INTERNATIONAL BURCH UNIVERSITY

FACULTY OF ENGINEERING AND NATURAL SCIENCES

DEPARTMENT OF INFORMATION TECHNOLOGIES

World Suicides Analytics

BIG DATA ANALYTICS

Ibrahim Muzaferija

MENTOR

Assist. Prof. Dr. Zerina Masetić

SARAJEVO

2019

Table of Contents

Table of Contents **1**

Abstract **2**

Introduction **3**

Methods and Materials **4**

Data preparation 4

Dataset Creation 6

Results **7**

Discussion **14**

Conclusion **17**

References **19**

Abstract

Global suicide trends of a 35-year long period are presented, highlighting particularities in means of suicide patterns across different regions of the world. The global data is examined with regards to age, gender, population, suicide numbers, suicide rate, and in relation to economic factors (i.e. gross domestic product) of countries. Moreover, trends in suicide rates over time are presented on the population level, as well as country, age group and gender levels.

Introduction

Suicide is one of the major health problems and among leading causes of death in most countries in the world [1]. The effects of suicide go beyond the person who acts to take own life, it can have a lasting effect on family, friends, and communities.

Suicide often comes from a deep feeling of hopelessness. The inability to see solutions to problems or to cope with challenging life circumstances may lead people to see suicide as the only option to what is really a temporary situation. Depression is a key risk factor for suicide, while others include psychiatric disorders, substance abuse, chronic pain, a family history of suicide, and a prior suicide attempt. Impulsiveness often plays a role among adolescents who take their life. If a person expresses thoughts or plans of suicide, it's essential to initiate a conversation to fully explore the person's thoughts and emotions, and also follow up with the person over time [2].

Edwin Shneidman's [3] theory of suicide describes psychache (i.e., emotional or psychological pain) as the primary motivator of an attempt. He states that suicide occurs when an individual's threshold for tolerating psychological pain is surpassed and that this threshold varies across individuals.

According to the WHO (World Health Organisation), close to 800 000 people commit suicide every year, which is one person every 40 seconds. Effective and evidence-based interventions can be implemented at population, sub-population and individual levels to prevent suicide and suicide attempts. There are indications that for each adult who died by suicide there may have been more than 20 others attempting suicide [4]. Those attempts are not going to be analyzed in this paper but may have significance in understanding the suicides as a global phenomenon.

In this paper, the focus will be suicide analytics on the population level in order to extract possible suicide patterns, risks, and factors that occurred throughout the ages of 1979 and 2016. The dataset [5] was acquired from WHO and contains information from 141 countries. It is structured in a way that each data sample has a country name, year of recording, age group (young, young adult, adult, senior adult, senior, old), gender, and total population related to age group and gender.

Furthermore, the analysis will try to present statistics and find connections between the suicides and major happenings in those years, as well as the behavior of age groups in different countries over the years.

Methods and Materials

This research is conducted using RapidMiner - a data science software [6], and Python3 programming language [7]. The suicide data, as mentioned above, was obtained from WHO which is an agency of the United Nations that is concerned with international public health. It was established in 1948 and is a member of the United Nations Development Group, making it a reliable source of information.

Data preparation

In order to achieve reliably meaningful results, data needs to be refined and its consistency analyzed. Dataset doesn't contain duplicate values, but there were missing values in suicide and population column. Filling missing suicide values by an average of previous and next year values would corrupt the data, so those missing examples were discarded. Missing population values could be filled with an average value from previous and next year, but countries that are missing population values are missing the values from every year, thus those examples were discarded as well.

country	year	sex	age	suicides	population
Albania	1989	female	35-54 years	7.0	288600.0
Albania	1989	female	5-14 years	0.0	321900.0
Albania	1989	female	55-74 years	1.0	149600.0
Albania	1989	female	75+ years	0.0	37000.0

Figure 1 - Initial dataset sample

Next, there are 141 countries and they have an uneven number of samples. Some countries have around 450 samples, while others have only 12. This can result in having an increased number of suicides in countries with more examples and vice versa. To avoid such oscillation, shown in *Figure 2*, countries having below an average number of samples were discarded.

Number of samples over the years has inconsistency as well (*Figure 3*), starting and ending years count the least samples, while the overall number of samples increased over time. This might result in an increase of suicides over the years but can be avoided by crossing the graphs. Two ending years (2015 and 2016) were discarded due to the insufficient number of samples.

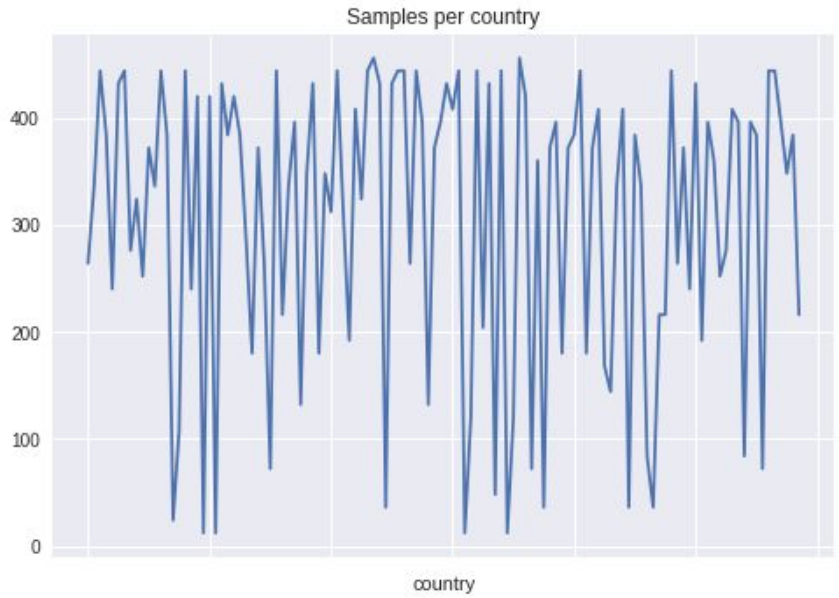

Figure 2 - Oscillation of samples per country

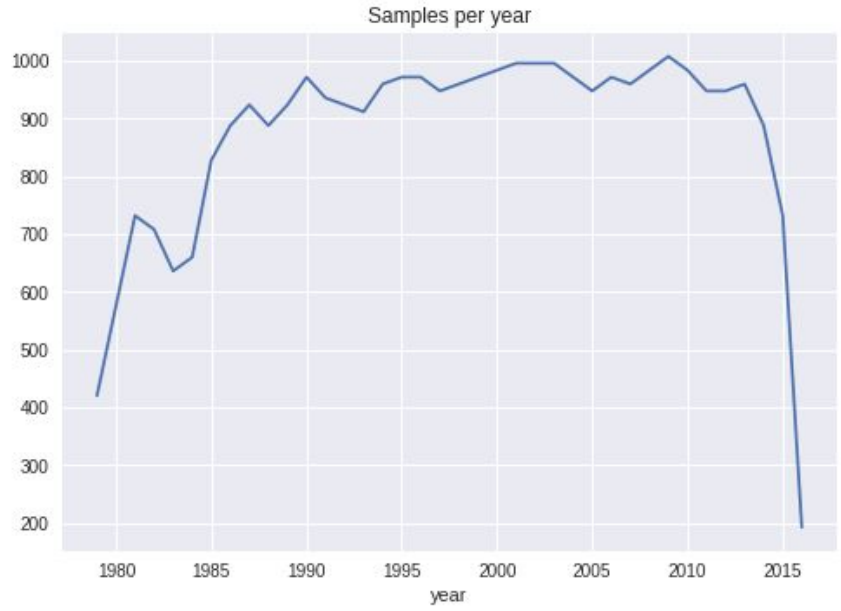

Figure 3 - Oscillation of samples per year

Furthermore, the number of suicides is proportional to the size of the population. Countries with bigger population size will have a higher number of suicides, making the comparison in the number of suicides between countries and age groups uneven. To normalize the suicide numbers, the suicide rate is obtained by calculating the percentage of the population that committed suicide, that is dividing the number of suicides by population size and multiplying it by hundred.

Dataset Creation

In the data preparation phase, data cleansing was performed by removing samples with missing values, countries with below average sample size, and years with the lowest sample size. Also, new data was derived by calculating the suicide percentage.

After the data preparation phase, the dataset size decreased by 35% and numbers 28,380 samples. The table below depicts the dataset attributes description, as well as statistical information.

Table 1 - Final dataset attribute description

Attribute	Data type	Range	Missing values	Distinct values	Unique values	Statistics
index	integer	[0,28379]	0	28380	28380	—
country	nominal	Antigua and Barbuda, Argentina, (...)	0	73	0	Least: Grenada (288) Most: Italy, (...) (432)
year	integer	[1979, 2014]	0	35	0	Least: 1979 (408) Most: 2009 (876)
sex	nominal	male, female	0	2	0	Least: male, female (14190) Most: male, female (14190)
age	nominal	5-14 years, 15-24 years, (...)	0	6	0	Least: 5-14 years, (...) (4730) Most: 5-14 years, (...) (4730)
suicides	integer	[0,22338]	0	0	0	Min: 0 Max: 22338 Average: 264.78
population	integer	[3.890000e+02, 4.380521e+07]	0	0	0	Min: 3.890000e+02 Max: 4.380521e+07 Average: 1.908851e+06
percentage	float	[0,0.219224]	0	0	0	Min: 0.000000 Max: 0.219224 Average: 0.014317

Results

Analytics results are presented in four sections: gender, population, suicides, and suicide percentage, each reflecting a standpoint in comparison to other subsequent parts.

In relation to gender, the leftmost graph in the figure below shows summed up suicide percentages over the 35-year period. The graph in the middle illustrates how many suicides were committed over the years in total, while the rightmost graph shows the total population size over the years.

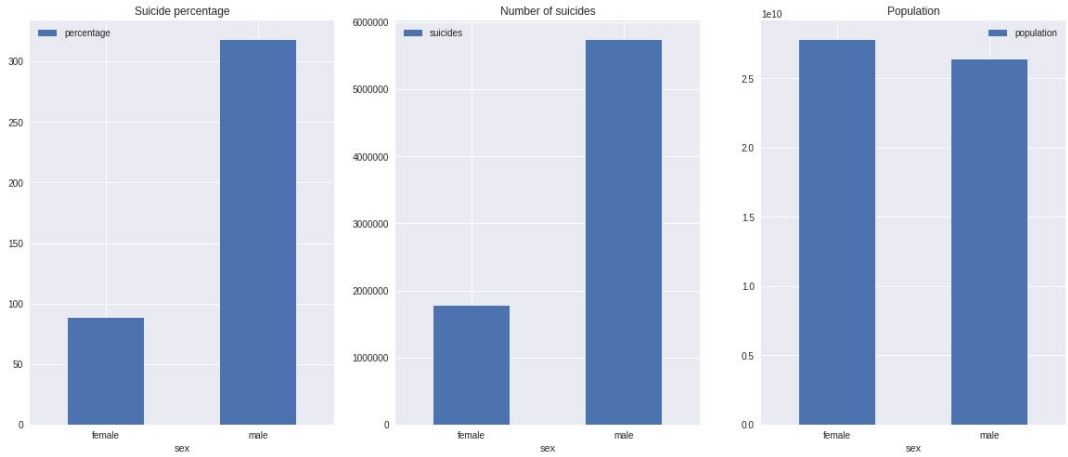

Figure 4 - Percentage, suicides and population size per gender

Population varies in size between age groups, as the left graph in the figure below shows, most numbered age group is between 35 and 54 years, while the least numbered age group is above 75 years. Total population size for each year is depicted in the right graph.

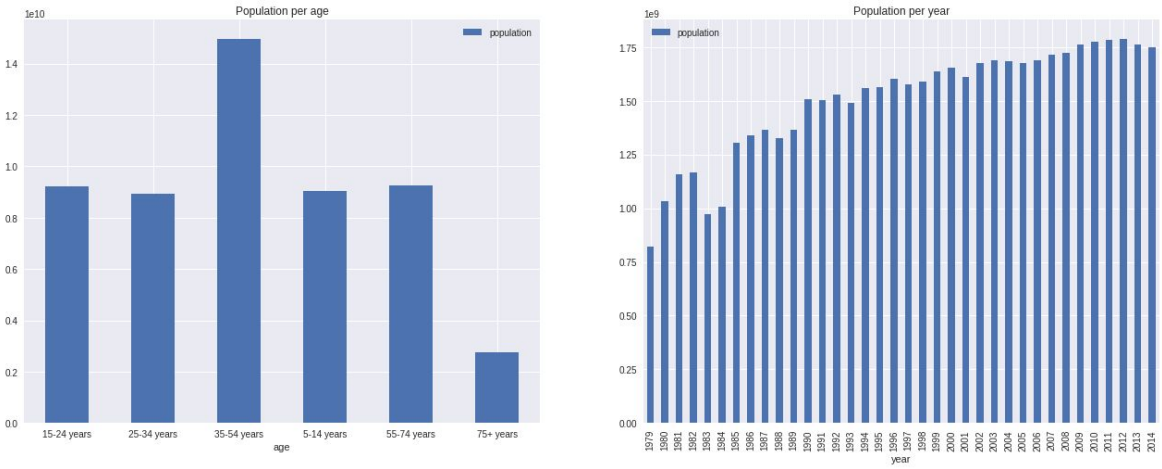

Figure 5 - Population size per age group and year

Population size varies per countries as well, shown in *Figure 6*, where ten countries have the population as twice as bigger than the average: USA, Brazil, Russian Federation, Japan, Mexico, France, Italy, Germany, United Kingdom, and Thailand.

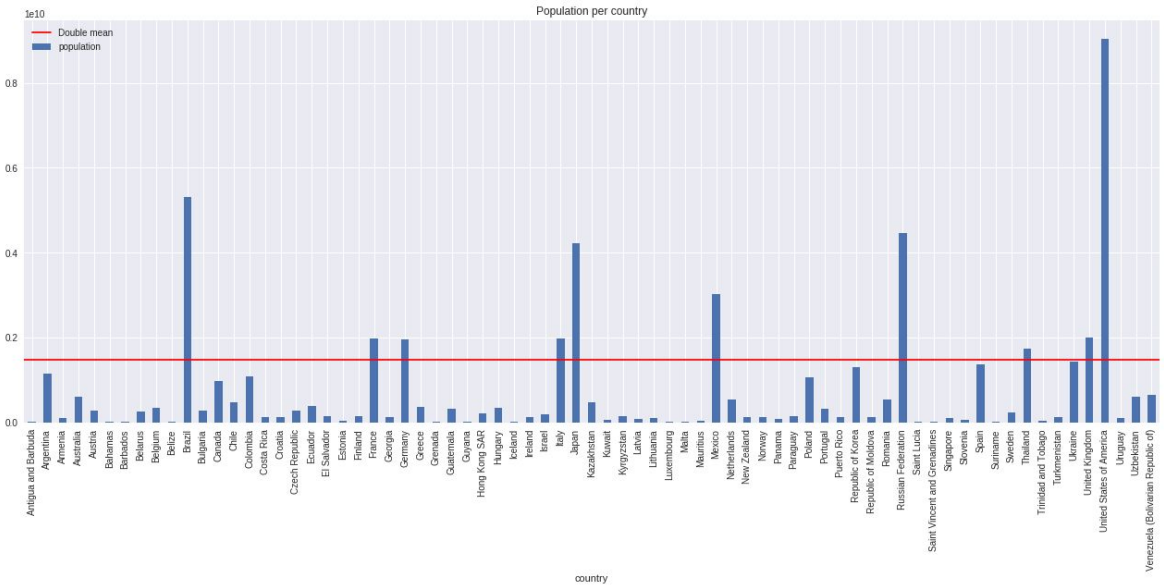

Figure 6 - Population size per country

In relation to the suicides that occurred between the ages of 1979 and 2014, the highest fraction belongs to the age group of 35-54 years and 55-74 years, respectively. The number of suicides per year is shown in the right graph of the same figure below.

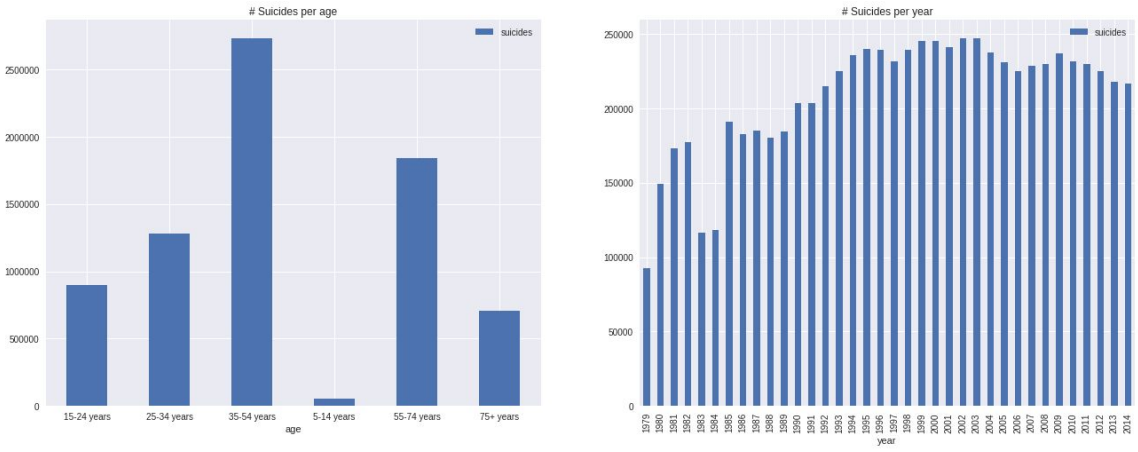

Figure 7 - Number of suicides per age group and year

The total number of suicides per country is shown below. It is noticeable that eight countries have suicides above twice as average: Russian Federation, USA, Japan, France, Ukraine, Germany, Republic of Korea, and Brazil, respectively.

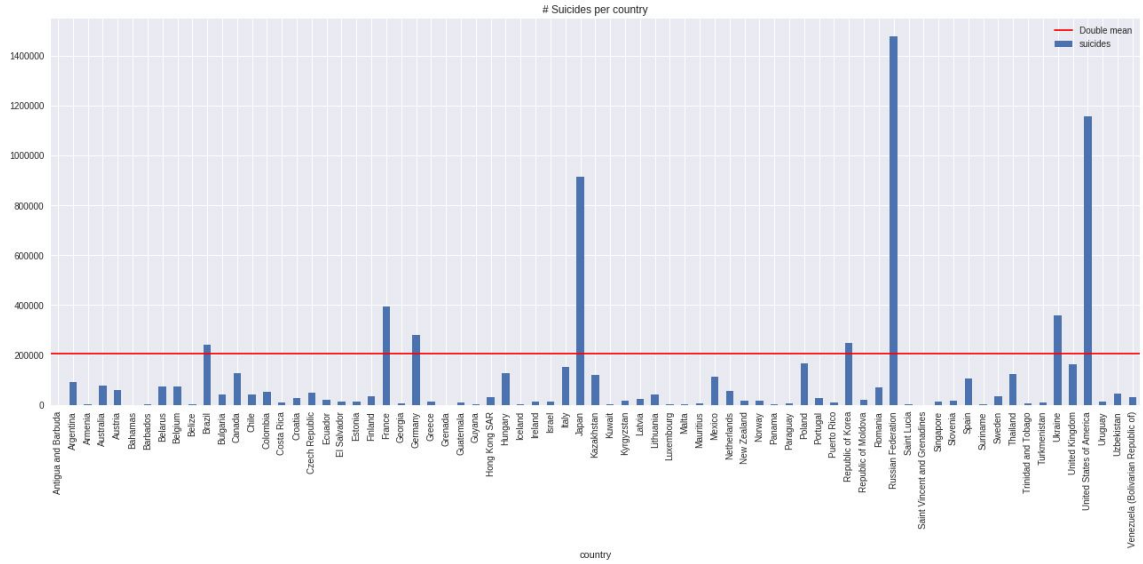

Figure 8 - Number of suicides per country

The total number of suicides per year and sex yields interesting results. Males have a much higher number of suicides committed, peaking in the nineties, while females have low steady counts of committing the suicide.

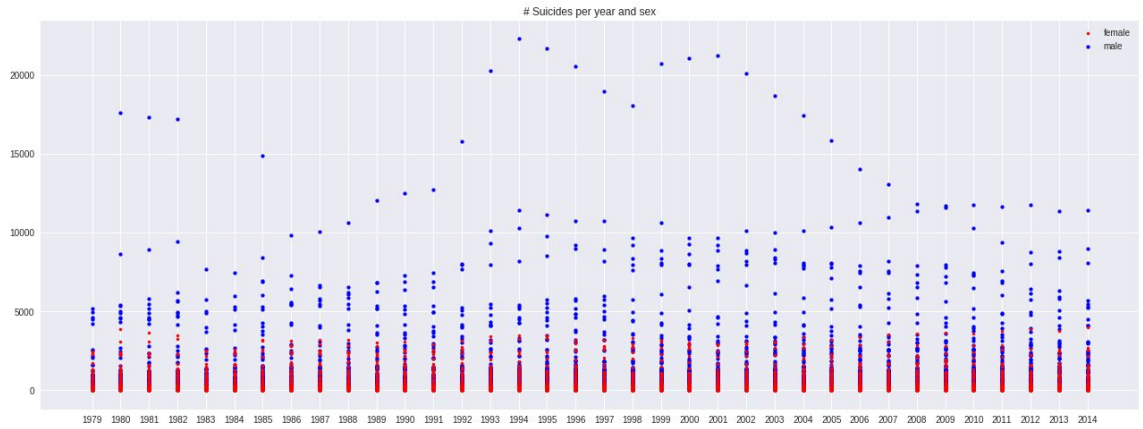

Figure 9 - Number of suicides per year and sex

By introducing age groups to the previous graph, the peak was taken by the age group of 35-54 years and followed by the next older age group, then the younger group. Other age groups have a steady rate of suicide numbers.

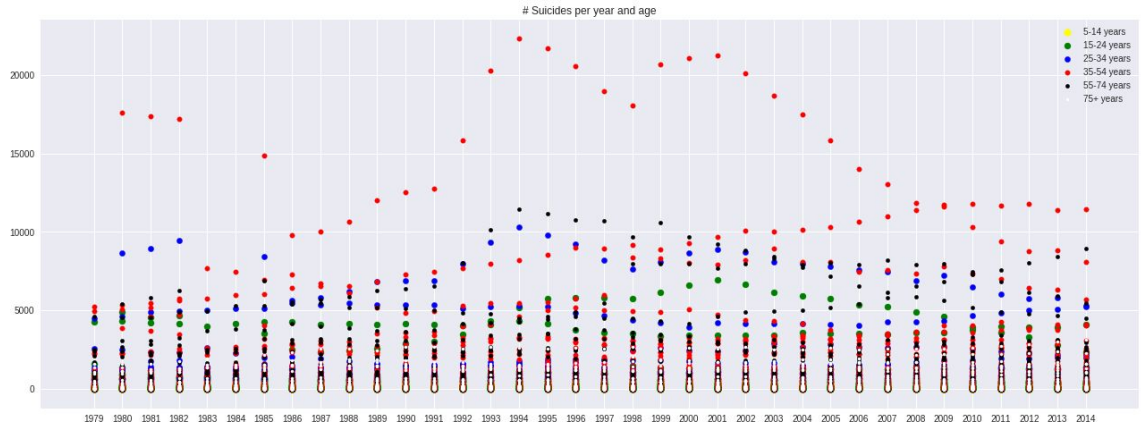

Figure 10 - Number of suicides per year and age

As previously mentioned, by inducing the suicide percentage, that is, calculating the percentage of the population that committed suicide, more accurate results can be obtained. In the figure below, the left graph depicts suicide percentage in age groups, while the right graph illustrates the suicide percentage per years.

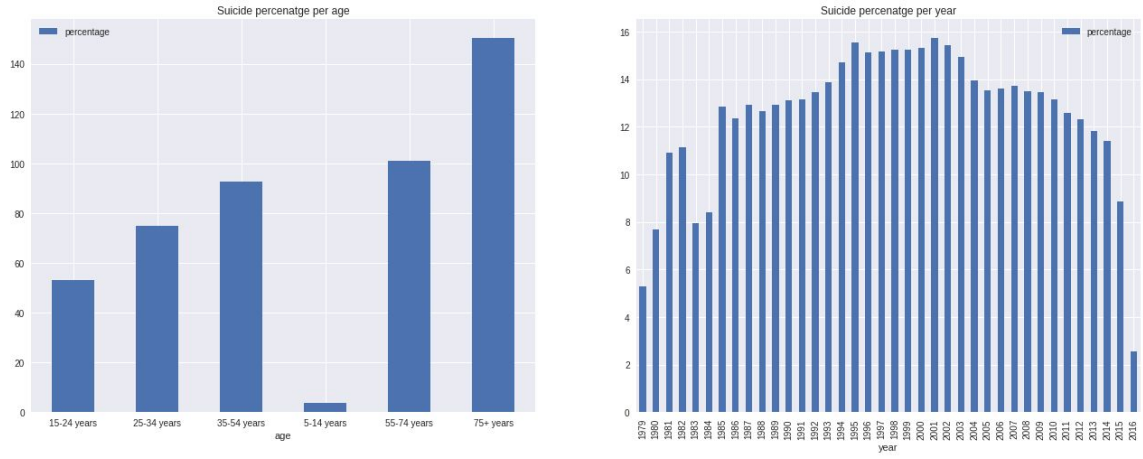

Figure 11 - Suicide percentage per age group and year

Previously, in *Figure 8*, the number of suicides per country was shown and eight countries were above double average. The next figure shows suicide percentage per country, having seven countries that are above double average: Hungary, Lithuania, Russian Federation, Latvia, Estonia, Kazakhstan, and Slovenia, respectively.

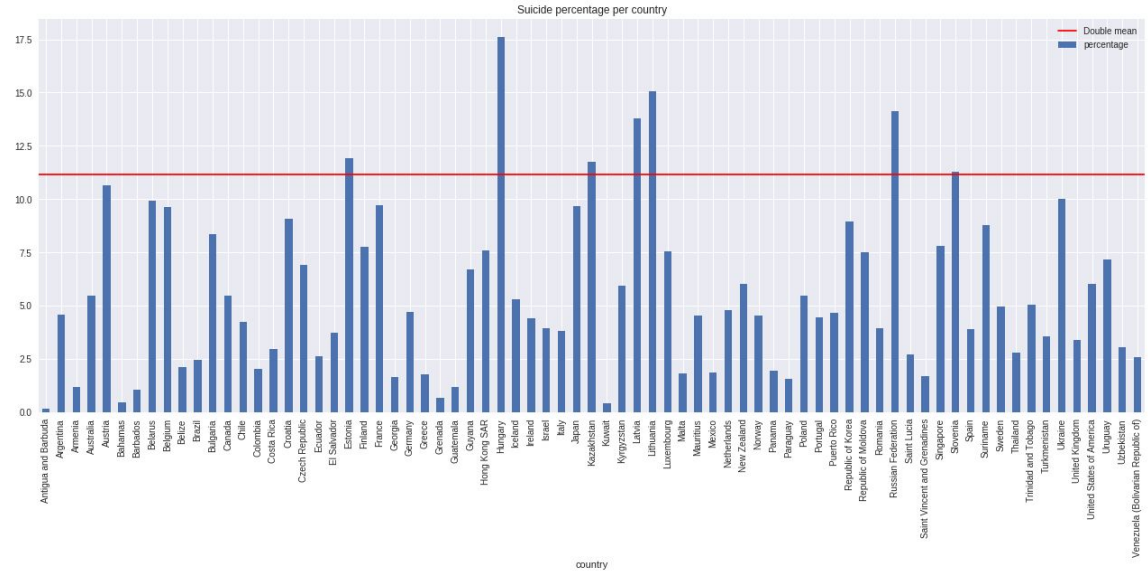

Figure 12 - Suicide percentage per country

While *Figure 9* illustrates the number of suicides per year and sex, and *Figure 10* illustrates the number of suicides per year and age, having a peak in the nineties, two figures below illustrate the same results but in relation to suicide percentage, confirming the peak and adding standalone cases of increased suicide percentages in ages before and after.

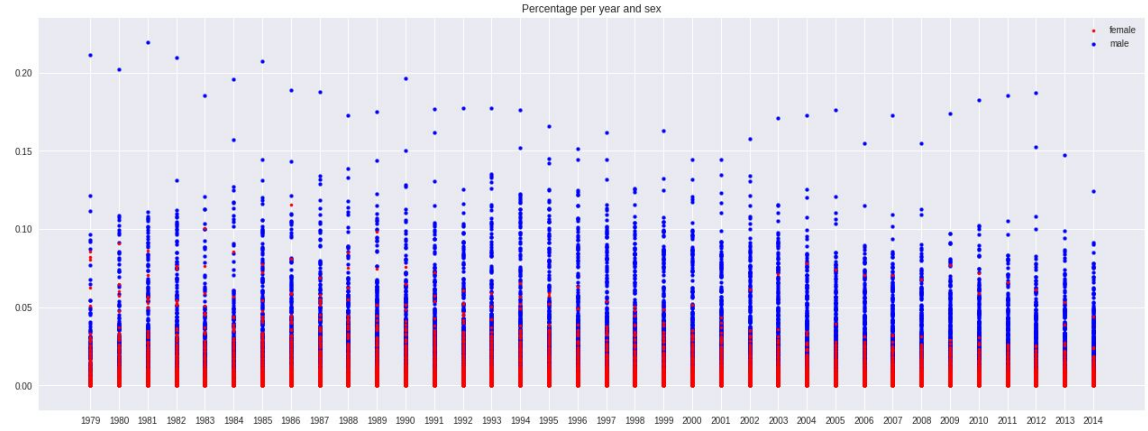

Figure 13 - Suicide percentage per year and sex

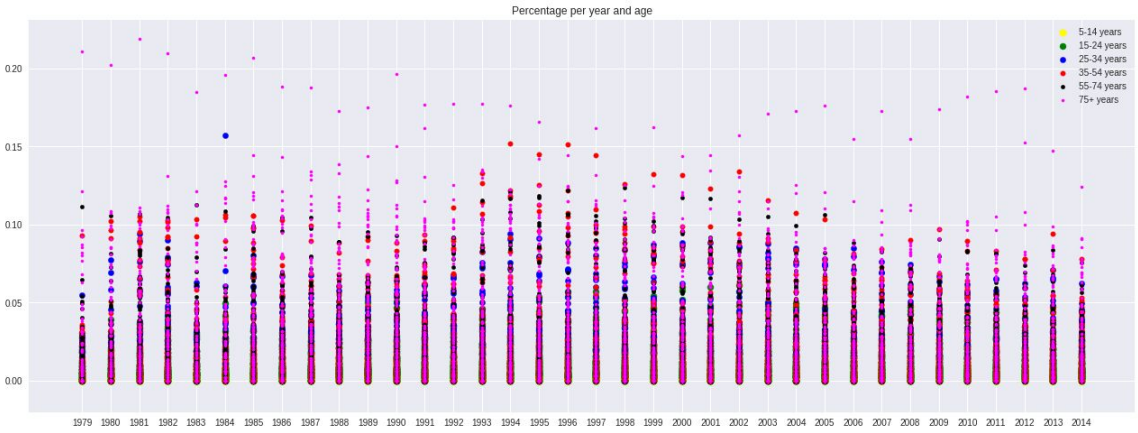

Figure 14 - Suicide percentage per year and age

The most interesting result was obtained by crossing gender, age groups, and suicide percentages with years. The figure below depicts suicide percentages for each year, divided into groups by age, and colored by the gender. Increases and decreases of suicides by year are easily noticeable for each age group, as well as gender.

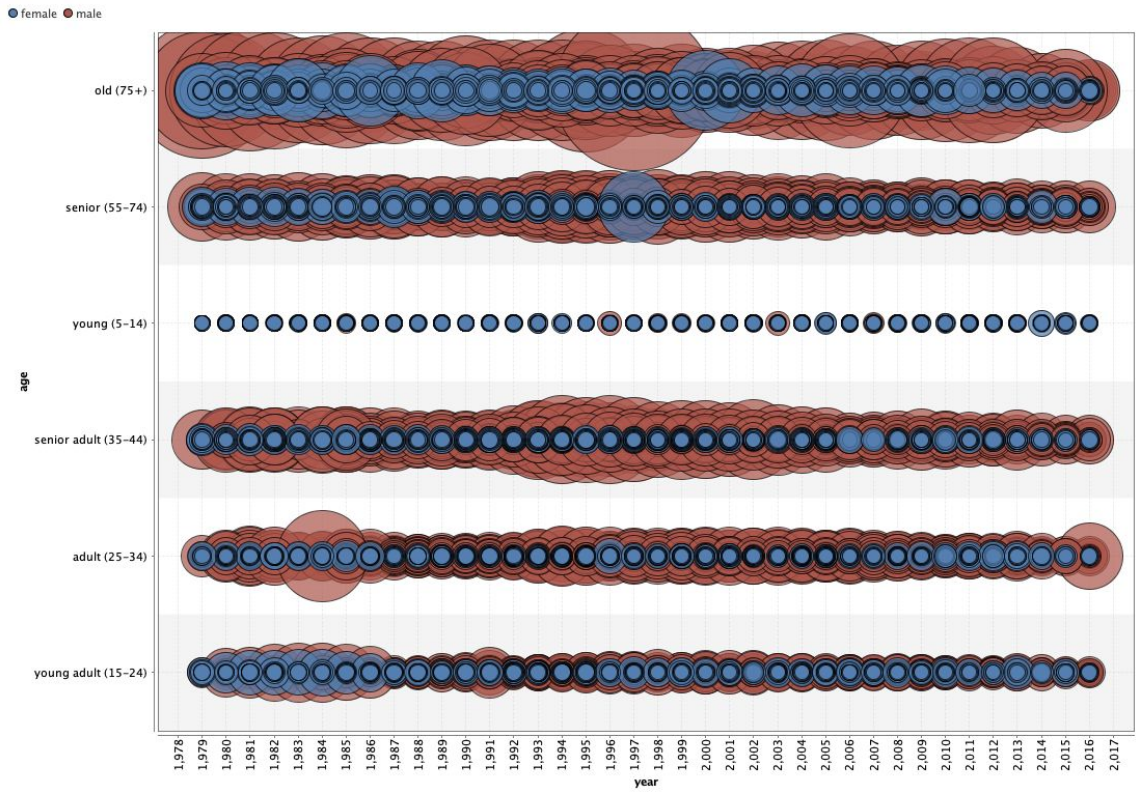

Figure 15 - Suicide percentage per age group, gender, and year

In order to find a connection that could result in an increase or decrease of suicide percentages in countries, an external dataset was introduced, containing the GDP data for each country and each year. The figure below illustrates total GDP per capita of each country, having three lines that

represent the double mean, mean, and a half mean, in order to classify countries as having low GDP and as having high GDP per capita.

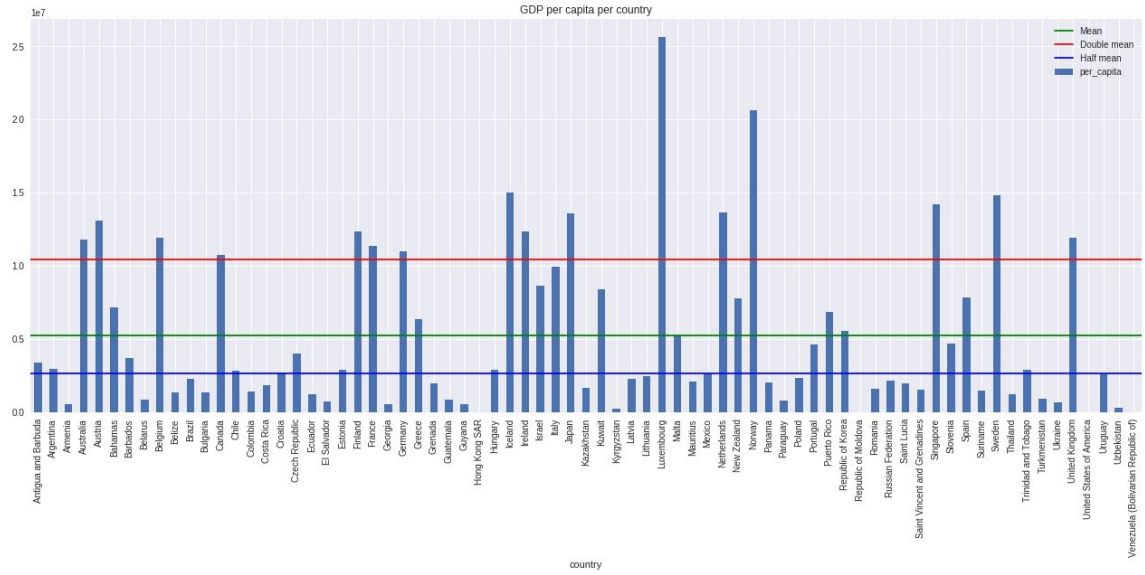

Figure 16 - GDP per capita per country

While the yearly GDP per capita graph contains high variations in data, making it hard to interpret, more easily interpreted and more informative graph, containing stacked GDP on top of suicide percentages per country is shown below.

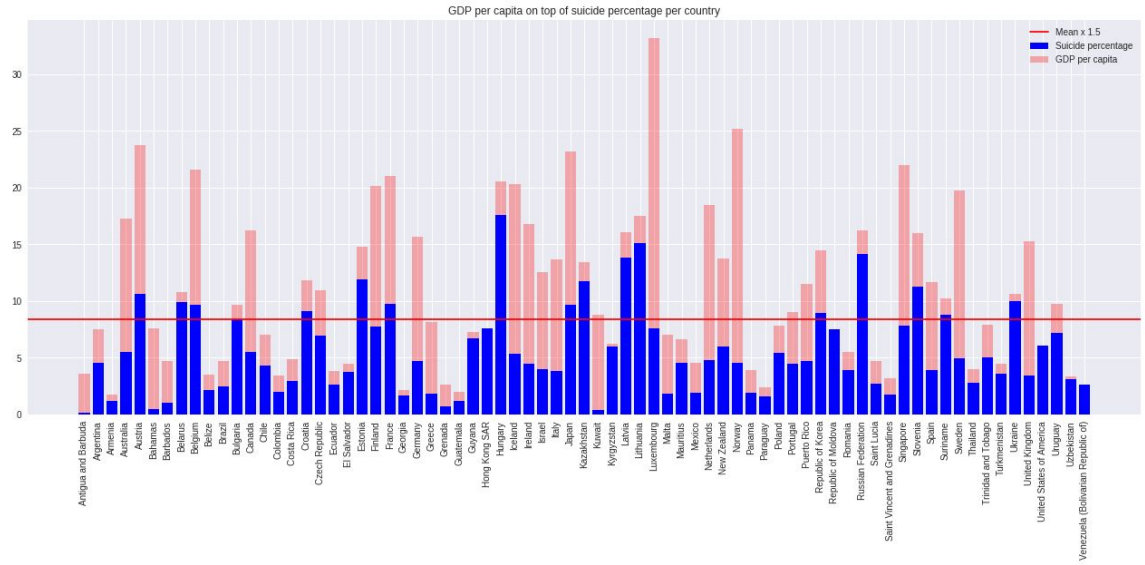

Figure 17 - Country GDP per capita on top of suicide percentage

Discussion

In relation to gender, the leftmost graph in *Figure 4* shows that male suicide percentage is more than three times higher compared to female suicide percentage. The comparison between the number of suicides committed shows a slight increase for females, as the graph in the middle shows, due to female population size being slightly bigger than males - the rightmost graph. But why is the male suicide rate much higher than females? Daniel Freeman, a researcher, and professor of clinical psychology [8] explains that reason might be the lethality of suicide attempts. Women who attempt suicide tend to use nonviolent means, such as overdose, while men often use firearms or hanging, which are more likely to result in death.

While the right graph in *Figure 5* shows the total population size over the years, inconsistency is noticeable in the years 1983 and 1984, having a sudden decrease in population size. After observation of sample sizes gathered for each year, the similar inconsistency is found, resulting in the inconsistency of population size. The figure below depicts deviation in population size caused by the number of samples for each year. Two previously mentioned years with a sudden decrease in the population are connected with the sudden decrease in the number of samples for the same years. While the number of samples is inconsistent, it is still noticeable that population size increased over the years.

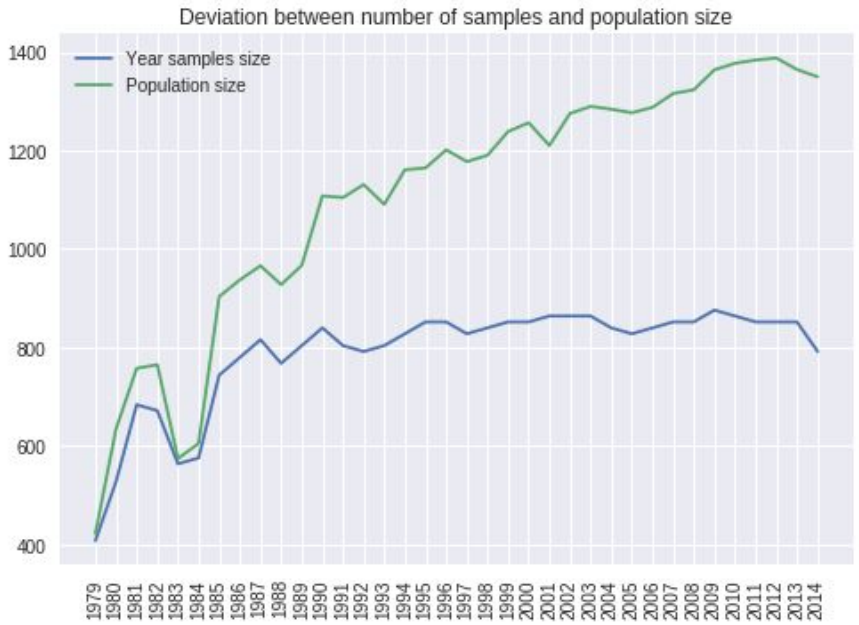

Figure 18 - Deviation in population and year samples size

Ten countries whose population size exceeds double mean value (*Figure 6*) have the highest population size. These ten countries don't have the most numbered population in the world [9],

as not all countries are present in the dataset, but are indeed amongst the highest populations in the world.

The highest number of suicides is present in the age group of 35-54 years (shown in the left graph of *Figure 7*), this result is due to the same age group having the highest fraction of the population. Moreover, the percentage of an age group that committed the suicide is presented in *Figure 11* and clearly shows that highest suicide rate is presented in the age group of 75+ years, followed by age group of 55-74 years, which means the older the people get, the more likely they will commit suicide.

The number of suicides per year shows noticeable inconsistency (right graph in *Figure 7*). Again, by crossing the number of samples by the number of suicides, as shown in the figure below, inconsistency in linked and it is shown that the number of suicides had an increase between the ages 1989-1995, 1997-1999, and 2006-2009 and a decrease between the ages of 2002-2006 and 2009-2014.

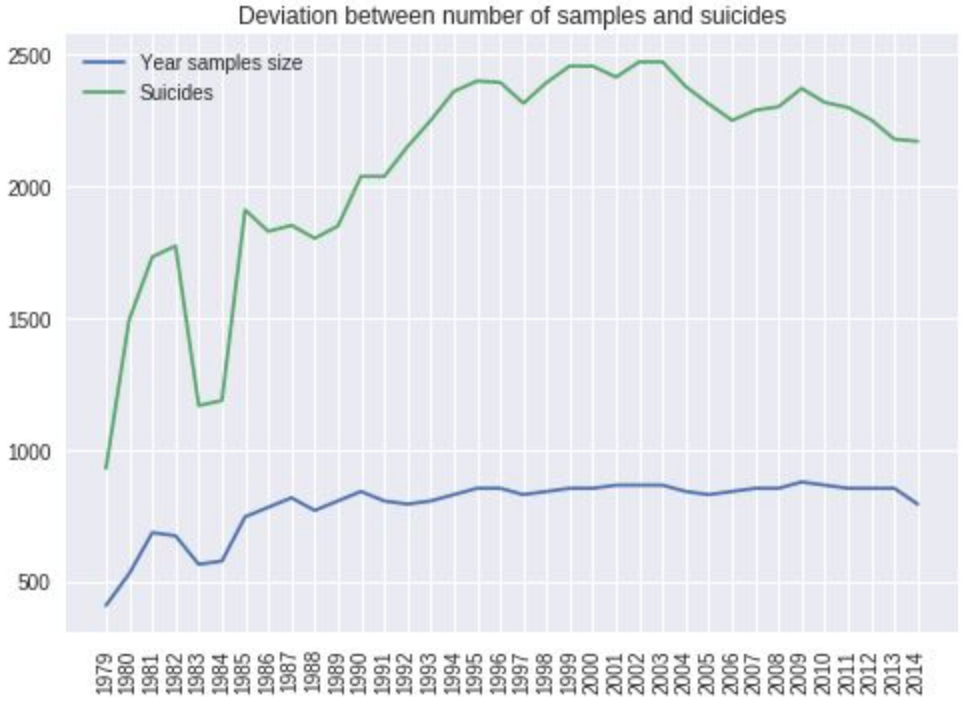

Figure 19 - Deviation in suicides and year samples size

Figure 8 shows the number of suicides per county, eight having numbers twice above the average. Some of those countries have highly numbered population which might yield highly numbered suicides, while on the other hand countries having smaller populations size yield a smaller number of suicides, making them negligible in comparison to prior. More accurate results, presenting suicide rates per country, are shown in *Figure 12*.

The suicide percentage per year shows the same consistency as in previous per year graphs. By crossing the yearly number of samples with suicide percentage per year, the inconsistency is linked again. Also, a slight increase in suicide rate is noticed in the ages between 1985 and 1995, while a strong decrease in suicide rate follows afterward.

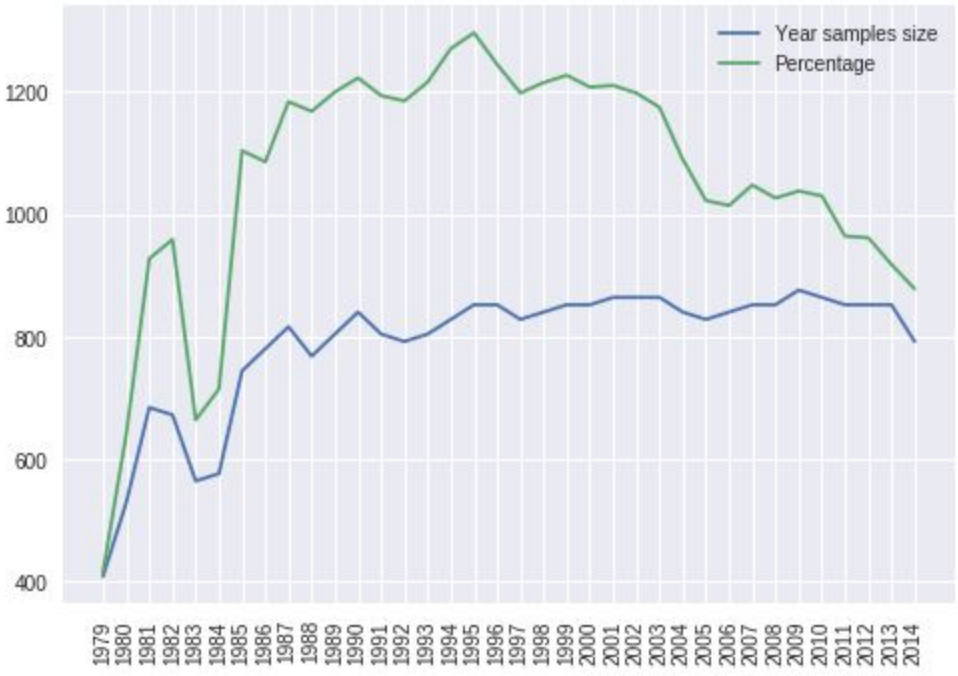

Figure 20 - Deviation in suicide percentage and year samples size

Conclusion

The graph in the middle of *Figure 4* shows that males commit suicide over three times more than females. Figures 9,13, and 15 confirm these results.

From the beginning of the 80s, the overall suicide rate was increasing and had its peak in the 90s, after which was strongly decreased, as shown in figures 13,14,15, and 20.

Countries with the highest number of suicides are the Russia, USA, and Japan, while the countries with the highest suicide rate are Hungary, Lithuania, Russia, and Latvia. These results can be confirmed in figures 8 and 12.

Figure 13 and *Figure 14,* depicting suicide percentage per year and sex, and per year and age respectively, show a steady decrease in the suicide rate in the ages after 1995. These figures also confirm what was previously noticed, that is, male suicide percentage is much higher than females, and people in the age group of 75+ years have the highest suicide percentage of all age groups.

In order to find a peak in suicide percentage for each age group, *Figure 15* was made by crossing suicide percentage with age groups, gender and years, resulting in an easily interpreted graph. It is confirmed that the age group of 5-14 years has the lowest fraction of suicides for both sexes, almost negligible in comparison to other groups.

Age group of 15-24 years had a peak for both sexes in the early 80s, and second, much smaller peak, in the late 90s. Males belonging to this age group constantly occupy a slightly higher fraction of the suicides.

Third age group, that is 25-34 years, had a peak of male suicides in ‘83, while an increase in male suicides was between the ages of 1993 and 2009, followed by a slight steady decrease. Females in this age group occupy around three times smaller fraction of suicides in comparison to males and have a steady rate of committing the suicide, with slight increases in the early 80s and between the ages of 2008-2014.

Age group of 35-54 years, despite being the one with the most numbered population, doesn’t have the highest suicide rate. Males have almost four times higher suicide rate with an increase that began in the 90s and ended in the early 00s. Females have a small steady rate of committing suicide without meaningful increases.

Forth age group, 55-74 years, shares the same pattern and a similar, but slightly smaller, increase period as the previous age group. Females belonging to this age group had increased suicides in the early 80s and late 10s, with a peak in the 1997 year.

Age group of 75+ years has the smallest population fraction and, on the contrary, has the highest suicide rate for both males and females. Relation of male suicide rate to females is, as in previous groups, around three times higher. Males had a decrease in the 80s and an increase in the 00s, with a peak being in the 1997 year. Females had decreased suicide rates only in the 90s, while the peak was in the year 2000.

GDP per capita has a certain impact on suicide rate in countries, as *Figure 17* shows, 23% of countries with a high suicide rate has high GDP per capita, while 77% of countries with high suicide rate has low GDP per capita. This indicates that, in countries with lower GDP per capita, a higher fraction of the population is more likely to commit the suicide.

References

[1] “Sci-Hub: устраняя преграды на пути распространения знаний.” [Online]. Available: https://sci-hub.tw/10.1146/annurev-clinpsy-021815-093204. [Accessed: 12-Jan-2019]

[2] “Suicide | Psychology Today,” *Psychology Today*. [Online]. Available: https://www.psychologytoday.com/basics/suicide. [Accessed: 08-Jan-2019]

[3] E. S. Shneidman, “Commentary: Suicide as Psychache,” *J. Nerv. Ment. Dis.*, vol. 181, no. 3, pp. 145–147, 1993 [Online]. Available: http://dx.doi.org/10.1097/00005053-199303000-00001

[4] “WHO | Suicide data,” Nov. 2018 [Online]. Available: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/. [Accessed: 06-Jan-2019]

[5] “WHO Suicide Statistics.” [Online]. Available: https://www.kaggle.com/szamil/who-suicide-statistics. [Accessed: 08-Jan-2019]

[6] T. Wentworth, “Visual Workflow for Predictive Analytics | RapidMiner© Studio,” *RapidMiner*, 07-Aug-2018. [Online]. Available: https://rapidminer.com/products/studio/. [Accessed: 12-Jan-2019]

[7] “Welcome to Python.org,” *Python.org*. [Online]. Available: https://www.python.org/about/. [Accessed: 12-Jan-2019]

[8] D. Freeman and J. Freeman, “Why are men more likely than women to take their own lives?,” *the Guardian*, 21-Jan-2015. [Online]. Available: http://www.theguardian.com/science/2015/jan/21/suicide-gender-men-women-mental-health-nick-clegg. [Accessed: 13-Jan-2019]

[9] “Ten Countries with the Highest Population in the World.” [Online]. Available: https://www.internetworldstats.com/stats8.htm. [Accessed: 16-Jan-2019]