**STAT 3345.501 Team Project**

*Analyzing Social and Economic relation*

*with Suicide rates*

**Team Comet 13**

**Jacob Oates, Mehedi Toufiqe, Sriniketh Potlapalli**

**April 8, 2020**

**1. Introduction**

This report will give an overview of our project and discuss social and economic relations with suicide rates. Our project was to practice using the R language and RStudio to analyze a large set of data and observe interesting findings from the data set. For our project, we use a large data set from a publicly accessible website, “Kaggle.com”. The name of the data is “Suicide Rates Overview 1985 to 2016” [1]. The dataset contains information from three different references, the World Health Organization, the United Nations Development Program, and the World Bank, which provides information on the details of those who committed suicide from 1985 to 2016. Our dataset contains 27,820 total observations and 12 variables as follows: *country* (name of the country), *year* (indicating year), *sex* (categorized by two genders), *age* (age categorized by years), *suicides\_no (*number of suicides), *population* (number of population), *suicides\_100k\_pop* (number of suicides per 100,000 population), *country\_year* ( country associated with the year), *HDI\_for\_year* ( HDI score for specific year), gdp\_for\_year ( GDP score for specific year), *gdp\_per\_capita* ( GDP score of per capita), *generation* ( name of the generation). One important fact to note is that this dataset does not contain information for all countries, so our findings are only concerning the countries found in this dataset only.

Suicide is defined as the act of intentionally causing one’s death. The most shocking statistics regarding suicide show that suicide is the 10th leading cause of death worldwide. Concerns about suicide in recent years include a 56% increase in deaths by suicide by those aged 12-24 from Generation-Z [2]. The question that comes to our mind almost immediately after hearing this is “What contributes to this large amount of suicides and why are they increasing every year? Are most developed countries are having less suicide? Since our data contains Human Development Score (HDI) scores in every country, it will be interesting, not to mention useful, in finding some analysis. The Human Development Index (HDI) is a composite index of life expectancy, education, and per capita income indicators, which are used to rank countries into four tiers of human development [3]. Another question that comes to mind when looking at this dataset is the effect of factors such as sex, generation, GDP’s effect on suicide? We start with cleaning the data and selecting a specific subset for plotting our images. We came up with six different major questions to analyze the social and economic relation to suicide. We answer each question in detail and explicitly state our findings from the plots. At the end of the report, on the conclusion we summarize the report and broadly discuss the analysis.

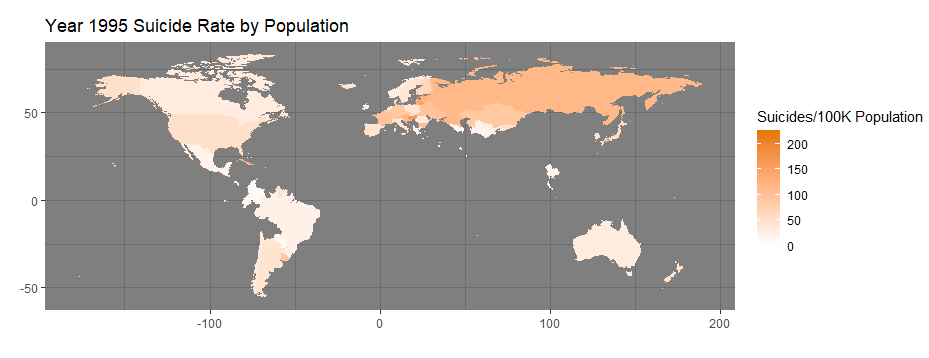
**2. Data Cleaning**

This dataset was pretty clean when we found it. Though there were just a couple things we had to change around to make it work for us. We started off by renaming the columns of the data to best suit our definitions of each column. Then we added a column to the data to put the countries into their respective regions. After that we defied the column’s string so that we could use the data properly. Then we filtered the data every five years, starting in 1995, for the world map graph. Lastly, all we had to do was define the four different decades that the data was in.

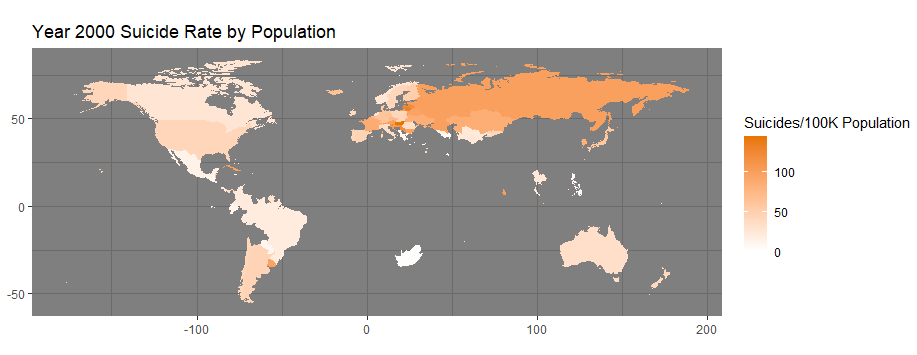
**3. Questions and Findings**

In this section, we will be analyzing the data by answering 6 major questions. For each question, a graph was made to best represent the data relevant to the question, and to show correlation between different columns of the data. We discussed and thought of our questions over many days and arrived at these: Which countries have the greatest suicide rates by population? Is there a relationship between the number of suicides and the GDP per capita and does this relations depend on which generation the victims of suicide belonged to? Which generation is more prone to suicide? How have the trends of suicides changed over time in the United States by gender? Which countries have the greatest suicide counts? Do GDP per capita and suicide rates have any impact on the top countries with the highest Human Development Index (HDI) score?

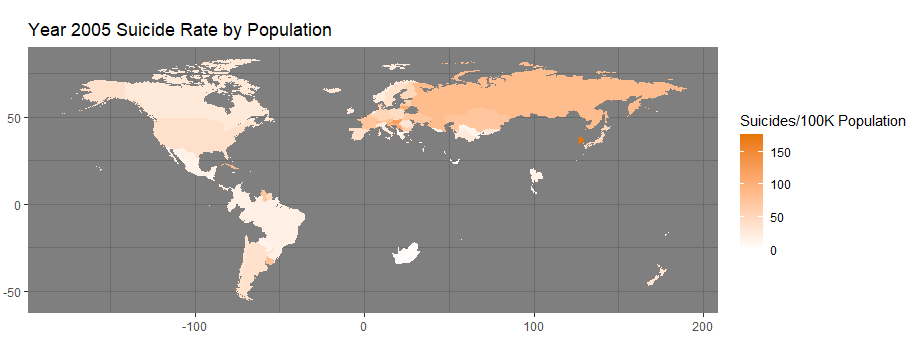
**3.1 Which countries have the greatest suicide rates by population?**

****

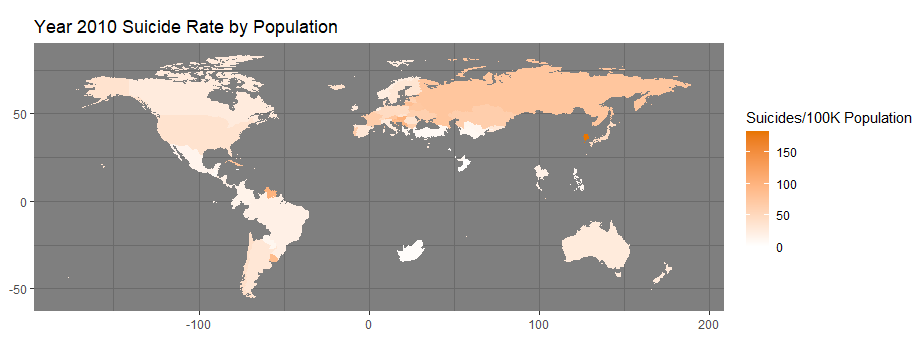
*Figure 3.1.a: Year 1995 Suicide Rate (Suicides/100K Population)*

****

*Figure 3.1.b:Year 2000 Suicide Rate (Suicides/100K Population)*

****

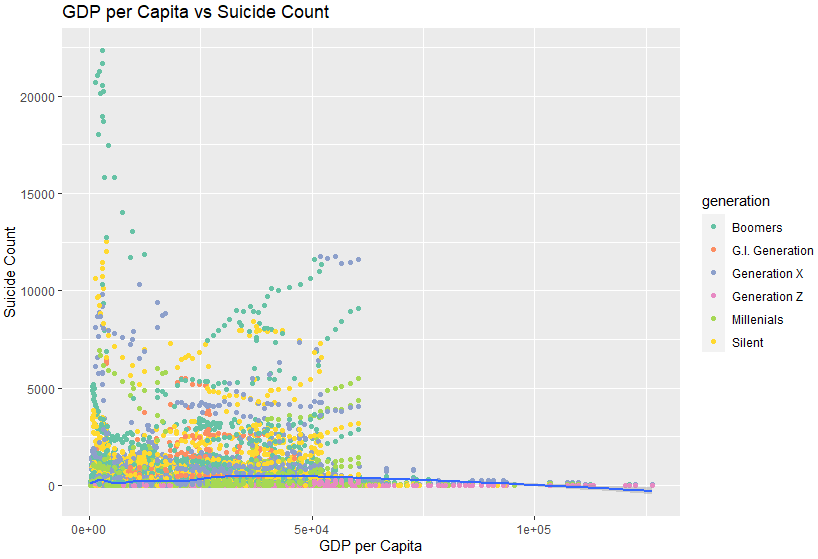
*Figure 3.1.c: Year 2005 Suicide Rate (Suicides/100K Population)*

****

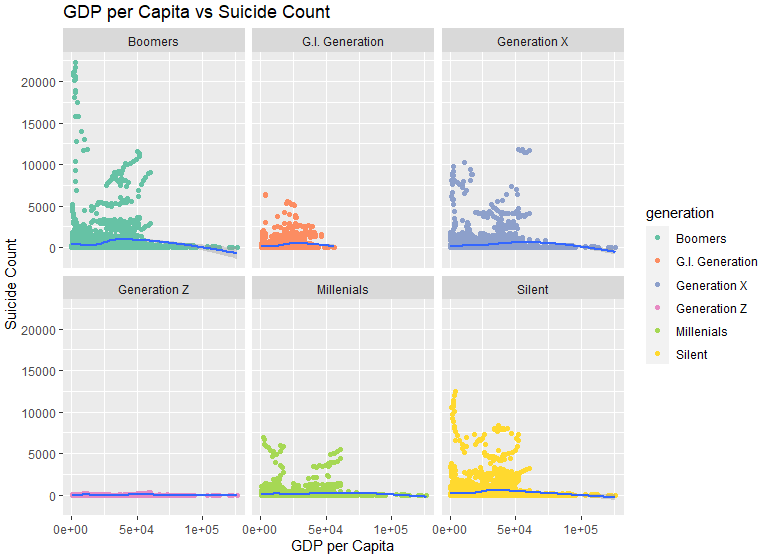
*Figure 3.1.d: Year 2010 Suicide Rate (Suicides/100K Population)*

Figures 3.1.a-d each show a snapshot of the suicides per 100K population by a specific year. Notice how some countries do not show up. This is because our dataset did not have information for all countries; it only provided us with 101 out of 253 regions of the world (these numbers were obtained via the country column of the suicide data and the region column of the world map data). While we had the option to append another dataset onto this one to complete the graph, we ultimately decided not to after realizing that these datasets did not contain the same amount/level of information as our original one. For example, only our dataset showed how many suicides were attributed to each generation, age, etc. Looking at the Figures 3.1.a-d, we can immediately tell, by paying attention to the scale and the colors on the map, that the suicide numbers per 100K population have been steadily decreasing fom 1995 - 2010. This may be due to the fact that the economy is getting better over time, which could indicate a lower amount of suicides, but it is too early to tell. While the graph gives us information on the suicide rates by population, it does not indicate the sheer number of suicides. To analyze this, we must look more in depth into the data.

**3.2 Is there a relationship between the number of suicides and GDP per capita? Does this relationship depend on which generation the victims of suicide belonged to?**



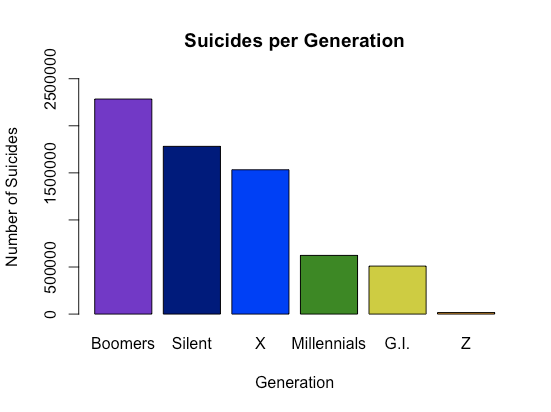
*Figure 3.2.a: Plot of GDP per Capita vs Suicide Count*



*Figure 3.2.b: Plot of GDP per Capita vs Suicide Count, Faceted by Generation*

Figure 3.2.a & 3.2.b both display a scatterplot showing the relationship between the number of suicides and GDP per capita, with Figure 3.2.a not being faceted so you can see the data as a whole, compared to Figure 3.2.b, where we can separate the data by generation and compare them to each other. The Figure 3.2.a reveals that there is a slight negative correlation between the number of suicides and the GDP per capita, while Figure 3.2b showed that there was a slight negative correlation only for select generations, while showing that the GDP per capita impacted suicide numbers significantly for others. Another interesting observation we have made is that a higher GDP per Capita will only indicate a decreasing number of suicides after a specific GDP range. In fact, the suicide numbers could even rise until they reach the favorable range. This fact was especially surprising, considering that we thought that a higher GDP would indicate a lower amount of suicides, due to factors including better infrastructure in the country and quality of living, and that this would be universal across all GDP per Capita, not just above a certain number. Interestingly, we see some distinct outliers for most of the generations. For example, the Boomer generation had abnormally high suicide numbers between the range of approximately 1000 - 1500 GDP per Capita. This may be due to times of economic stress in countries that fall within this range of GDP.

**3.3 Which generation is more prone to suicide?**

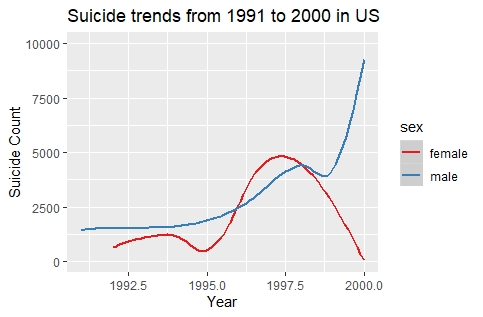
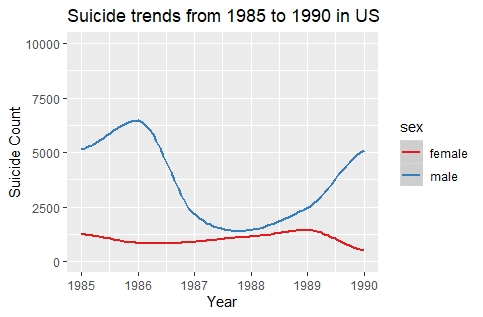
****

*Figure 3.3: Boxplot of Suicides per Generation*

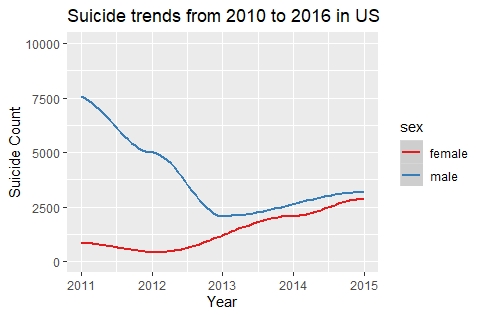
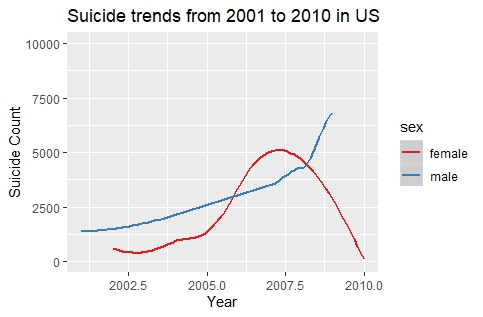
This graph shows the relationship between the generations and the amount of recorded suicides from the years 1985 to 2016. The graph almost goes in order of the generation. The Silent Generation (Silent) did come before the Baby Boomer Generation (Boomers), however the majority of the Boomers suicides were from Russia between 1990-2009, which was right after the Cold War and the fall of communism. With the Boomers growing up, the greater part of their life under communism could have been a factor in their choice to commit suicide as their way of life was coming to an end. The Greatest Generation (G.I.) and iGen/Genz (Z) are the oldest and youngest generations on this graph, so it was no surprise that they had the least amount of suicides compared to the other generations. The youngest G.I.’s in this data were 61 years old in 1985, which is the main factor that led them to being the second lowest. While the oldest Z’s are just turning 21 in 2016 and the youngest turning 4, that would lead them to being the smallest in this graph. Then lastly we have Generation X (X) and the Millennials (Millennials). X came before the Millennials so it’s no surprise to see them one place above the Millennials.

In conclusion, the two generations on the extreme opposites of this dataset have the lowest number of suicides. The two generations that were completely adults (above 21) in this dataset had the most number of suicides recorded. The two generations that were born slightly before or during this dataset’s time period had the third and fourth most number of suicides.

**3.4 How have the trends of suicides changed over time in the United States by gender?**

****

*Fig 3.4.a Fig 3.4.b*

****

*Fig 3.4.c Fig 3.4.d*

*Figure 3.4 : Plot of Suicides counts among gender in the United States*

Figures 3.4.a-d show the trends of suicides among male and female over the decades in the United States. On Fig.3.4.a: The plot is showing the suicide trend among male and female from 1985 to 1990. It’s very surprising to see that male suicide count is also higher than the female. In 1986, there is an increase in suicide counts in males. Among female suicide count, the trend is pretty stable. Only small increases in 1989. Since our data only have the value from 1985, so it’s hard to analyze the suicide trend in the ’80s.

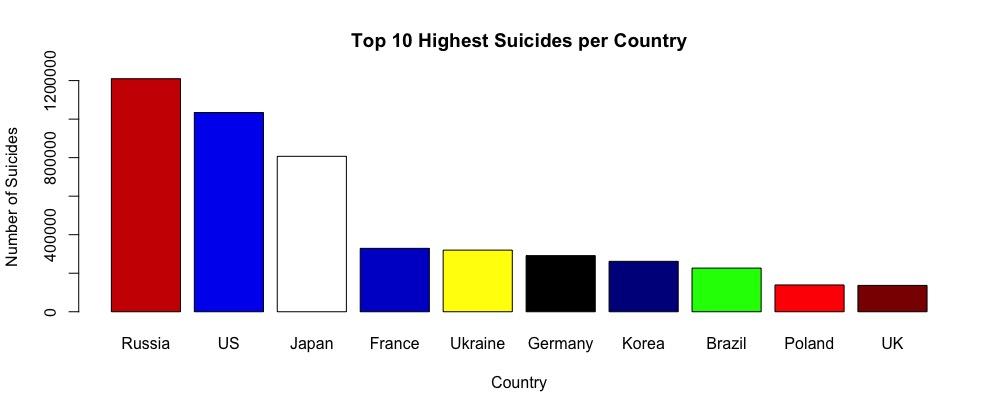
In Fig.3.4.b: The plot is showing the suicide trend among male and female from 1991 to 2000. As same as the previous decade, the male suicide count is increasing. From 1995, the female suicide count is increasing and dropping in the late ’90s. In 1997, the female suicide count overshadows male suicide count. But in the late 90’s because of political and economical swift, we see that male suicide count increasing rapidly. Also, in the late ’90s because of women's rights and health movement, we see that female suicide count decreasing.

In Fig.3.4.c: The plot is showing the suicide trend among male and female from 2001 to 2010. As same as the previous decade, the male suicide count is increasing every year. In 2008, the female suicide count overshadows male suicide count. Because of the economic recession in 2008 in the United States, we see that a spike in suicide rates both among males and females. But from 2009, the female suicide count drops while the male suicide count is rising.

In Fig.3.4.d: The plot is showing the suicide trend among male and female from 2010 to 2016. The suicide trend among male makes huge drops in 2013. Because of political and economic development, the male suicide trend finally drops. But surprisingly, the female suicide trend rises in 2013. Since 2013, the female suicide count rising and the male suicide count is also rising. Since our data only value is only limited to 2016, so it’s very hard to analyze the suicide trend in recent time.

We can finally come to conclude that from 1985 the male suicide count is always rising except in 2013 and the female suicide count neither increasing nor decreasing constantly. We also see that because of an economic, social issue in a specific year does play a huge role in the suicide trend. So, the trends of suicides changed over time in the United States by gender could be affected by socio-economic issues.

**3.5 Which countries have the greatest suicide counts?**

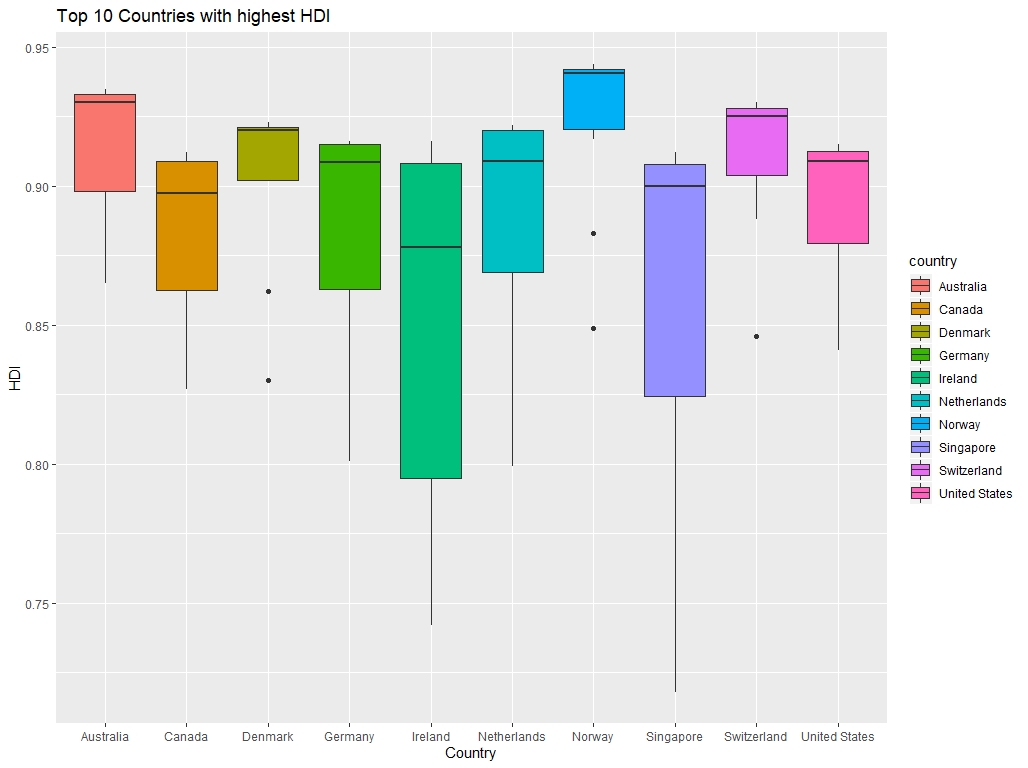


*Figure 3.5: Boxplot of Top 10 Highest Suicides per Country*

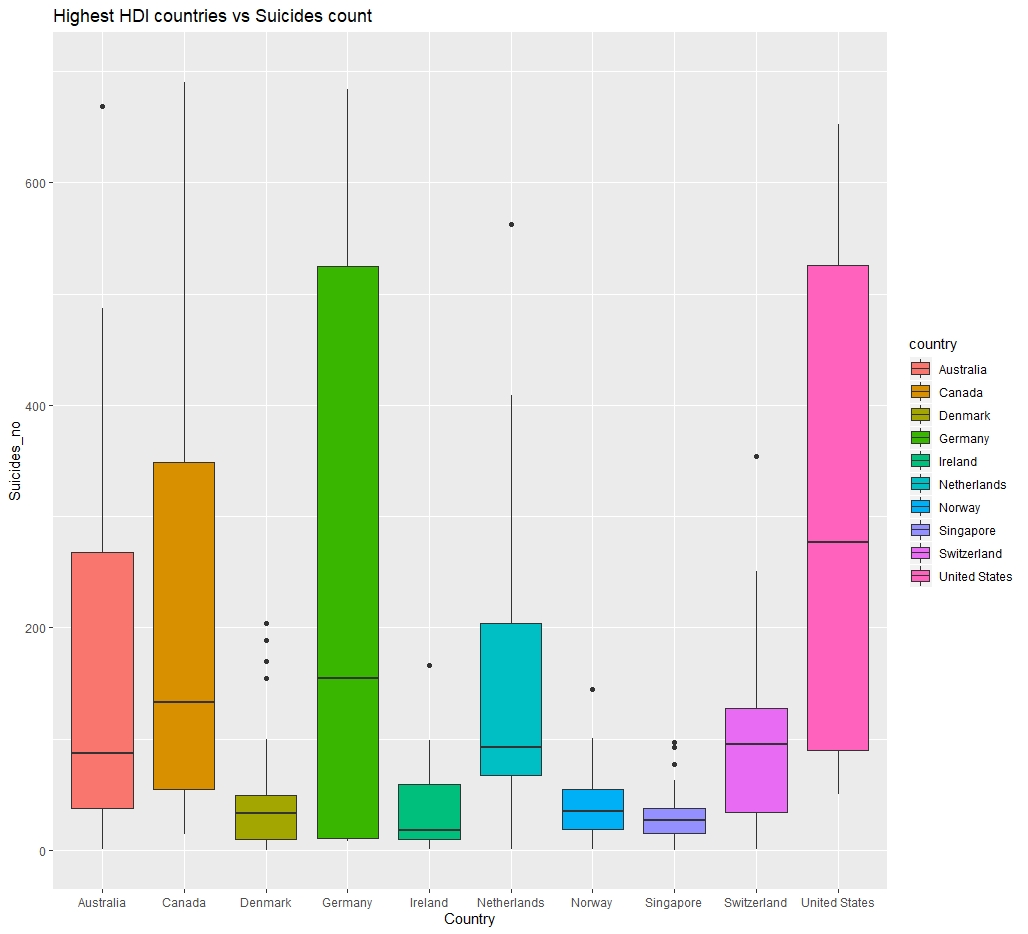
The above boxplot shows the top 10 countries with the highest number of suicides recorded in the dataset. As you can see by the graph, Russia is the top country, and followed by the United States, Japan, France, Ukraine, Germany, Korea, Brazil, Poland, and the United Kingdom. Russia being at the top isn’t super surprising because of the loss in the Cold War and the death of communism during this time period. The United States (US) also isn’t a huge surprise either because they were the other world super power during this time period. However, from the graph, Japan’s number of suicides is also pretty high compared to the rest of the countries on the list. Most of Japan’s highest sucide numbers fall in the 21st century. After doing some research outside the dataset, this steep increase in suicide numbers is due to over population of the country and higher demands at the place of work. These lead to health issues which is the main cause of suicide. The rest of this list, minus Poland and the United Kingdom (UK), have similar numbers recorded. I believe this is due to overpopulation at different time periods during this data set. The UK and Poland had similar numbers compared to a lot of countries left off this graph, barely edging them out to be in the top 10.

Overall, a generic summary of this graph could be the countries with the highest populations during this time period could correlate to the higher number of suicides. However, other countries with high population numbers, such as China and India, weren’t on this list. As you can tell from Figure 3.1.a-d, these countries were left off the list and were not recorded due to them being absent from our dataset. This leads to the data being skewed towards the medium populations countries making this graph.

**3.6 Do GDP per capita and suicide rates have any impact on the top countries with the highest Human Development Index (HDI) score?**

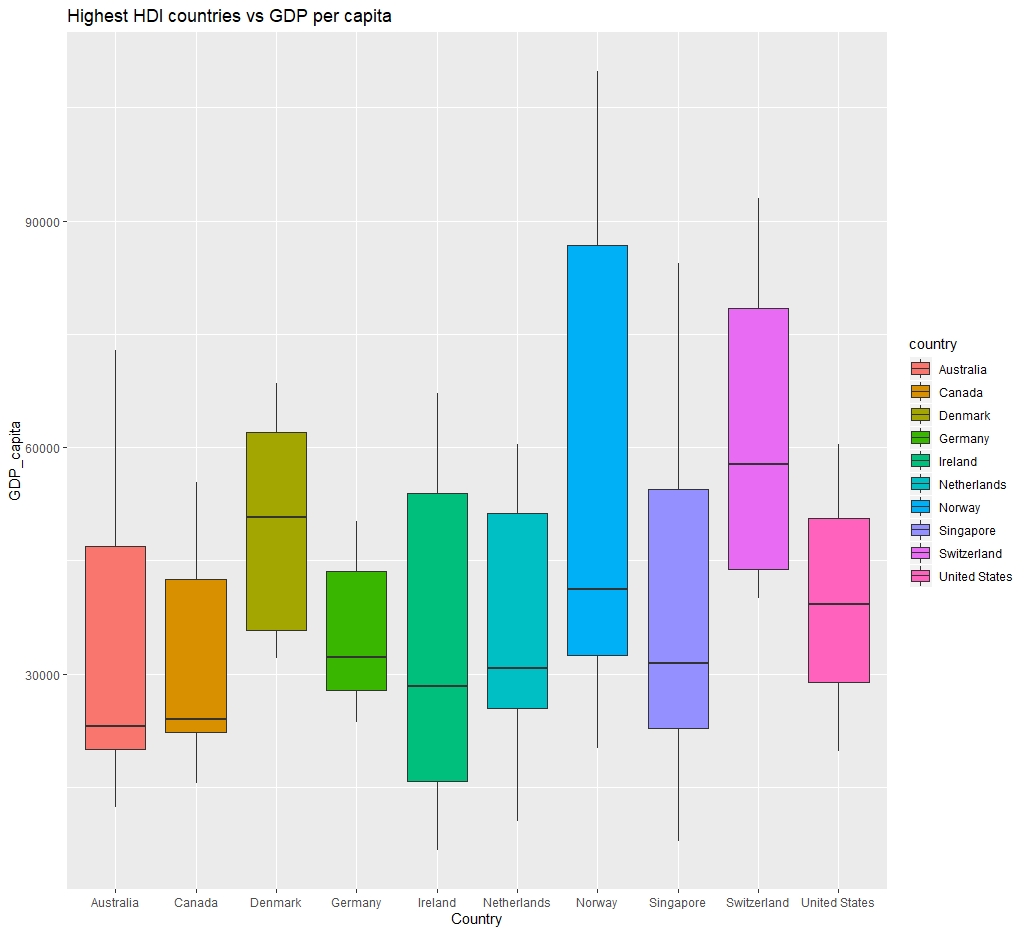
*Figure 3.6.a : Boxplot of Top 10 countries with the highest HDI*

The above figure shows the boxplots of the top 10 countries with the highest HDI. The countries are as follows: Australia, Canada, Germany, Ireland, Netherlands, Norway, Singapore, Switzerland, and the United States. As we can see, the countries are appearing in alphabetical order rather in increasing or decreasing order. Because some countries are similar flag colors, we choose a different color pattern to identify these ten countries. From the plot, we can see that Norway has the highest HDI and Ireland is the lowest HDI among others. The top three are; Norway, Australia, and Switzerland. Denmark, Germany, Netherlands and the United States are at the middle level. Canada, Singapore, and Ireland are at a lower level.

**

*Figure 3.6.b : Boxplot of the highest HDI countries vs suicide counts*

The above figure shows that boxplots of the top 10 countries with the highest HDI compared with suicides count. As we see, the United States has the highest number of suicide counts followed by Germany, Canada. In the middle level, we see that four countries have a similar number of suicide counts. Those countries are; Australia, the Netherlands, and Switzerland. Denmark, Norway, and Singapore have similar suicide counts at the lower level. Ireland has the least suicide count. If we compare Fig.3.6. a and Fig.3.6. b, we get some interesting results. Ireland has the least HDI among other countries but they also have fewer suicide counts. Norway has the highest HDI and also, they have fewer suicide counts.

**

*Figure 3.6.c : Boxplots of the highest HDI countries vs GDP per capita*

The above figure shows the boxplots of the top 10 countries with the highest HDI compared with GDP per capita. GDP means “Gross Domestic Product” that measures everything that a country produces in a year. The GDP implies a country's overall economic situation. GDP per capita is a measurement of the total economic output of a country divided by the number of people. It's used to compare the standard of living between countries and over time [4]. From Fig.3.6.c, we can see that Switzerland has the highest GDP per capita and Australia has the lowest GDP per capita. Denmark has the second-highest GDP per capita. Norway and the United States have similar GDP per capita. Germany, Netherlands, Singapore, and Ireland are at the middle level. Australia and Canada.

If we compare Fig3.6.a-c we find some interesting results; Switzerland, Denmark, and Norway have a high HDI score and they also have high GDP per capita score. Their suicides count is also lower than in other countries. If we look onto the United States, Germany, and Canada, they have a large amount of suicides. It is very important to understand that their geographic location and population is also larger. But on the scale of HDI and GDP per capita, the United States, Germany, and Canada scores are poor. Australia, Netherland, and Singapore have high HDI but their GDP per capita score is poor and Australia both have smaller suicide counts. We can say that, despite country population, Switzerland, Denmark and Norway has a stable social and economic system and because of that their suicide counts are low. The United States, Germany, Canada are in Top highest HDI but their GDP per capita score is low compared to other countries. Finally, GDP per capita and suicide rates have an impact on suicide rates with the top ten highest Human Development Index (HDI) score.

**4. Additional Findings**

**4.1 How HDI is changing between the top five countries from early 1900 to recent time?**

we can analyze the countries with the highest HDI vs corresponding year. From 1990 to the recent decade four countries; Norway, Australia, Switzerland, and Netherlands are constantly increasing. White Denmark HDI is also increasing but, in some years, the score is decreasing. Not only each country's HDI is increasing but also their GDP per capita is high [Fig.3.6.c]. And their suicide rate is constantly low. Again, this makes sense, because countries with high HDI score and high GDP per capita have lower suicide rates compared to other countries.

**4.2 Is there a relationship between the number of suicides and GDP per year? Does this relationship depend on which generation the victims of suicide belonged to?**

Our findings for this question were consistent with our findings from question 3.2. This makes sense because GDP per year should be some multiple of GDP per capita, since they both use GDP in their calculation. One interesting thing to note, however, is the fact that the graph has most of its points to the leftmost side of the graph. Again, this makes sense, especially after considering the fact that the GDP per year is lower than the GDP per capita.

**5. Conclusion**

After our thorough analysis of the dataset, we extrapolated many interesting findings. First, we observed that the suicide rates were actually decreasing as time was passing. We attributed this to likely be the byproduct of increasingly favorable economic factors. Next, we see that as the GDP per capita is increasing, the amount of suicides decrease, albeit with the caveat that the amount of suicides could actually start increasing in the lower to mid range of GDP per capita. Then, we found the amount of suicides by generation. However, we also realized that while this data is valuable, it does not really take the oldest and youngest generation that much into account, due to the limited amount of different ages of those who belonged to those two generations in the dataset. For the next three questions, we look at some of the top countries which have the most importance in this dataset. We find that in the United States, male suicide counts are generally higher than female suicide counts. The male suicides have been almost constantly increasing, whereas the female suicides over time varies quite a lot. Next, we see the top 10 countries in suicide counts. While analyzing each of the countries, we see some countries, such as japan, who have had a higher suicide rate in recent years, which could actually be attributed to a higher economic status, which is somewhat consistent with our findings in 3.2, where a higher GDP could actually result in more suicides. Finally, in our last two plots, we see that the suicide rates are lower when the HDI is higher in the top 10 countries, the HDI is lower when the population is larger, and the fact that these countries with larger populations, despite having a good GDP, their suicide rates are higher.

Taking all of these findings into account, as well as doing some additional research, we can see that all of our findings make logical sense. As the GDP, and more importantly, HDI increases, the suicide rates are lower, which makes sense in Japan’s case. According to our additional findings, we see that GDP and HDI generally increase over time, at least for the top 10 countries in our dataset, indicating that as socioeconomic factors become more favorable, the suicide rate decreases. We also see that males have a higher suicide count than females in general. This conclusion is further reinforced when taking the generations into account, and factoring in the major world events that happened during the prime of each generation, the sheer numbers and the suicide rates certainly make sense.

**APPENDIX**

**6. Code for Data Cleaning**

#CODE FOR DATA CLEANING/INITIALIZING

#load libraries

library(ggplot2)

library(maps)

library(dplyr)

#load relevant datasets into Rstudio in Rstudio

mas <- read.csv("master.csv")

world\_map <- map\_data("world")

#rename column names

colnames(mas) <- c("country", "year", "sex", "age", "suicides\_no", "population",

"suicides\_100k\_pop", "country\_year", "hdi", "gdp\_year",

"gdp\_capita", "generation")

#Add columns

region <- mas$country

mas <- cbind(mas, region)

mas$region <- as.character(mas$region)

#Data Cleaning for whole data set

#converting all columns with numbers to numeric

mas$year <- as.numeric(mas$year)

mas$suicides\_no <- as.numeric(mas$suicides\_no)

mas$population <- as.numeric(mas$population)

mas$suicides\_100k\_pop <- as.numeric(mas$suicides\_100k\_pop)

mas$hdi <- as.numeric(mas$hdi)

mas$gdp\_year <- as.numeric(mas$gdp\_year)

mas$gdp\_capita <- as.numeric(mas$gdp\_capita)

#Figure 3.1

#Replace region names that differ from the corresponding region names

#in the world\_map dataframe

mas$region[which(mas$region == "United States")] <- "USA"

mas$region[which(mas$region == "Republic of Korea")] <- "South Korea"

mas$region[which(mas$region == "Russian Federation")] <- "Russia"

mas$region[which(mas$region == "Saint Vincent and Grenadines")] <- "Grenadines"

mas$region[which(mas$region == "Saint Kitts and Nevis")] <- "Saint Kitts"

world\_map <- left\_join(world\_map, mas, by = "region")

#filter data by selected year

#Year 1995 data

world\_1995 <- filter(world\_map, world\_map$year == 1995)

#Year 2000 data

world\_2000 <- filter(world\_map, world\_map$year == 2000)

#Year 2005 data

world\_2005 <- filter(world\_map, world\_map$year == 2005)

#Year 2010 data

world\_2010 <- filter(world\_map, world\_map$year == 2010)

#Figure 3.3

#make subset of each generation

x <- mas[mas[ , "generation"] == "Generation X", ]

sil <- mas[mas[ , "generation"] == "Silent", ]

gi <- mas[mas[ , "generation"] == "G.I. Generation", ]

boom <- mas[mas[ , "generation"] == "Boomers", ]

mil <- mas[mas[ , "generation"] == "Millenials", ]

z <- mas[mas[ , "generation"] == "Generation Z", ]

#find sum of suicide numbers per generation

nx <- sum(x$suicides\_no)

nsil <- sum(sil$suicides\_no)

ngi <- sum(gi$suicides\_no)

nboom <- sum(boom$suicides\_no)

nmil <- sum(mil$suicides\_no)

nz <- sum(z$suicides\_no)

#vector of total rows of each generation

numofgen <- c(nx, nsil, ngi, nboom, nmil, nz)

#names of each generation

names(numofgen) <- c("Gen X", "Gen Silent", "G.I.", "Boomers", "Millennials", "Gen Z")

#Figure 3.4

#Data cleaning for Figure 3.4.a

#selecting united stated suicide rates from 1985 to 1990

decade1 <- filter(mas, year == as.factor(c(1985:1990)), country == "United States")

#Data cleaning for Figure 3.4.b

#selecting united stated suicide rates by male and female from 1991 to 2000

decade2 <- filter(mas, year == c(1991:2000), country == "United States")

#Data cleaning for Figure 3.4.c

#selecting united stated suicide rates by male and female from 2001 to 2010

decade3 <- filter(mas, year == c(2001:2010), country == "United States")

#Data cleaning for Figure 3.4.d

#selecting united stated suicide rates by male and female from 2010 to 2016

decade4 <- filter(mas, year == c(2011:2016), country == "United States")

#Figure 3.5

#Data cleaning for 3.5

#subsets of top 10 suicide\_no countries

rus <- mas[mas[ , "country"] == "Russian Federation", ]

us <- mas[mas[ , "country"] == "United States", ]

jap <- mas[mas[ , "country"] == "Japan", ]

ukr <- mas[mas[ , "country"] == "Ukraine", ]

kor <- mas[mas[ , "country"] == "Republic of Korea", ]

ger <- mas[mas[ , "country"] == "Germany", ]

fra <- mas[mas[ , "country"] == "France", ]

bra <- mas[mas[ , "country"] == "Brazil", ]

po <- mas[mas[ , "country"] == "Poland", ]

uk <- mas[mas[ , "country"] == "United Kingdom", ]

#find the sum of every row in suicides\_no column

totrus <- sum(rus$suicides\_no)

totus <- sum(us$suicides\_no)

totjap <- sum(jap$suicides\_no)

totukr <- sum(ukr$suicides\_no)

totkor <- sum(kor$suicides\_no)

totger <- sum(ger$suicides\_no)

totfra <- sum(fra$suicides\_no)

totbra <- sum(bra$suicides\_no)

totpo <- sum(po$suicides\_no)

totuk <- sum(uk$suicides\_no)

#vector of each sum

numofkills <- c(totrus, totus, totjap, totukr, totkor,

totger, totfra, totbra, totpo, totuk)

#name of countries

names(numofkills) <- c("Russia", "US", "Japan", "Ukraine", "Korea", "Germany",

"France", "Brazil", "Poland", "UK")

#Figure 3.6

#Data cleaning for Figure 3.6

#selecting top ten countries with the highest HDI

top\_ten <- filter(mas, country == c("Norway", "Australia", "Switzerland", "Denmark",

"Netherlands", "Germany", "Ireland", "United States", "Canada",

"Singapore"))

**7. Code for Data Analysis**

#Figure 3.1

#Year 1995 map

ggplot(world\_1995) +

geom\_polygon(mapping = aes(x = long, y = lat, group = group, fill = suicides\_100k\_pop)) +

coord\_quickmap() +

scale\_fill\_gradient(name = "Suicides/100K Population",

low = "white", high = "#e87500") +

theme(axis.text.x = element\_blank(), axis.ticks.x = element\_blank()) +

labs(x = "", y = "", title = "Year 1995 Suicide Rate by Population") +

theme\_dark()

#Year 2000 map

ggplot(world\_2000) +

geom\_polygon(mapping = aes(x = long, y = lat, group = group, fill = suicides\_100k\_pop)) +

coord\_quickmap() +

scale\_fill\_gradient(name = "Suicides/100K Population",

low = "white", high = "#e87500") +

theme(axis.text.x = element\_blank(), axis.ticks.x = element\_blank()) +

labs(x = "", y = "", title = "Year 2000 Suicide Rate by Population") +

theme\_dark()

#Year 2005 map

ggplot(world\_2005) +

geom\_polygon(mapping = aes(x = long, y = lat, group = group, fill = suicides\_100k\_pop)) +

coord\_quickmap() +

scale\_fill\_gradient(name = "Suicides/100K Population",

low = "white", high = "#e87500") +

theme(axis.text.x = element\_blank(), axis.ticks.x = element\_blank()) +

labs(x = "", y = "", title = "Year 2005 Suicide Rate by Population") +

theme\_dark()

#Year 2010 map

ggplot(world\_2010) +

geom\_polygon(mapping = aes(x = long, y = lat, group = group, fill = suicides\_100k\_pop)) +

coord\_quickmap() +

scale\_fill\_gradient(name = "Suicides/100K Population",

low = "white", high = "#e87500") +

theme(axis.text.x = element\_blank(), axis.ticks.x = element\_blank()) +

labs(x = "", y = "", title = "Year 2010 Suicide Rate by Population") +

theme\_dark()

#Figure 3.2

#GDP per Capita, faceted

ggplot(data = mas) +

geom\_point( aes(x = gdp\_capita, y = suicides\_no, color = generation)) +

geom\_smooth( aes(x = gdp\_capita, y = suicides\_no)) +

facet\_wrap(~generation) +

labs(x = "GDP per Capita", y = "Suicide Count",

title = "GDP per Capita vs Suicide Count") +

scale\_color\_brewer(palette = "Set2")

#GDP per capita, not faceted

ggplot(data = mas) +

geom\_point( aes(x = gdp\_capita, y = suicides\_no, color = generation)) +

geom\_smooth( aes(x = gdp\_capita, y = suicides\_no)) +

labs(x = "GDP per Capita", y = "Suicide Count",

title = "GDP per Capita vs Suicide Count") +

scale\_color\_brewer(palette = "Set2")

#Figure 3.3

#Data Plotting for Figure 3.3

#bar plot of data

barplot(sort(numofgen, decreasing = TRUE), main = "Suicides per Generation", xlab = "Generation", ylab = "Number of Suicides", ylim = c(0,2500000), col = c("purple3", "navy", "blue", "green4", "yellow3", "orange2"))

#Figure 3.4

#Data Plotting for Figure 3.4.a

#plotting the data by year on x-axis and suicide count on the y-axis and using gender as the trend

ggplot(data = decade1, mapping = aes(x = year, y = suicides\_no)) +

geom\_smooth(mapping = aes(color = sex)) + scale\_y\_continuous(limits = c(0,10000)) +

labs(x = "Year", y = "Suicide Count", title = "Suicide trends from 1986 to 1990 in US") +

scale\_color\_brewer(palette = "Set1")

#Data Plotting for Figure 3.4.b

#ploting the data by year on x-axis and suicide count on the y axis and using gender as the trend

ggplot(data = decade2, mapping = aes(x = year, y = suicides\_no)) +

geom\_smooth(mapping = aes(color = sex)) +

scale\_y\_continuous(limits = c(0,10000)) +

labs(x = "Year", y = "Suicide Count", title = "Suicide trends from 1991 to 2000 in US") +

scale\_color\_brewer(palette = "Set1")

#Data Plotting for Figure 3.4.c

#ploting the data by year on the x axis and suicide count on the y-axis and using gender as the trend

ggplot(data = decade3, mapping = aes(x = year, y = suicides\_no)) +

geom\_smooth(mapping = aes(color = sex)) +

scale\_y\_continuous(limits = c(0,10000)) +

labs(x = "Year", y = "Suicide Count", title = "Suicide trends from 2001 to 2010 in US") +

scale\_color\_brewer(palette = "Set1")

#Data Plotting for Figure 3.4.d

#ploting the data by year on x-axis and suicide count on the y-axis and using gender as the trend

ggplot(data = decade4, mapping = aes(x = year, y = suicides\_no)) +

geom\_smooth(mapping = aes(color = sex)) +

scale\_y\_continuous(limits = c(0,10000)) + labs(x = "Year", y = "Suicide Count",

title = "Suicide trends from 2010 to 2016 in US") +

scale\_color\_brewer(palette = "Set1")

#Figure 3.5

#Data Plotting for 3.5

#barplot of data

barplot(sort(numofkills, decreasing = TRUE), main = "Top 10 Highest Suicides per Country",

xlab = "Country", ylab = "Number of Suicides",

col = c("red3", "blue2", "white", "blue3", "yellow", "black", "blue4", "green", "red", "red4"))

#Figure 3.6

#Data Plotting for Figure 3.6.a

# creating box plots for each county associated with their HDI

#since some country have the same flag color, need a different color scheme

# using various color rather than their flag color

ggplot(data = top\_ten, mapping = aes(country,y = hdi , fill = country)) +

geom\_boxplot(notch = FALSE, na.rm = TRUE, notchwidth = 0.2) +

labs(x = "Country", y = "HDI", title = "Top 10 Countries with highest HDI")

#Data Plotting for Figure 3.6.b

#using the top ten countries with the highest HDI

#creating box plots for each country associate with their suicides\_no

ggplot(data = top\_ten, mapping = aes(country,y = suicides\_no , fill = country)) +

geom\_boxplot(notch = FALSE, na.rm = TRUE, notchwidth = 0.2) +

labs(x = "Country", y = "Suicides\_per\_100k", title = "Highest HDI countries vs Suicides count")+

scale\_y\_continuous(limits = c(0,700))

#Data Plotting for Figure 3.6.c

#using the top countries with the highest HDI

#creating box plots for each country associated with their gdp\_capita

ggplot(data = top\_ten, mapping = aes(country,y = gdp\_capita , fill = country)) +

geom\_boxplot(notch = FALSE, na.rm = TRUE, notchwidth = 0.2) +

labs(x = "Country", y = "GDP\_capita", title = "Highest HDI countries vs GDP per capita")

**8. Code for Additional Data**

#4.1

#Selecting Top five countries with highest HDI

top\_five <- filter(mas, country == c("Norway", "Australia", "Switzerland", "Denmark", "Netherlands"))

#ploting x as year and y as HDI

#adding trend line of countres to analyze

ggplot (data = top\_five, mapping = aes(x = year, y = hdi)) +

geom\_smooth(mapping = aes(color = country))+

labs(x = "Year", y = "HDI", title = "Countries with the Highest HDI vs year") +

scale\_color\_brewer(palette = "Set1")

#4.2

#Gdp per year, faceted

ggplot(data = mas) +

geom\_point( aes(x = gdp\_year, y = suicides\_no, color = generation)) +

geom\_smooth( aes(x = gdp\_year, y = suicides\_no)) +

facet\_wrap(~generation) +

labs(x = "GDP per Year", y = "Suicide Count",

title = "GDP per Year vs Suicide Count") +

scale\_color\_brewer(palette = "Set2")

#GDP per year, not faceted

ggplot(data = mas) +

geom\_point( aes(x = gdp\_year, y = suicides\_no, color = generation)) +

geom\_smooth( aes(x = gdp\_capita, y = suicides\_no, color = generation)) +

labs(x = "GDP per Capita", y = "Suicide Count",

title = "GDP per Capita vs Suicide Count") +

scale\_color\_brewer(palette = "Set2")

**References**

**[1]** [**https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016**](https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016)

**[2]<https://www.businessinsider.com/cdc-teenage-gen-z-american-suicide-epidemic>**

**[3]<https://en.wikipedia.org/wiki/Human_Development_Index>**

**[4]**[**https://www.thebalance.com/real-gdp-per-capita-how-to-calculate-data-since-1946-3306028**](https://www.thebalance.com/real-gdp-per-capita-how-to-calculate-data-since-1946-3306028)