**TERM PROJECT REPORT**

**Global Suicide Trends and Analysis**

**SCS\_3250\_028**

**By Team 5:**

Vaishali Siddeshwar

Chandni Sehgal

Gaurav Rattan Singla

Jie Ren

Javier Rodriguez

Duane Robinson

Liangliang Shi

Wing Ho

July 23rd, 2019

*A comprehensive analysis to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum.*

Contents

[**1.** **Introduction** 2](#_Toc14796657)

[**2.** **Objective** 2](#_Toc14796658)

[**3.** **Data Description** 2](#_Toc14796659)

[**4.** **Data Preparation** 5](#_Toc14796660)

[**5.** **Global Analysis** 7](#_Toc14796661)

[5.1 Global Trend 7](#_Toc14796662)

[5.2 By Continent 8](#_Toc14796663)

[5.3 By Gender 9](#_Toc14796664)

[5.3 By Age 11](#_Toc14796665)

[5.4 By Country 12](#_Toc14796666)

[5.4.1 Overall 13](#_Toc14796667)

[5.4.2 Linear Trends 15](#_Toc14796668)

[5.5 Gender differences, by Continent 18](#_Toc14796669)

[5.6 Gender differences, by Country 19](#_Toc14796670)

[5.7 Age differences, by Continent 21](#_Toc14796671)

[5.8 As a country gets richer, does its suicide rate decrease? 22](#_Toc14796672)

[5.9 Do richer countries have a higher rate of suicide? 22](#_Toc14796673)

[5.10 The 5% highest risk instances in history 24](#_Toc14796674)

[**6.** **Trends in developed countries** 26](#_Toc14796675)

[6.1 Overall Trend 26](#_Toc14796676)

[6.2 By Gender 27](#_Toc14796677)

[6.2.1 Male & Female Rates (over time) 27](#_Toc14796678)

[6.2.2 Young to Middle-Aged Men 28](#_Toc14796679)

[6.2.3 Men - Ages 15-54 Combined 29](#_Toc14796680)

[6.4.2 Men - Ages 15-24, 25-34 & 35-54 29](#_Toc14796681)

[6.3 Recent Trends in developed countries 31](#_Toc14796682)

[6.3.1 Proportion of suicides that are Men 31](#_Toc14796683)

[6.3.2 Age Rates 32](#_Toc14796684)

[6.3.3 Male & Female Rates (for different age categories) 32](#_Toc14796685)

[**7.** **HDI and Suicide Rate** 33](#_Toc14796686)

[**8.** **Generation and Suicide Rate** 34](#_Toc14796687)

[**9.** **2007-08 Recession** 36](#_Toc14796688)

[**10** **External Factor** 37](#_Toc14796689)

[**11** **Conclusion** 38](#_Toc14796690)

[**12** **References** 38](#_Toc14796691)

# **Introduction**

Every individual is responsible for their actions and outcomes. However, the society and the socio-economic environment can have a huge impact on an individual’s life.

Suicide is global and according to World Health Organization, “Close to 800 000 people die due to suicide every year, which is one person every 40 seconds” (WHO, 2019).

Particularly, most of the victims are aged between the ages of 15 – 49.

In the report below, Suicide Rate data has been explored to understand what variables act as dominant factors and what is the correlation between these variables, that forces people to commit suicide.

# **Objective**

The dataset is explored to understand what significance each variable could potentially have on an individual’s life. It is further explored to see if any of these variables have any correlation with each other.

For ex: GDP which represents the economy of a country and its residents; does GDP have any correlation with number of suicides? In other words, if the GDP of a country goes up, the number of suicides should go down and vice-versa. Did the suicide rate during the recession 2007-08.

Furthermore, the significance of gender variable and its correlation with number of suicides is also explored.

# **Data Description**

The data was obtained from Kaggle

Kaggle consolidated the information from four other datasets that are linked by time and place. This was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum.

We can summarize the data as follows:

1. The dataset contains information on suicide rates between 1985 and 2015 for 100 countries in the world.
2. Year 2016 is included in the data, but it is assumed to be incomplete.

The data is in csv format and has 27820 x 12 (rows x columns)

Below are the descriptions of all the features:

**Categorical Features:**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Data Type** | **Explanation** |
| *Country* | String | There are 101 countries in this dataset. |
| *Year* | Integer | This ranges from 1985 to 2016 |
| *Sex* | String | Male/female gender |
| *Age* | String | Five intervals i.e. 5-14, 15-24, 25-34, 35-54, 55-74 and 75+ |
| *Generation* | String | There are six generations included in this dataset:  *G.I. Generation*: Born 1920 or before  *Silent Generation*: Born 1921 - 1945  *Boomers Generation*: Born 1946 - 1964  *Generation X*: Born 1965 - 1976  *Millennials Generation*: Born 1977 - 1995  *Generation Z*: Born 1996 – TBD |

**Numerical Features:**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Data Type** | **Explanation** |
| *Population* | Integer | Number of people in Country during the given Year |
| *Number of Suicides* | Integer | Number of suicides in Country during the given Year |
| *Suicides per 100k people* | Integer | Number of suicides divided by the population size and multiplied by 100,000 for a Country during the given Year |
| *GDP for year* | Integer | Gross Domestic Product, a measure of the market value for a country-year combination. |
| *GDP per capita* | Integer | Obtained by dividing the GDP by the total population of the country for that year. |
| *HDI* | Decimal | Human Development Index, an index that measures life expectancy, income and education for a country for a given Year. |

So, every sample combination of Country, Year, Sex and Age, which is a subsample of the population. An example demographic is - Canadian males of age between 15 and 25 in 1996.

The limitations of data are:

1. No **reasons for suicide** mentioned
2. data available only for 93 countries
3. No data for populous countries like India, China.
4. HDI is missing for 2/3rd data

We have integrated the suicide dataset with temperature dataset.

The Temperature dataset is also obtained from Kaggle.

Below table shows the variables of Temperature dataset:

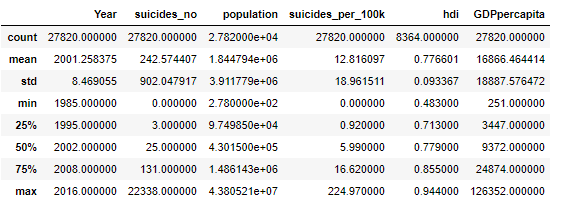
|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Data Type** | **Explanation** |
| Date | Date | in YYYY-MM-DD format ranging from 1743-11-01 to 2013-09-01 represented for each country |
| AverageTemperature | Float | Temperature of the given Country during a specific Year |
| AverageTemperatureUncertainty | Float | Temperature of the given Country during a specific Year |
| Country | String | Name of the Country |

We have dropped the following features from Temperature dataset: AverageTemperatureUncertainty

The common variables between suicide rate dataset and temperature dataset were country and year. The **year** in temperature dataset had to be cleaned up to match the year range and format.

# **Data Preparation**

In the first step of Data Preparation we loaded the dataset from csv file and renamed a few features for convenience. We analyzed the below statistics of the data



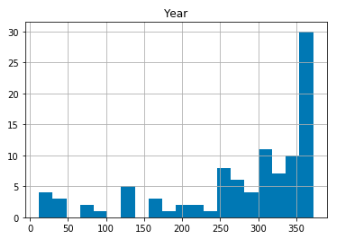
We can see that none of variables except **HDI has missing values**. The below output confirms that ~70% of HDI values are missing.



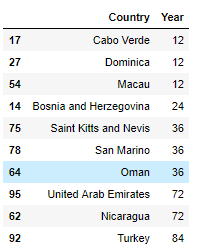
This means, there are 19456 NaNs for HDI feature out of 27820. i.e. We cannot fill so many values as this will bias our analysis. Hence, NaNs are **NOT** replaced. Only available values of HDI are used.

Next, we are dropping the columns that are not required for our analysis. The dropped feature is ‘country-year’ as it is redundant. We have also formatted the data in Age column for better plotting. We have changed the datatype of features-Age and Generation from String to Categorical.

Now, we found out if any countries have very less data. Below histogram gives us a clue that there are few countries with very less information.



We have listed below, the countries that have only 5 or less than 5 years of information.



We have removed them from our analysis. The Countries removed are Cabo Verde, Dominica, Macau, Bosnia and Herzegovina, Saint Kitts and Nevis, San Marino, Oman.

At this point, the total number of rows and columns left for analysis was 27,492 rows and 11 columns. In short, after data cleaning and preparation, only 1.18% of data was lost, which in turn indicates that dataset obtained had good quality of information for rest of the analysis.

We added a new Column 'Continent' for analyzing trends for each Continent.

We added a 'Decade' column to analyze trends for 3 decades from 1985 - 2015

HDI categorized as - Very High, High, Medium and Low representing Human Development Index categories

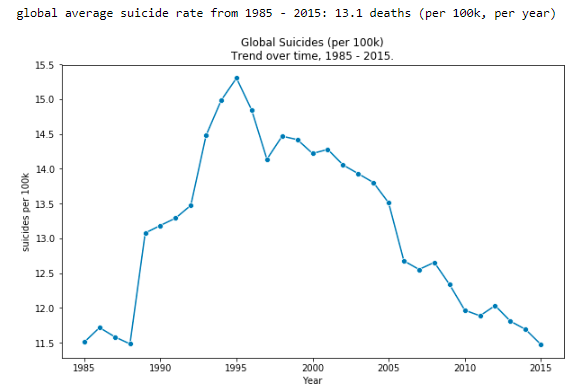
# **Global Analysis**

This is the highest level of analysis that includes all the Countries, all age groups, both the gender and during the timespan 1985-2015.

## 5.1 Global Trend

To begin with, one of the biggest challenges is to determine whether the death of a person in a country is recorded with how much accuracy. In other words, it is hard to determine if the death of an individual that was recorded in the country was actually suicide or a natural death. This uncertainty about this is especially high in developing or poor countries where the system to record such information may or may not be present.

Hence, it is good to get the overall bigger picture about the trend of suicide rate across the world from the year of 1985 until 2015. In order to do this, as the first level of analysis, we have plotted the trend of suicide rate globally. Every data point in the plot denotes suicide/100k population for all the countries and all age groups and both gender in a given year.



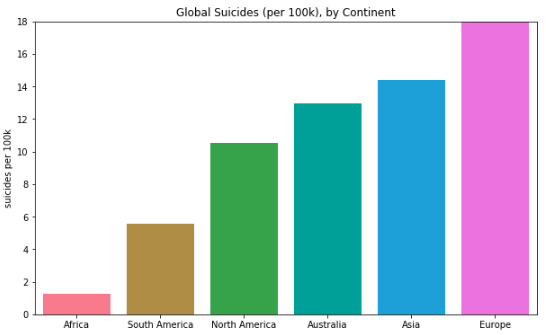
Below are insights drawn from the above plot:

* global average suicide rate from 1985 - 2015: 13.1 deaths (per 100k, per year).
* Peak suicide rate was 15.3 deaths per 100k in 1995**.**
* Rate has decreased steadily, to 11.5 per 100k in 2015 (~25% decrease).
* Rates are only now returning to their pre-90’s rates.
* Insights before 90’s seem to be relatively weak as there is not enough data available. Only 5 data points were available for the trend before 90’s.

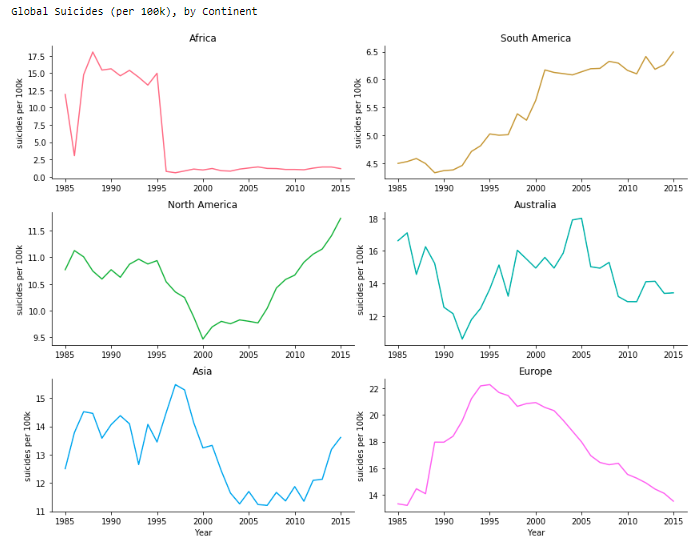
At the Global level, we got a rough idea that, the suicide rate (in general) is showing a downward trend over the years,

## By Continent

Now, we are in second level of analysis. We are interested to know the suicide rate for each continent. Firstly, let us see the counts of suicide rate per continent as a bar graph.



Let us also examine the trends in the Continents over the span of years 1985-2015.

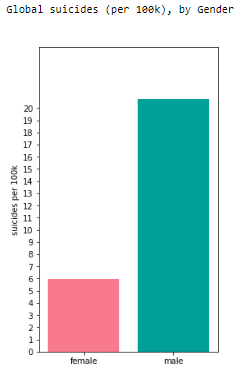


From the above two plots drawn at the Continent Level, we have the following insights:

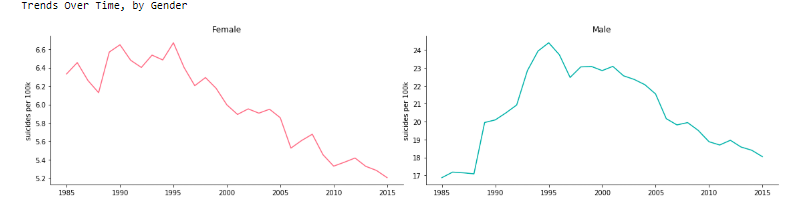
* European rate highest overall, but has steadily decreased ~40**%** since 1995.
* The trendline for Africa is due to poor data quality - just 3 countries have provided data.
* Americas' upward trends are more concerning.

## 5.3 By Gender

We now have to find out which Gender is more susceptible to suicides. To do this, first, we plot a bar graph showing suicide counts for both the Gender in all the countries over the span of 1985-2015.



Next, we also want to know the trend of suicide rates collectively for all countries over the years, separately for each Gender.

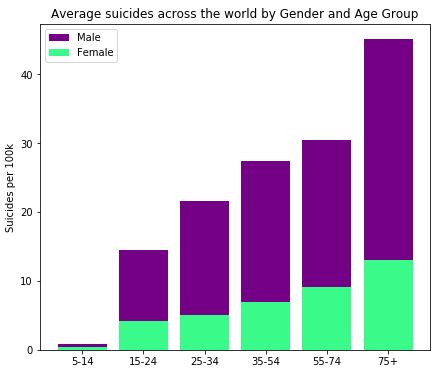
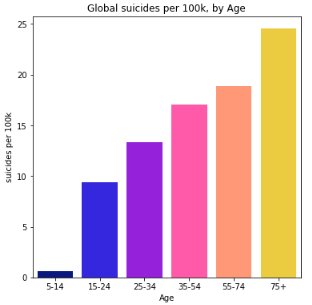


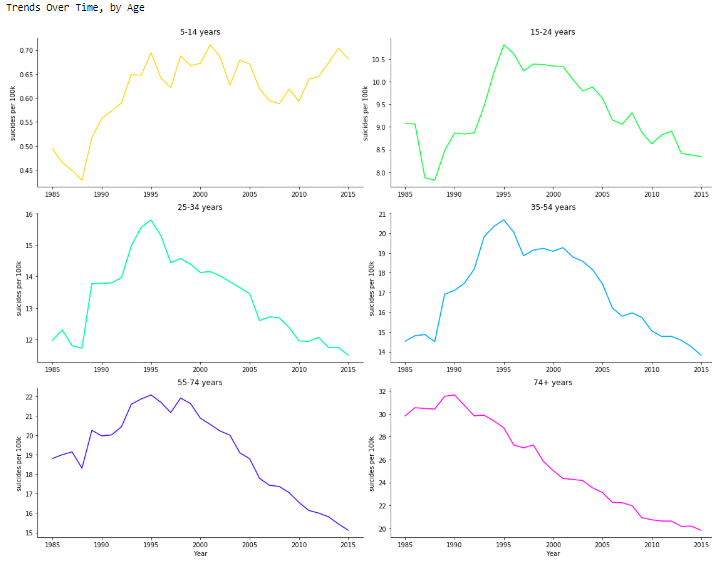
By seeing the Trends for both Gender and the Bar Graph we understand that,

* Globally, the rate of suicide for men has been ~3.5x higher.
* Both male & female suicide rates peaked in 1995, declining since.
* This ratio of 3.5: 1 (male: female) has remained relatively constant since the mid 90’s.
* However, during the 80’s this ratio was as low as 2.7: 1 (male: female).

## By Age

Now, we are interested to know globally, which age Group is more susceptible to suicide. Firstly, we plot a bar graph of counts of suicides for each Age Group from all the countries, from both genders and over the years 1985-2015. Also plot, how many of the suicides are contributed by Women.



We also plot the trends for each age group separately over the years.

The key insights from the above 3 plots are:

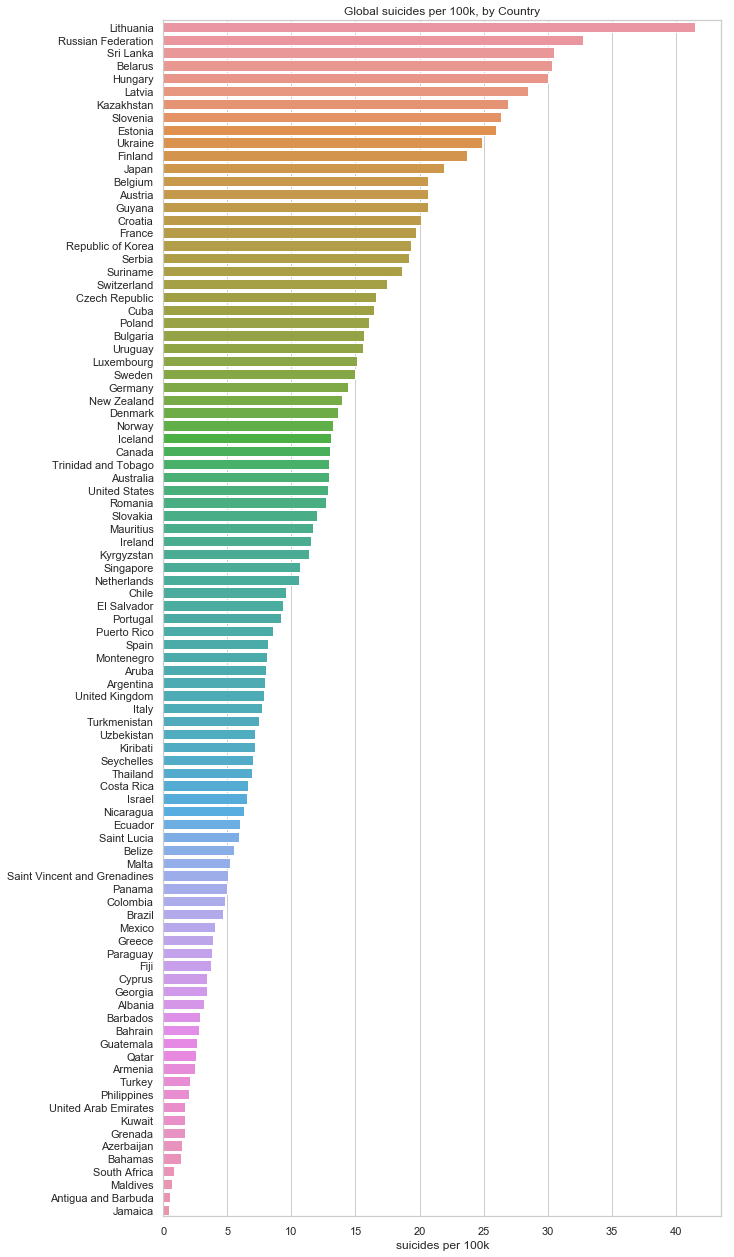
* Globally, the likelihood of suicide increases with age
* Men across all age groups have higher suicide rate.
* After 1995, suicide rate for everyone aged >= 15 has been linearly decreasing
* The suicide rate of those aged 75+ has dropped by more than 50% since 1990
* Suicide rate in the ‘5-14’ category remains roughly static and small (< 1 per 100k per year)

## By Country

Now, we are in third level of analysis. Here we analyze suicide rate for each country for all Age Groups and both genders over the years 1985-2015. We also analyze the Yearly trends of each Country.

### 5.4.1 Overall

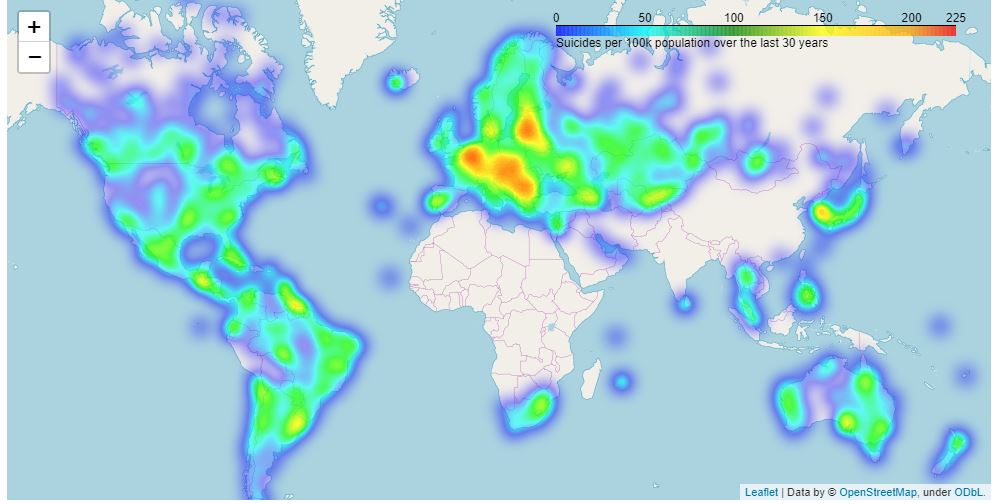
Here we first plot the count of suicide rate for each country for all Age Groups and both genders over the years 1985-2015

.

Here are the insights for the above graph:

* Lithuania’s rate has been highest by a large margin: > 41 suicides per 100k (per year)
* Large overrepresentation of European countries with high rates, few with low rates

Below is a geographical heat map of the suicide rates between the timeframe of this analysis - **note the lack of data for Africa and Asia**, and bear in mind that 10 countries have been removed due to insufficient data.



Southwestern European nations have highest suicidal rates (~199/100K), while the rest of Europe are in line with the Americas (75 to 120 / 100K).

It’s important to note that looking at figures at a global/continent level might not truly be representative of the globe/continent for these reasons.

Comparing the raw suicide rates of countries may also lead to some issues - the definition of suicide (and the reliability that a death is recorded as suicide) will likely vary between countries.

However, trends over time (within countries) are likely to be reliable. We address this next.

### Linear Trends

We are interested in how the suicide rate is changing over time within each country.

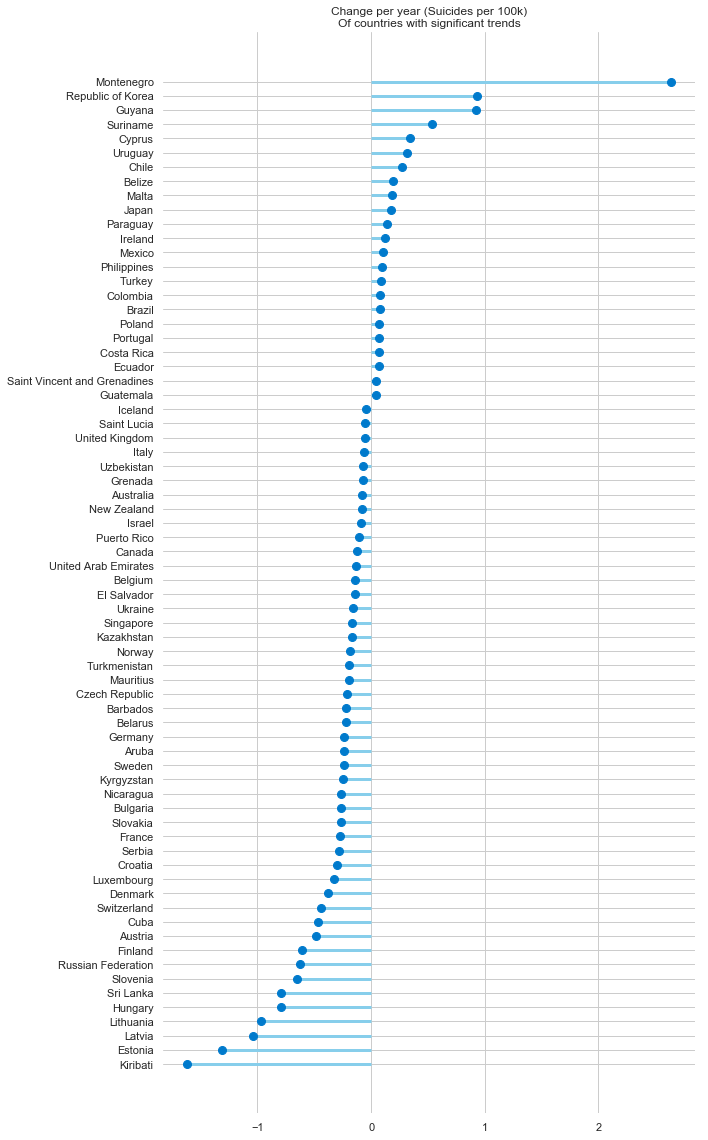
In other words: as time goes on, we look for countries where the suicide rate is linearly increasing or decreasing over time. These can then be rank ordered by their rate of change as time goes on.

The graph on the next page shows the linear increase/decrease in “rate of change of suicide rate” for each country over the years 1985-2015. We can see that as time progresses, most of the countries have decrease in their suicide rate.

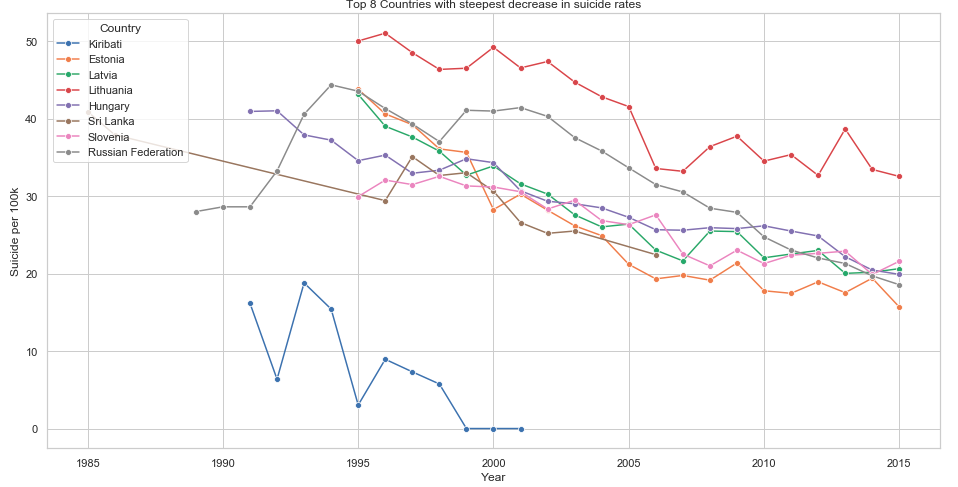
We have calculated the percentage of total countries showing decreasing trend as 67%.

Below are the insights that we can make from the below graph:

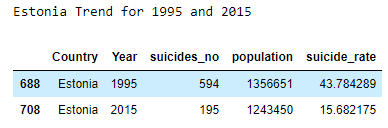
* ~1/2 of all countries suicide rates are changing linearly as time progresses
* ~2/3 of these countries are decreasing
* Overall, this is painting a positive picture



Let us analyze the trends for few countries with Steepest decrease.



We have calculated ‘Estonia’s trend between the years 1995 and 2015 as:



Following are our findings from the above plot:

* Estonia shows the most positive trend - every year, ~1.31 less people (per 100k) commit suicide - the steepest decrease globally
* Between 1995 and 2015, this drops from 43.8 to 15.7 per 100k (per year) - a 64% decrease
* The Russian Federation trend is interesting, only beginning to drop in 2002. Since then it has decreased by ~50%.

Let us analyze the trends for few countries with Steepest increase.

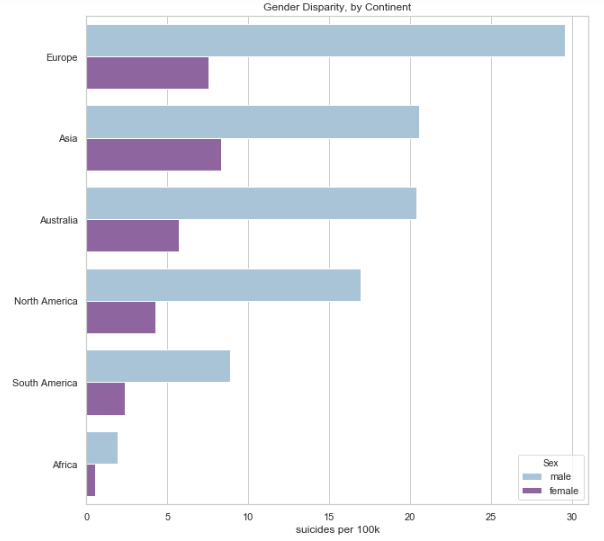


From the above graph we have the following insights:

* Korea shows the most concerning upward trend - an increase in suicide of 0.931 people (per 100k, per year) - the steepest increase globally
* Guyana is similar, at + 0.925 people (per 100k, per year)
* Between 1998 and 1999 Guyana’s rate increased enormously (5.3 to 24.8),
* The historical data for Guyana seems questionable - it’s known for very high suicide rates but the spike seems unnatural

## Gender differences, by Continent

Now, we will check the suicide rate for men from each Continent. Firstly, we plot the suicide rate for men vs. women over the years 1985-2015.

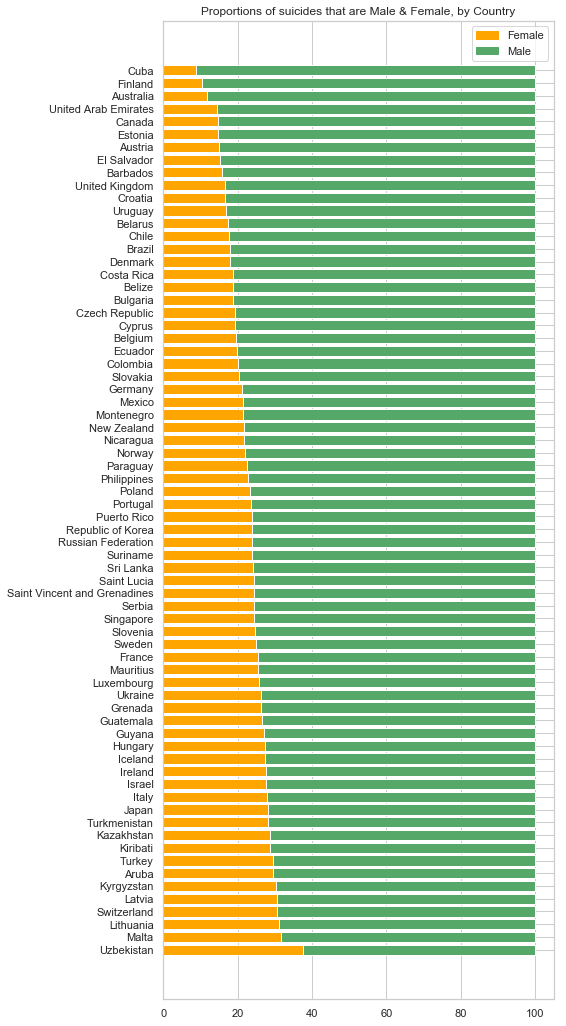


From the above plot, we get the following insights:

* European men were at the highest risk between 1985 - 2015, at ~ 30 suicides (per 100k, per year)
* Asia had the smallest overrepresentation of male suicide - the rate was ~2.5x as high for men
* Comparatively, Europe’s rate was ~3.9x as high for men

## Gender differences, by Country

We will now plot the proportion of suicides that are Male and Female for each country of all Age Groups and for the entire period of 1985-2015.

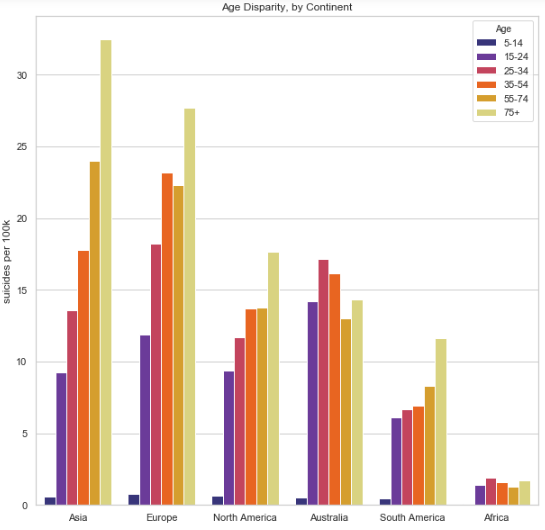


Below are our thoughts on the above graph:

* The overrepresentation of men in suicide deaths appears to be universal, and can be observed to varying extents in every country
* As per the gender paradox on suicidal behavior, whilst women are more likely to suffer from depression and suicidal thoughts, men are more likely to die from suicide

## Age differences, by Continent

We are interested to know how the suicide rate varies amongst Age groups for different Continents for both Genders and for all countries and for the entire period of 1985-2015.



The above plot confirms that - For the Americas, Asia & Europe (which make up most of the dataset), suicide rate increases with age.

## As a country gets richer, does its suicide rate decrease?

It depends on the country - for almost every country, there is a high correlation between year and GDP per capita, i.e. as time goes on, GDP per capita linearly increases.

We calculated the Pearson correlations between ‘year’ and ‘GDP per capita’ within each country, then summarized the results:

The mean correlation was 0.878, indicating a very strong positive linear relationship.

This basically means that looking within a country and asking “does an increase in wealth (per person) have an effect suicide rate” is pretty similar to asking “does a countries suicide rate increase as time progresses”.

This was answered earlier in (5.4.2) - it depends on the country. Some countries are increasing with time, most are decreasing.

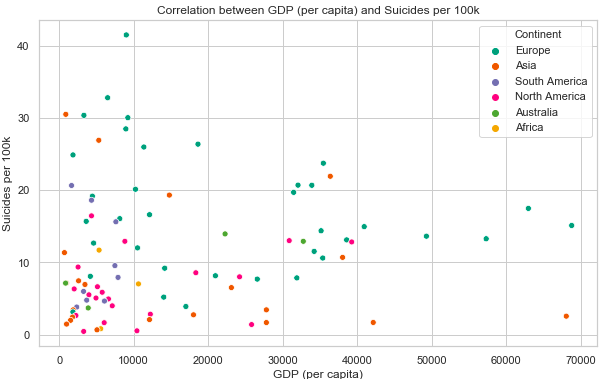
So, instead, we ask a slightly different question in 5.9.

## Do richer countries have a higher rate of suicide?

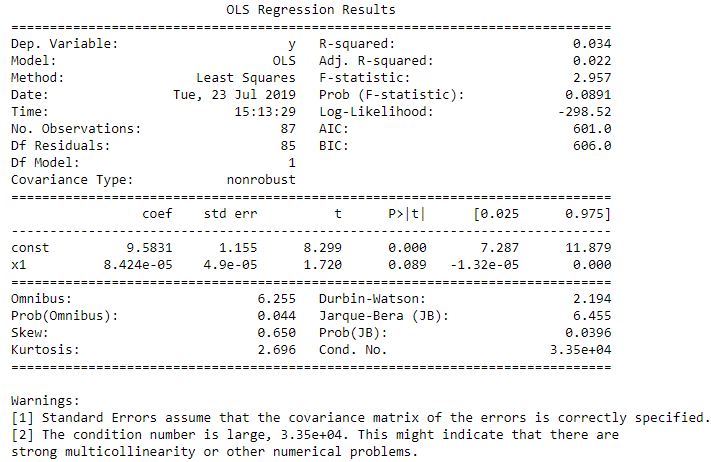
Instead of looking at trends within countries, here we take every country and calculate their mean GDP (per capita) across all the years in which data is available.

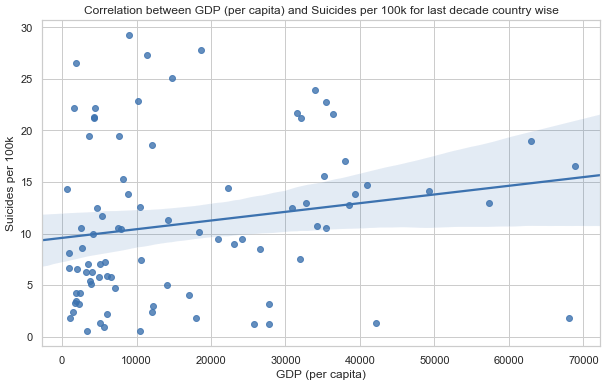
We then measure how this relates to the countries suicide rate across all those years.

The end result is one data point per country, intended to give a general idea of the wealth of a country and its suicide rate.



There are quite a few high leverage & residual countries that could have a significant impact on the fit of regression line (e.g. Lithuania, top left). So, we excluded those outliers and plotted the graph. We also now assessed the statistics of the model.





The p-value of the model is **0.034** < 0.05. This means we can reject the hypothesis that a countries GDP (per capita) has no association with its rate of suicide (per 100k).  
The r-squared is **0.0544**, so GDP (per capita) explains very little of the variance in suicide rate overall.

**What does all this mean?** 

There is a weak but significant positive linear relationship - **richer countries are associated with higher rates of suicide**, but this is a weak relationship which can be seen from the graph above.

This line of best fit is represented by the equation below, where:

* Suicides = Suicides per 100k
* GDP = GDP per capita (in thousands, USD)

Suicides=8.7718+0.1115∗GDP

This means that, at a country level and over the time frame of this analysis (1985 - 2015), **an increase of GDP (per capita) by $8,967 was associated with 1 additional suicide, per 100k people, per year.**

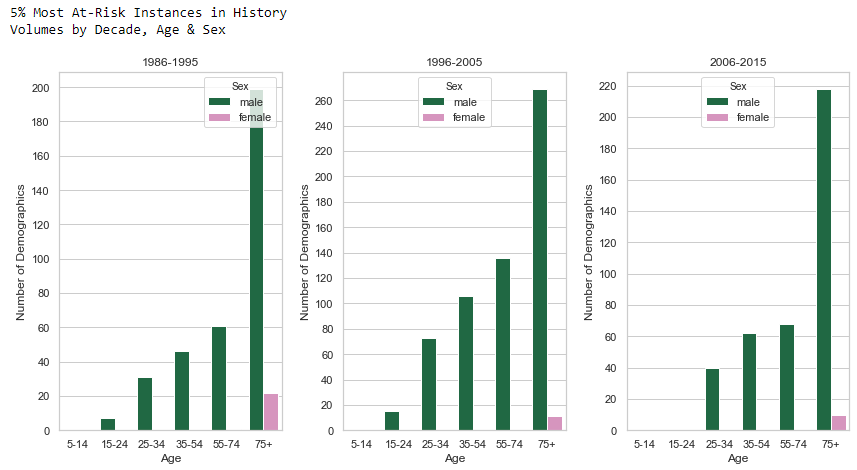
## The 5% highest risk instances in history

We will filter out data from 1985 only and look at what happens in the 3 decades following.

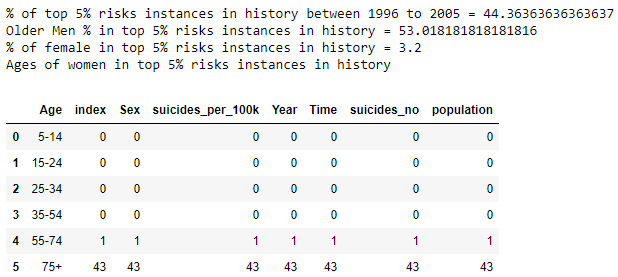
Here we are interested in the 5% highest risk (suicides/100k) demographics between 1986 and 2015.

We define a demographic as a year in a particular country, for some combination of sex & age. e.g. ‘United States, 2010, Female, 15 - 24’.

In order for a demographic to be in the top 5% for historic suicide rates, it would require a suicide rate exceeding 49.86 (per 100k) in that year.



We have calculated a few interesting facts



From the above plot and calculations, can infer that,

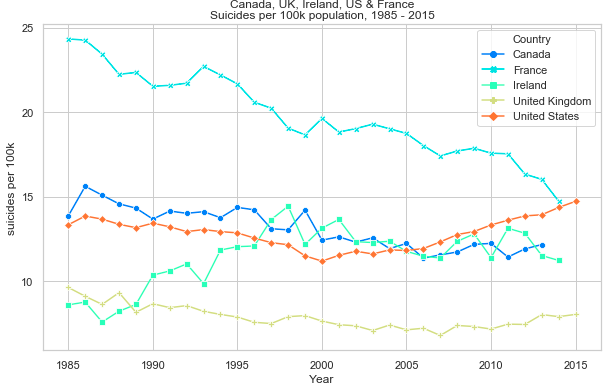
* 44.5% of these ‘high risk’ instances occurred between 1996 and 2005.
* ~50% were in the 75+ age category.
* 96.9% were a male demographic.
* Of the 3.2% (42 instances) that were for women, 41/42 of these were in the 75+ demographic.

# **Trends in developed countries**

We think it would be useful to compare a few countries that people might think of as similar to the Canada (culturally, legally and economically). Hence, we have chosen UK, Ireland, US, UK, France along with Canada.

## 6.1 Overall Trend

To understand the Overall trend in developed countries, we have plotted the trend of suicide rates for all Age Groups and both Gender in the above mentioned 5 countries during the period -1985 to 2015.



Following are the points that can be determined using the above plot:

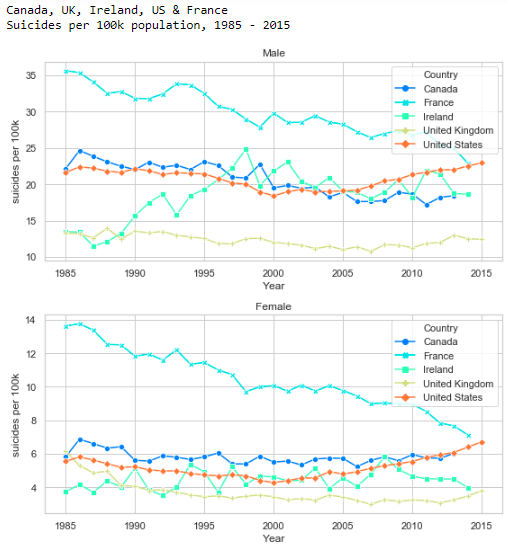
* The UK suicide rate has been consistently lowest since 1990, and has remained fairly static since ~1995.
* France has historically had the highest rate, but is now roughly equal with America.
* The US has the most concerning upward trend, linearly increasing by ~1/3 since 2000.
* Canada is showing a downward trend. Canada's suicide rate is consistently lowering been since 1995 and presently, is at its lowest rate since 1985.

## 6.2 By Gender

Now, we analyze the trends of suicide rate for male and female of all age groups.

### 6.2.1 Male & Female Rates (over time)

Here, we will be analyzing the trends of suicide rate for Male and Female separately for the developed countries for all age groups over the duration of 1985-2015.



Below are the insights found after going through the above 2 plots:

* For the UK, there’s no obvious increase in the suicide rate for men and that can also be observed to an equal extent in women.
* Rate for France has decreased to roughly rate same the US in 2015 for both Male and Female.
* The different trend lines for men & women in Ireland is unusual - in 1990, the male rate increases, but the same can’t be observed for females.
* For Canada, the suicide rate is decreasing roughly at the same rate.

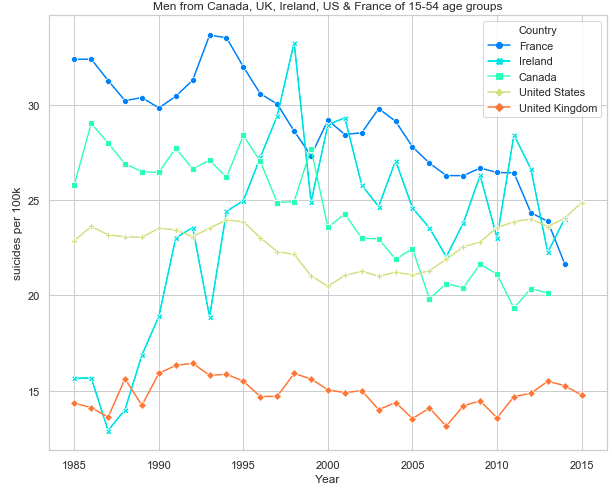
### 6.2.2 Young to Middle-Aged Men

Here we restrict our analysis to:

* Men
* Ages “15-24”, “25-34” & “35-54”

### 6.2.3 Men - Ages 15-54 Combined

First, we will plot the trends collectively for male from these 3 age groups from the chosen developed countries over the duration of 1985-2015.

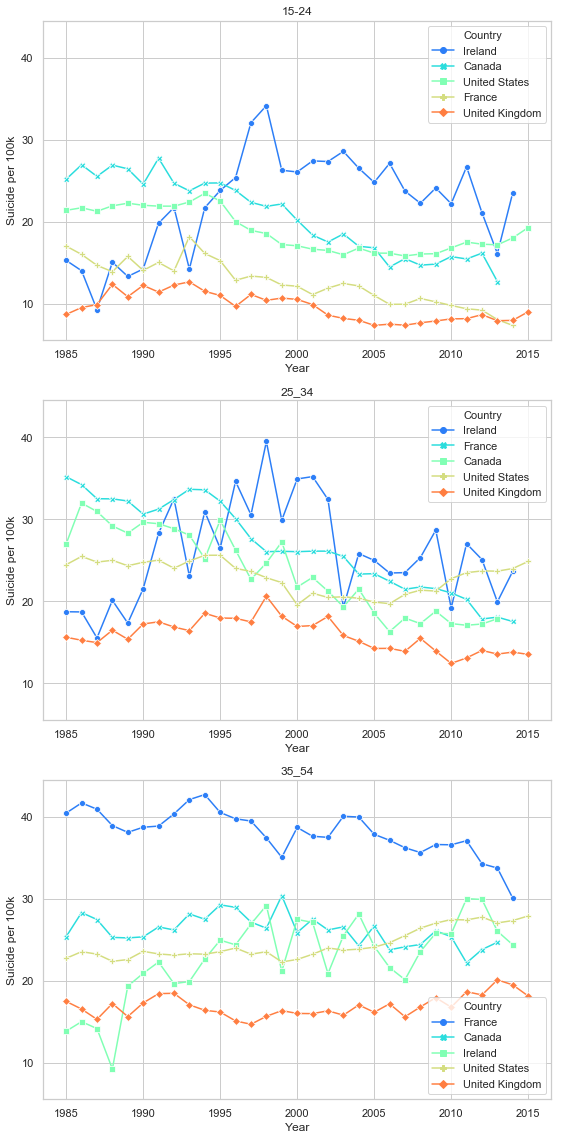


Here are the insights that we could derive:

* Ireland’s trend over the 1990’s was very concerning It went from ~14 (per 100k, per year) to 33 between 1988 and 1998.
* Again, the US shows the most obvious and concerning upward trend
* Comparatively, for young to middle-aged men, the UK seems fairly flat across time
* Canada shows a remarkable downward trend. The rate went from ~29 to ~19 from year 1995 to 2006 - a decrease of ~35%.

### Men - Ages 15-24, 25-34 & 35-54

We will plot the trends separately for male of these 3 age groups from the chosen developed countries over the duration of 1985-2015.



Below is the list of important insights:

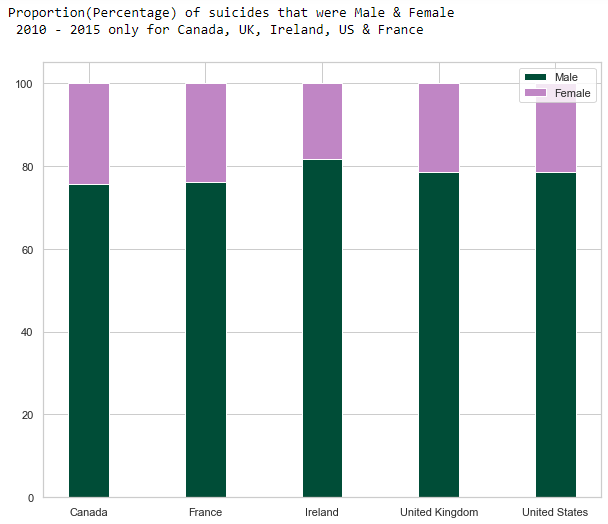
* Canada rates for men in the ‘15-24’ and ‘25-34’ categories appear to be decreasing, which is a positive thing.
* Canada rates for men in the ‘35-54’ categories appear flat & slightly decreasing.

## 6.3 Recent Trends in developed countries

For the purposes of these visualizations, we are interested in data from recent years (France, for example, has changed a lot), so we will **restrict the timeframe to 2010 onwards** for from Canada, UK, Ireland, America and France.

### 6.3.1 Proportion of suicides that are Men

Firstly, we are interested to know the ratio of suicide rate contributed by men from 2010 onwards from all the countries.

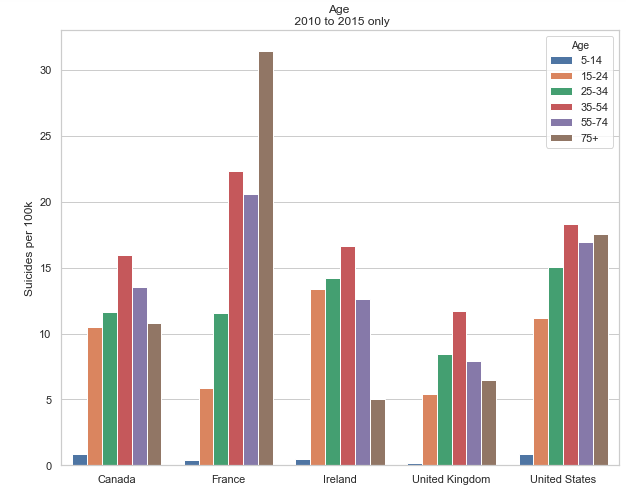


Below are the findings from the above plot:

* Similar pattern as seen throughout the analysis - men make up ~ 75% of deaths by suicide
* The highest proportion is in Ireland - ~82% male
* The lowest proportion is for Canada - ~75% male

### 6.3.2 Age Rates

We want to figure out the ratio of suicide rate contributed by each Age Group from 2010 onwards from all the countries.

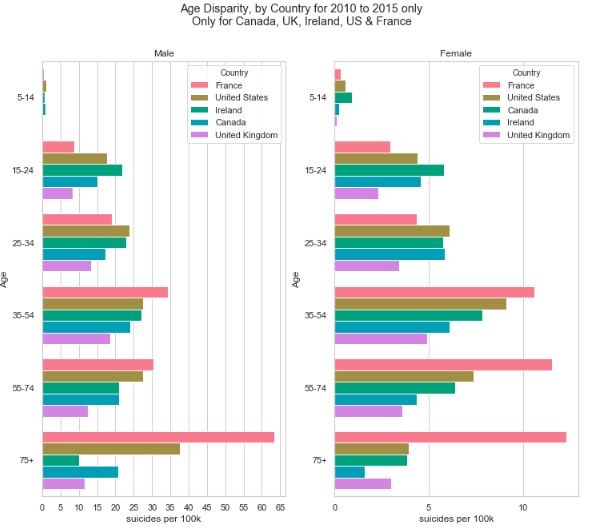


The insights from the above plot are:

* There’s a huge difference in the ‘trend’ of suicide rates as age varies within each country.
* Suicide rate increases with age for France, and the US (to a lesser extent).
* Those aged 35-54 at the highest risk in Canada, Ireland and the UK which follow closer to a gaussian distribution.

### 6.3.3 Male & Female Rates (for different age categories)

We want to figure out the ratio of suicide rate contributed by each Gender belonging to all Age Groups from 2010 onwards from all the countries.



Here are our findings derived from the above plot:

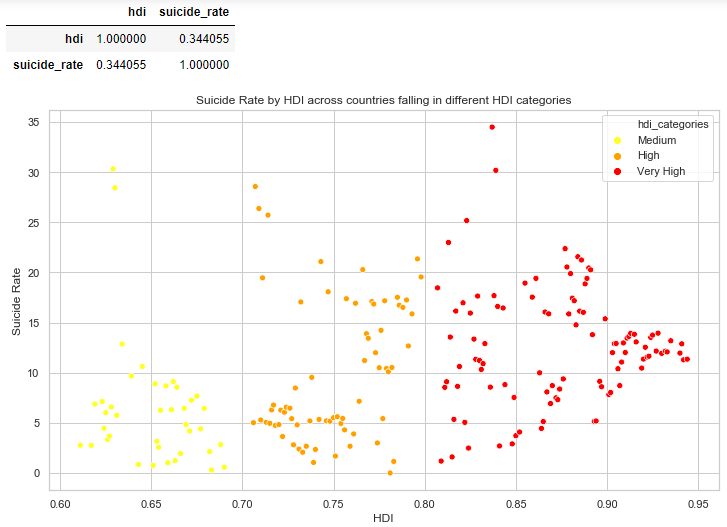
* In the US, suicide rate for men and women continues to increase with age, but the female rate decreases in old age.
* This weird disparity is only present in the US.
* Canada has the lowest suicide rate in every Gender-Age group.

# **HDI and Suicide Rate**

The Human Development Index (HDI) is a measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living.

A country scores a higher HDI when the above indicated parameters of human development is higher.

As HDI for year column contains missing values to around 70%, the data is filtered for recent years i.e., 2006 to 2016 and those countries with no HDI were removed. For the countries whose HDI is known, we have calculated the Correlation Coefficient between HDI and suicide rate and also plotted it.

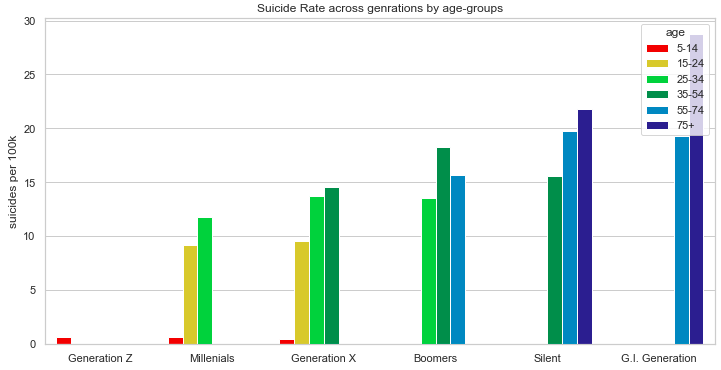


Our observation is:

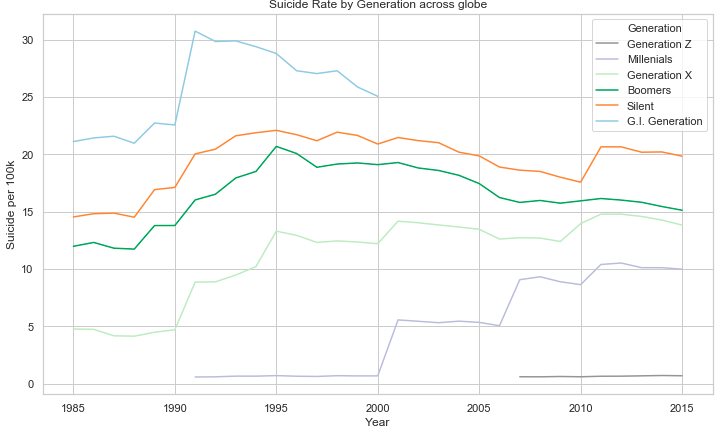
There exists a correlation of 0.43 (weak positive correlation) between suicide rates and HDI, indicating that suicide rates are high among countries with high human development index.

# **Generation and Suicide Rate**

In this section, we want to find out which generation is more susceptible to suicides and also analyze the trend over the years. Below is the plot which depicts the contribution of each Generation towards suicide rate from all the countries and from both genders for the duration of 1985-2015.



The trend of suicide rate over the years for each generation is shown below:



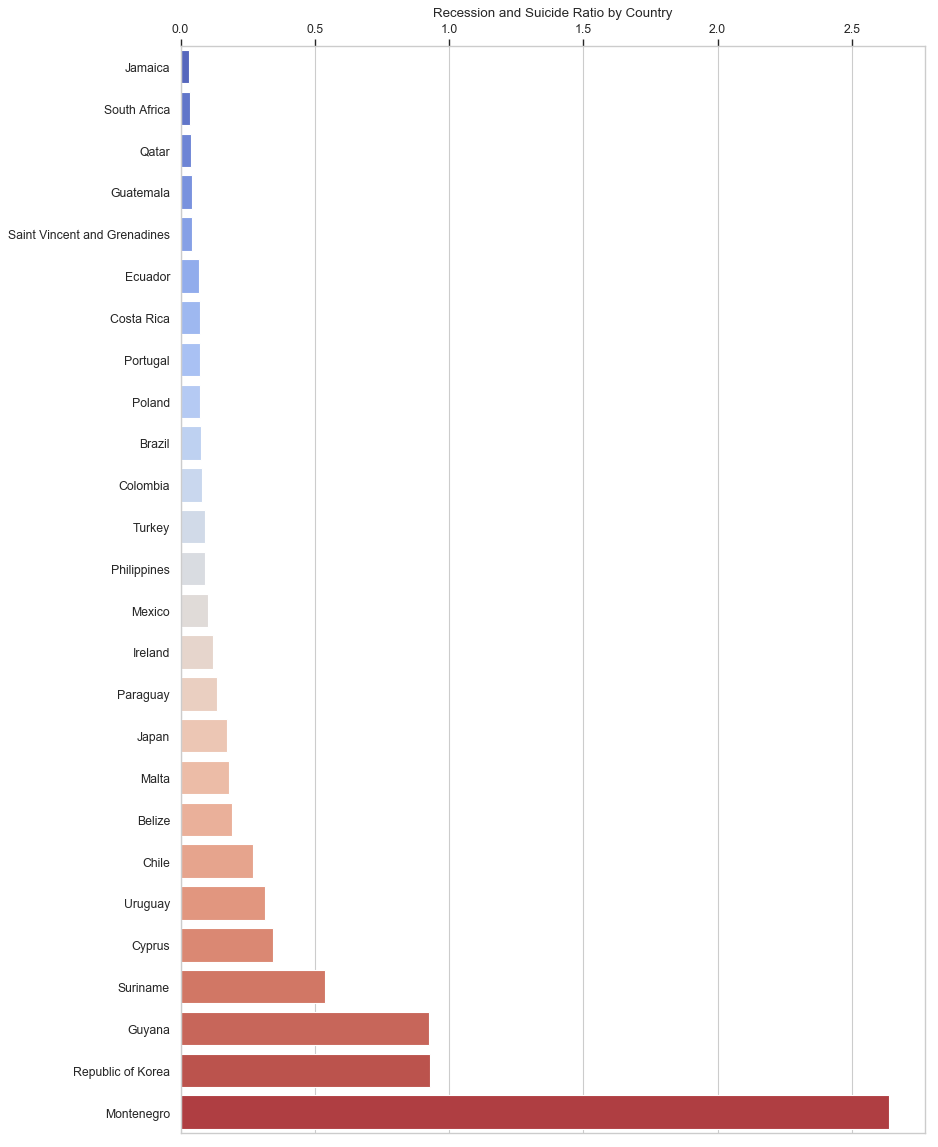
Some of the key findings are:

* We could say Boomers (born 1946-1964), Silent (born 1921-1945) and Generation X (born 1965-1976) are the most suicidal.
* Rate for Generation Z (born 1996 onwards) has plateaued since 2006 but, cannot be assertive due to lack of data.
* Rate for Millennials (born 1977 - 1995) is increasing at an alarming rate.
* Rate for GI Generation (born 1920 or before) appears to be decreasing after 1991 but, cannot be assertive due to lack of data.

# **2007-08 Recession**

To analyze if there was a rise in suicide during 2007-08 recession, we used the GDP data between years 2005 and 2008 and number of suicides during the same time period.

We plotted the Countries that showed an 'Increasing Suicide Rate Trend.'



The key insights from the above plot are:

* Out of 90 Countries available in the dataset, only 26 countries showed significant positive trend in suicide rate during the recession
* Also, a weak negative correlation is seen between years 2005 to 2008 and suicide rate is found during the same time frame.

# **External Factor**

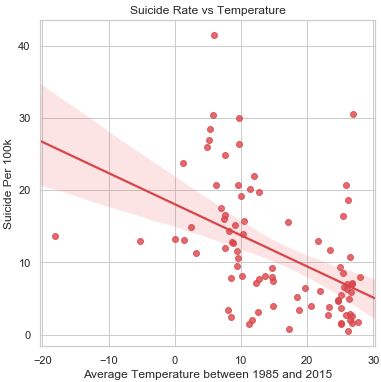
One can argue that temperature is related to sun exposure, sun exposure is related to Vitamin D and Vitamin D has been studied as a factor related to mental health and depression.

One can also argue that sun exposure is related to Melatonin, Melatonin is related to sleep quality and sleep quality is related to mental health.

That's why we decided to briefly explore these two factors together. Let's see what our data tells us.

The data was obtained on Kaggle and was processed and filtered to only show data by year and countries included in our suicide data. We merged the dataset with data from 'Climate Change: Earth Surface Temperature Data' to get the average.

In the below plot, every data point is a country’s average temperature and its suicide rate.



Below are our observations:

* For the countries in our data average temperature play a role!
* We found a weak negative correlation of ~-0.4 between Average Temperature of a Country and its suicide rate. This means that countries with warm temperature has lesser Suicide rate. But temperature is not the only factor.

# **Conclusion**

After analyzing the data, we can conclude that:

* Suicide rates are decreasing globally.
* Of those countries that show clear linear trends over time, 2/3 are decreasing.
* On average, suicide rate increases with age.
* This remains true when controlling for continent in the Americas, Asia & Europe, but not for Africa.
* There is a weak positive relationship between a countries GDP (per capita) and suicide rate.
* The highest suicide rate ever recorded in a demographic (for 1 year) is 225 (per 100k population).
* There is an overrepresentation of men in suicide deaths at every level of analysis (globally, at a continent and country level).
* Globally, the male rate is ~3.5x higher.
* Average temperature of a country has a weak negative correlation with its suicide rate. But this is not a Universal factor.

# **References**

United Nations Development Program. (2018). Human development index (HDI). Retrieved from <http://hdr.undp.org/en/indicators/137506>

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#](http://databank.worldbank.org/data/source/world-development-indicators)

[Szamil]. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from <https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook>

World Health Organization. (2018). Suicide prevention. Retrieved from <http://www.who.int/mental_health/suicide-prevention/en/>

<https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data#GlobalLandTemperaturesByCountry.csv>