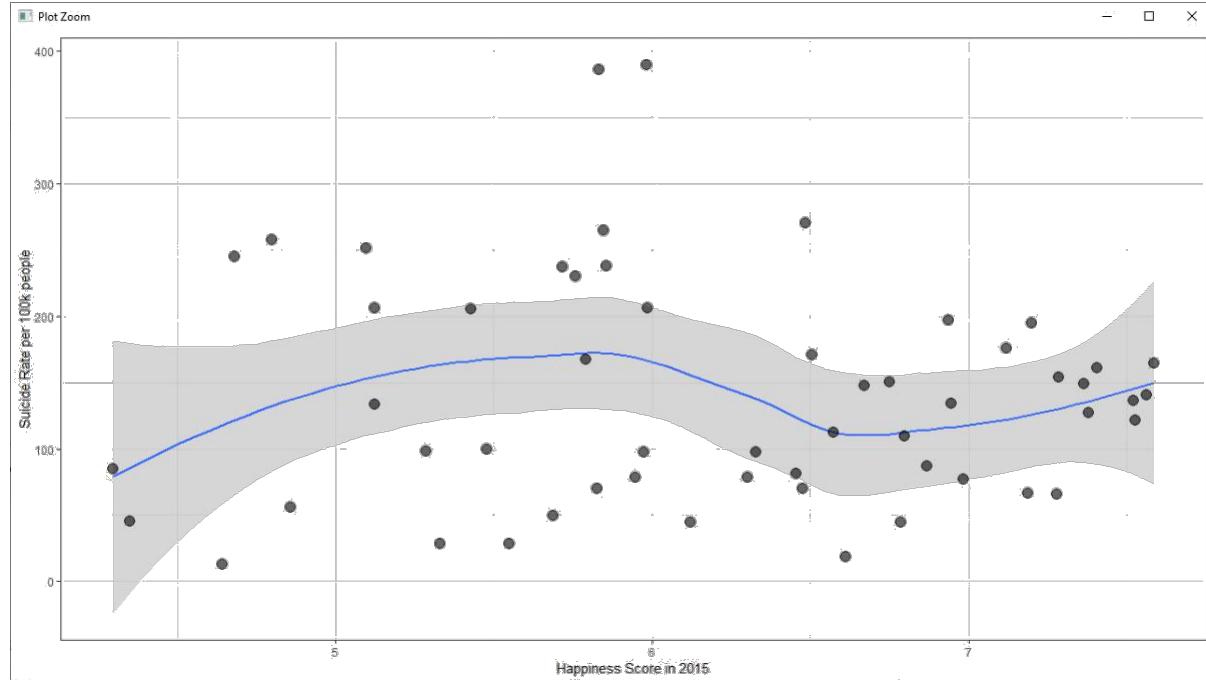


Figure 12: Suicide rate vs happiness for world.

In the below figure, I plotted the graph between suicide rate and happiness for male and female. It was observed that for females, suicide rate is independent of the happiness score. However, for male suicide rate is positively proportional to happiness score till happiness score reached approximately the value 6, after that suicide rate is negatively proportional to happiness score.

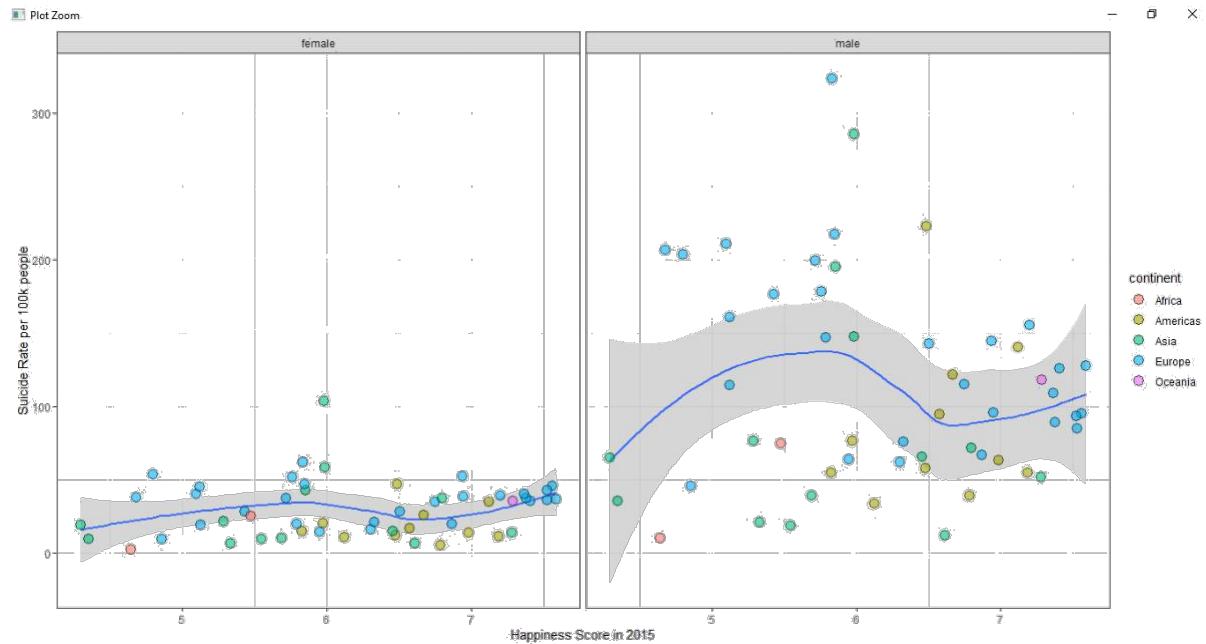
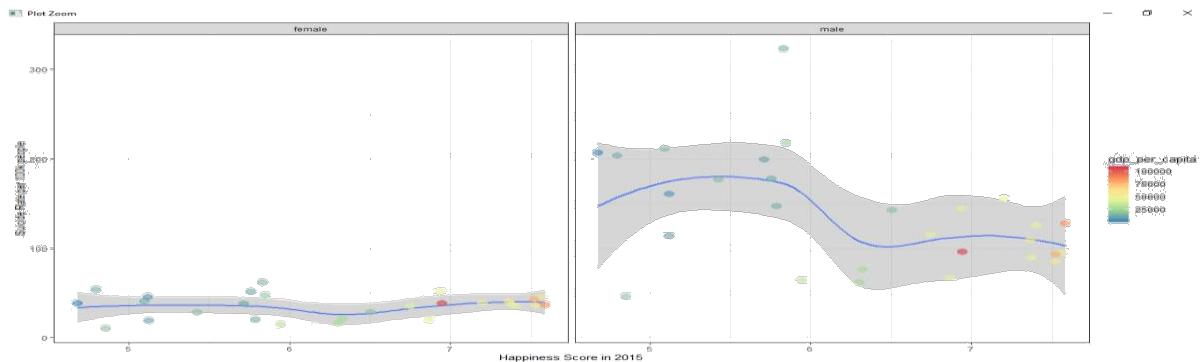
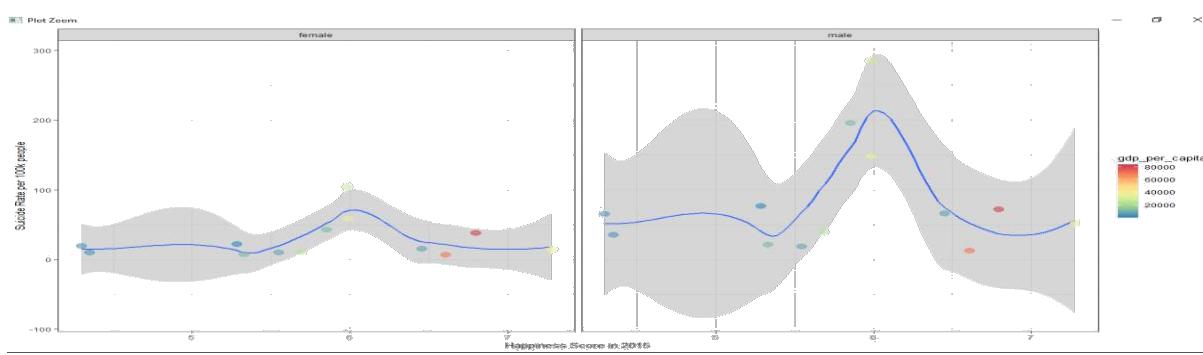
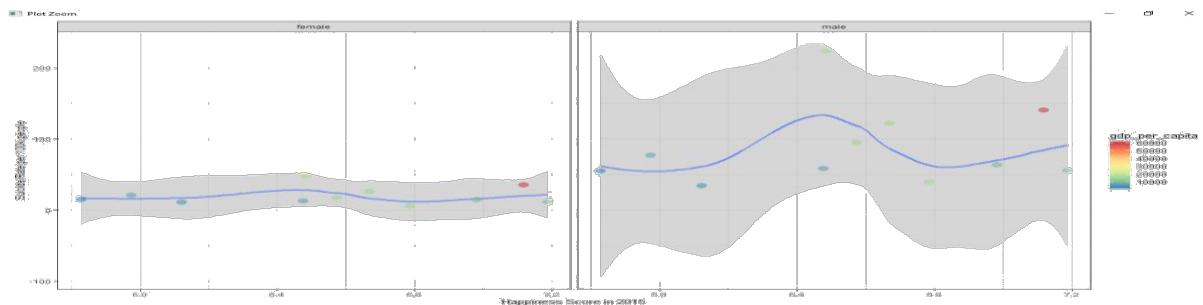


Figure 13: Female and male plot with suicide vs happiness.

Now more detailed analysis was done where we picked the data of male and female's suicide and happiness score of different in the year 2015. As usual, for female, there is no significant relationship between suicide and happiness score. However, for males, it was observed that there is a strong correlation between suicide rate and happiest countries. It was also observed that the suicide numbers in happiest countries are twice as much as it is in non-happy country, which is very much strange! It was found that Finland has the suicide rate of 13.43 with happiest countries rank 6. On the other hand, Italy with happiest countries rank 50 has a suicide rate of 6.53, almost half of that of Finland. Yes, very strange!

Europe:

Asia:

Americas:


Africa: (not much data available from Africa, hence plot is not useful)

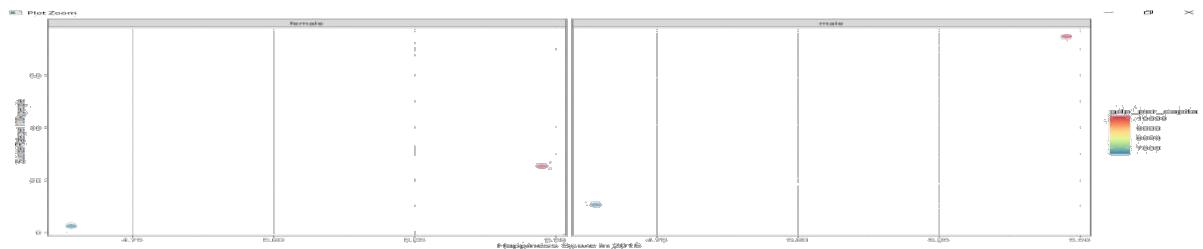


Figure 14: Suicide vs happiness for male and female for different continents.

What about suicide numbers by age group? Below, I have plotted suicide number graph of male and females and of different age groups. It can be observed that female suicide rates are lower than male, suggesting that females are much more emotionally stable than male. It can also be observed that the suicide rate is decreasing over the years for both male and female. However, the age group of 5-14 sees an increase in the suicide numbers, which is very disturbing as it is such an early age group. Also, the age group of 75+ males have a high suicide rate compared to the other age groups.

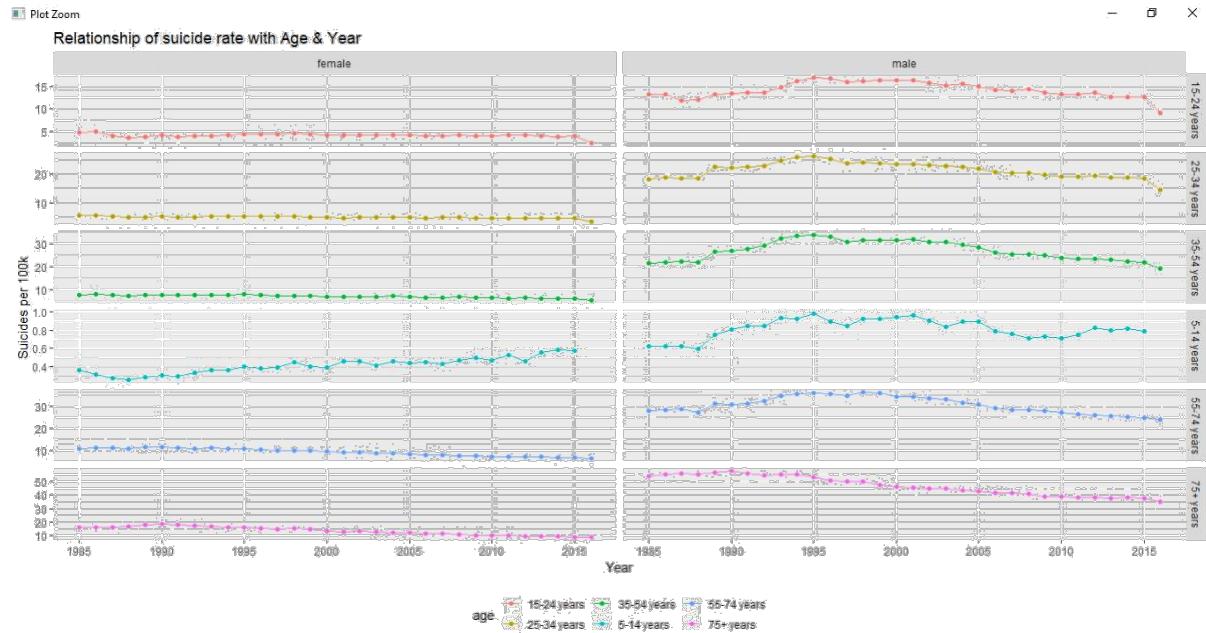


Figure 15: Suicide Rate over the Years for different Age groups of Male and Female

After diving further into the analysis, I found out from the below plot that the suicide rate increases with increasing GDP. Surprising, yes? However, one might argue that the below graph doesn't show a strong correlation between Suicide rate and GDP. However, it can be seen that suicide rates are much higher in rich countries than in poor (3rd world) countries. One of the reasons might be that the people in wealthy nations tend to be more depressed than those in poor countries, as per the saying: 'Money can't buy happiness'. Also it can be inferred that, as you get more wealthier, more types of problems arises, which are unforeseen.

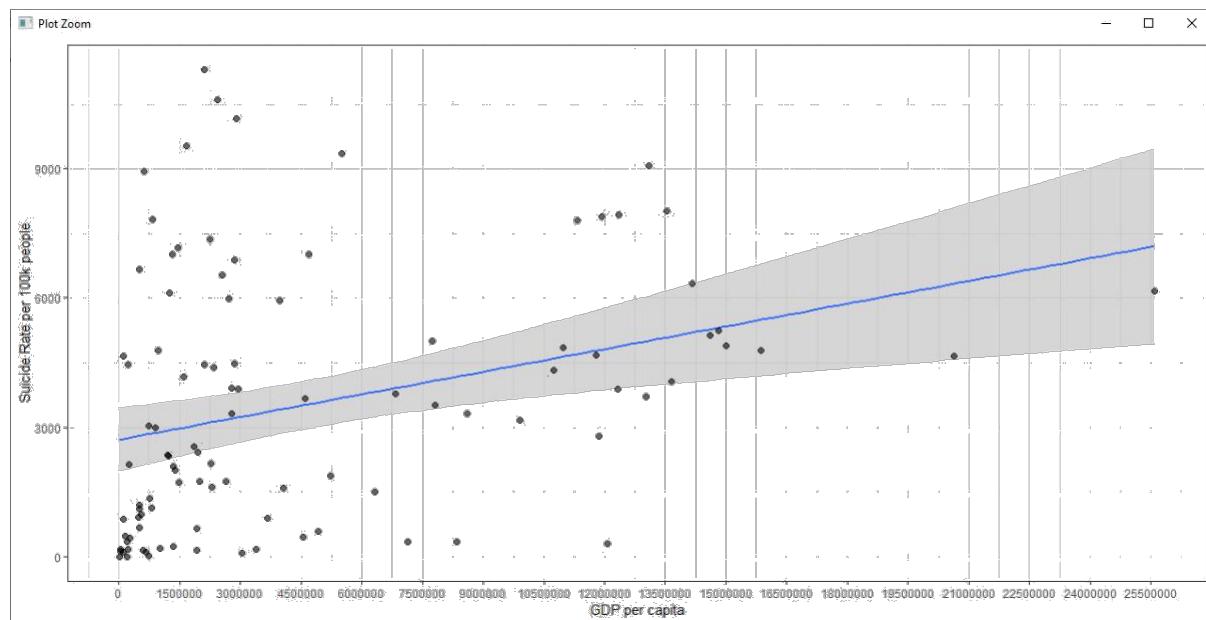


Figure 16: GDP per Capita vs Suicide Rate

4. Conclusion

To conclude this, I R and its different packages to be very useful for these kinds of detailed analysis. The answers to my 3 before stated questions were answered through this analysis. I found how the happiness score of a country and the suicide rate (suicide numbers) are associated with different factors. Through the analysis, it can be observed that the happier a country is, more will be the suicide rate, which completely shattered my previous generalised thinking. The analysis also broke my earlier thinking that more the GDP of a country is, the lesser will be the suicide rate but it is actually the opposite. Another interesting aspect was found that there is no strong evidence to support the relation between mental (emotional) health of a person for the suicide and GDP of a country. However though, happiness score of country largely depends on the life expectancy and family and generosity. But, Trust has very little impact on happiness score. To conclude I can now say, the wealthier a person becomes, he/she faces more issues and problems and hence, suicide rates increase.

5. Reflection

- I learned the various usage of the tool R and its packages (especially the visualisation functions).
- I learned that for correct analysis and findings, one must wrangle the data properly. The whole analysis and the conclusion will depend hugely on the quality of wrangling done on the data.
- I have learned the real usage of visualisation. I have learned how a graph can tell us more about the data and giving us a clear picture as you create more and fine-tuned plots. I have learned the usage of different kinds of plots and choose the best plot which fits best. For example, I used scatter 3D plot to show the relation between GDP, Family and Life expectancy.
- I have found the usage of my thinking skills. This is one area which cannot be trained by anyone. It segregates you from an average tech person. It completely depends on us as how to move forward with the analyse and it can only be achieved if you keep on asking questions from the project to yourself.
- The analysis still needs to be revised if provided with a fine-tuned data set. For example, statistics from different countries separately. Also, I believe that I could have searched more for a more recent dataset.
- Overall, it was a very interesting project to do, as it shattered many of my generalised thinking and I was happy to change the ideas of my fellow classmates, friends and family on how Suicide rate depends on GDP/Happiness of a country.

6. References

1. <https://ourworldindata.org/mental-health#link-between-mental-health-and-suicide>
2. <https://www.kaggle.com/unssdn/world-happiness>
3. [https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016\(master.csv\)](https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016(master.csv))
4. <http://www.sthda.com/english/wiki/ggplot2-quick-correlation-matrix-heatmap-r-software-and-data-visualization>
5. <https://rdrr.io/rforge/rworldmap/man/mapCountryData.html>
6. <https://www.kaggle.com/crizzi/are-happy-people-really-happy>
7. <https://www.rdocumentation.org/packages/car/versions/3.0-2/topics/scatter3d>
8. <https://www.kaggle.com/ashkash247/who-suicide-statistics-aesthetic-eda>
9. <https://www.kaggle.com/hanyan/visualizing-happiness>
10. <https://www.kaggle.com/ksvrd1/cluster-analysis-on-suicide-victims>