SUICIDE RATE ANALYSIS FROM THE YEARS 1985-2015 BASED ON G.D.P PER CAPITA, POPULATION, GENDER AND AGE

Submitted By:

Amruth Jaligama

Student Id: 11396245

Under The Guidance Of:

Professor Sadeghinaeenifard, Fariba

INFO 5709-Data Visualization and Communication

Department Of Information Science, University Of North Texas

April, 2020

ABSTRACT

Suicide is the act of killing one's self knowingly. There can be various reasons behind a person committing suicide such as depression, stressful situations, economic factors and many more. The main aim of this project is to know if socio-economic factors have an impact on suicide rate, to see if there is any difference in the suicide rates, suicide count based on gender, age. In this project I have analysed the suicide count of 18 countries selected from 101 different countries present in the dataset, the relationship between suicide rate and gdp_per_capita for the 18 selected countries, the suicide count & suicide rate based on gender, age, suicide rate and gdp_per_capita relationship in U.S.A Country for each year from the year 1985-2015. The data has been collected from Kaggle website, the dataset consists of data from the years 1985-2016 but in this project I am analysing the data from the years 1985-2015 only due to data value errors for the year 2016. Exploratory data analysis was performed to analyse the data in an organized way that is starting the analysis from a basic level and going into the low level details later, Time series analysis was performed to analyse the trends.

INTRODUCTION

According to the World Health Organization Suicide is a serious public health problem across the whole world which is preventable. It is one of the top 10 leading causes of death in developed and developing countries, it is estimated that approximately 1 million people die from suicides every year. Socio-Economic factors, personality disorders, Depression, historical factors are the main reasons due to which a person may commit suicide and some studies found out that Genetics also contribute to suicidal behaviour and suicides (Bondy, Buettner, & Zill, 2006).

Economic fluctuations also have an impact on suicide rate and many people believe that economic crisis can lead to increased suicide rates, a study of economic crisis on psychological well-being tells us that for 26 European countries and for the data collected from the years 1970-2007 for 1% increase in unemployment there was 0.79% increase in suicide rate for people aged less than 65 (Van Hal, 2015).In a Journal article published by (Agerbo, Sterne, & Gunnell, 2007) the impact of Marital status, employment status, the level of income on suicide rate was studied among individual studies and it was found that the suicide risk increased for individuals who lived alone, divorced, unemployed and had low income.

Suicide rates vary from one country to the other, from one region to the other due to many reasons some of them are Economy of that country, Social and Cultural factors. World Health Organization listed the average suicide rate for a region in the year 2016 that is it classified & grouped Countries based on their location into different regions and listed each regions average Suicide Rate in which Europe region had the most average suicides for 100,000 people with an average suicide rate of 15.4 which is the Average suicide rate for all the countries together in the Europe region for that year and Eastern Mediterranean had the lowest average suicides for

100,000 people with an average suicide rate of 3.5 which is the average suicide rate for all the countries together in the Eastern Mediterranean region for that year.

In a book published by the World Health Organization suicide numbers in thousands were listed for different regions for the year 2012 where each region is a collection of countries classified and grouped together into a region based on their economy by the World Bank in which it was found that Low middle income region had highest suicides in thousands with 333,000 suicides in the year 2012 and Low income region having the lowest suicides in thousands with 82,000 suicides in the year 2012 (World Health Organization).

Study & Analysis of suicides for Different countries by (Ying & Chang) found out that G.D.P (Gross Domestic Product) impacts suicide rates irrespective of the gender in most of the cases although there are some exceptions to this. When the economy of a country is good people of that country have less stress resulting in decreased suicide risks but when the economy of the country is not that good and people lose their jobs people of that country have more stress and depression which leads to increase in the risk of a person committing suicide. The same study found out that G.D.P doesn't have an impact on suicides for middle age females the reason being different age groups have different reasons for committing suicide and Gender also plays an important role when it comes to a family because people of different genders in the same family have different sources of stress and pressure.

In the research work carried out by (Mościck) it was found that the suicide rate for males was higher when compared to females. Even though Females have more suicidal thoughts and have more number of attempted suicides when compared to males but the completed suicides for males is more than female suicides due to various reasons some of the explanations say males choose more lethal methods when committing suicide, Some other explanations say that men are told to be strong from their childhood and told not to cry and are told that men shouldn't be

crying this results in a behaviour being developed which results in not sharing their worries, problems and not talking to a psychiatrist or counsellor when they have suicidal thoughts which results in increased suicides rates for males and coming to females they go to a psychiatrist or a counsellor if they have any suicidal thoughts which results in less suicide rate for females.

In a book written by (Durkheim & Suicide) it is stated that the suicide rate for males is greater when compared to females and it is stated that risk of suicide increases with age as socializing decreases which tells us that socializing is also a factor for suicides and it is also stated that suicide rate are higher in Scandinavian countries. There are different reasons for different age groups to commit suicide and the suicides vary for different age groups in different countries. From a study conducted by (Girard, 1993) and for the data collected for 49 countries from the years 1976-1980 it can be said that the suicide rates by age is highest for males whose age is above 65 and the suicide rate was the lowest for females of age group 25-44 although the suicide rate varies for countries in the continent of Asia and Africa for this countries the suicide rate is highest for people in the age group of 20-40 and suicide rates for younger women are greater than the suicide rates of young men due to various reasons some of the reasons being culture and influence of other people.

From the suicide analysis statistics for U.S.A during the years from 1997-2017 it was found that the suicide rate increased from 10.5 to 14.0, The suicide rates did not increase abruptly, the increase in suicide rates followed a pattern, During years 1999 to 2006 the suicide rates increased by 1% on average for each year and the suicide rate increased by 2% for each year from the years 2006-2017. The suicide rate is higher for males for both years 1999 and 2017 when compared to the suicide rates of females (Hedegaard, Curtin, & Warner, 2018).

According to an article written by (Weir, 2019) it is stated that for many countries the suicide rate decreased for the year 2012 when compared to the year 2000 but the suicide rate for U.S.A

increased for the year 2012 when compared to the year 2000. It is difficult to find the exact reason for increased number of suicides as there are many reasons behind people committing suicide. Few of the reason being U.S.A was a bit slow to develop preventive measures against suicides while other countries have taken a lot of measures since many years ago to prevent suicides, other reason being the lack of effective mental care, Many people in U.S.A having access to firearms. It is believed that restricting access to death means can also reduce the number of suicides for example In England people could kill themselves by the use of chemicals that were used for cooking such as carbon monoxide and coal, Once when the Country switched to less toxic gases for cooking the suicide rate decreased by 40%, In many of the Asian Countries people had access to pesticides and insecticides which are used in farming, As the countries in Asia made public access difficult to these toxic chemicals the suicide rates dropped in Asian countries. In U.S.A as the residential access of guns fell during 1990's committing suicides by using firearms decreased which resulted in a notable suicide drop but the suicides committed by using other means still remained the same.

METHODS

In this project I have performed exploratory data analysis & time series analysis, developed various hypothesis and questions, tested the hypothesis, and answered various questions by analysing and visualizing the data that we encountered while performing the process of exploratory data analysis and Time Series Analysis.

By performing Exploratory Data Analysis we start from the basic questions such as what is the suicide count for the selected 18 countries from 101 countries in the dataset from the years 1985-2015, what is the average suicide rate for the selected countries from 1985-2015, which countries have the highest & lowest average suicide rate ,what is the average gdp_per_capita for the selected 18 countries, which countries have the highest and lowest Average

gdp_per_capita in the selected countries list from the years 1985-2015. The 18 countries are selected in a way that most of the regions in the world are covered when the countries are classified and grouped into regions based on their locations, regions are divided into four those are North, South, East, and West. As we perform this process of EDA our questions become more and more specific going down the line, we develop various hypothesis and questions such as is there a relationship between G.D.P and suicide rate, does the suicide count increase with increase in the population and suicide count in the selected countries list from the years 1985-2015.

By performing Time Series Analysis for the Country U.S.A we analyse the change in trends of suicide rate, suicide counts from the years 1985-2015 based on G.D.P, Gender and Age.

DATA VISUALIZATION & VISUALIZATION TOOLS USED

When we generally look at numbers we don't get an idea regarding what that data tends to say but when we visualize the same data we understand what the data is actually stating by looking at the graphs, plots, charts, maps and images.

Data visualization is the representation of information and data by using different visual elements such as Charts, Graphs, Maps, Tables, and many more.

Data visualization tools are the applications that are used to visualize the data for many purposes such as to understand trends, to find patterns in data and many more. Good visualization tools are those which offer various visualization charts, are user friendly, can handle large data sets and data of different formats can also be visualized.

Tableau Desktop is the software being used to visualize and analyse the data in the dataset being used in this project. Tableau software was found in January, 2003 by Christian Chabot, Pat Hanrahan and Chris Stolte in Mountain view, California. Tableau is one of the best & most

powerful data visualization tools available in the market. Tableau applications range from representing simple data to visualizing complex business data sets. Tableau has many advantages some of them being able to create interactive visualizations quickly, ease of use, high performance but the main advantage of tableau is seen when it is used in business intelligence perspective. Tableau also has the functionality of mapping and uses longitude and latitudes to show custom geography.

DATA CLEANING & FILTERING

The dataset I am using for this project is collected from kaggle website. The dataset consists of information regarding suicides for 101 countries from the years 1985-2016.The dataset has the following fields country, year, sex, age, suicides_no, population, suicides/100k pop, country-year, HDI for year, gdp_for_year ($), gdp_per_capita ($), generation.

Dataset Link: https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016

The csv file that is being used in this project is master.csv.

I am selecting only 18 countries from the 101 countries present in the dataset to prevent the problem of over-crowding. These 18 countries are selected at random based on their geographical location in such a way that all the regions that are North, South, West and East are covered by the selected countries.

I am also excluding the year 2016 from the dataset as it has many value errors and many of the countries selected don't have proper data for the year 2016.

RESULTS

1. Visualizing the total suicides count from 1985-2015 in Figure 1 for the 18 selected countries to see which country has the highest suicide count and which country has the lowest suicide count

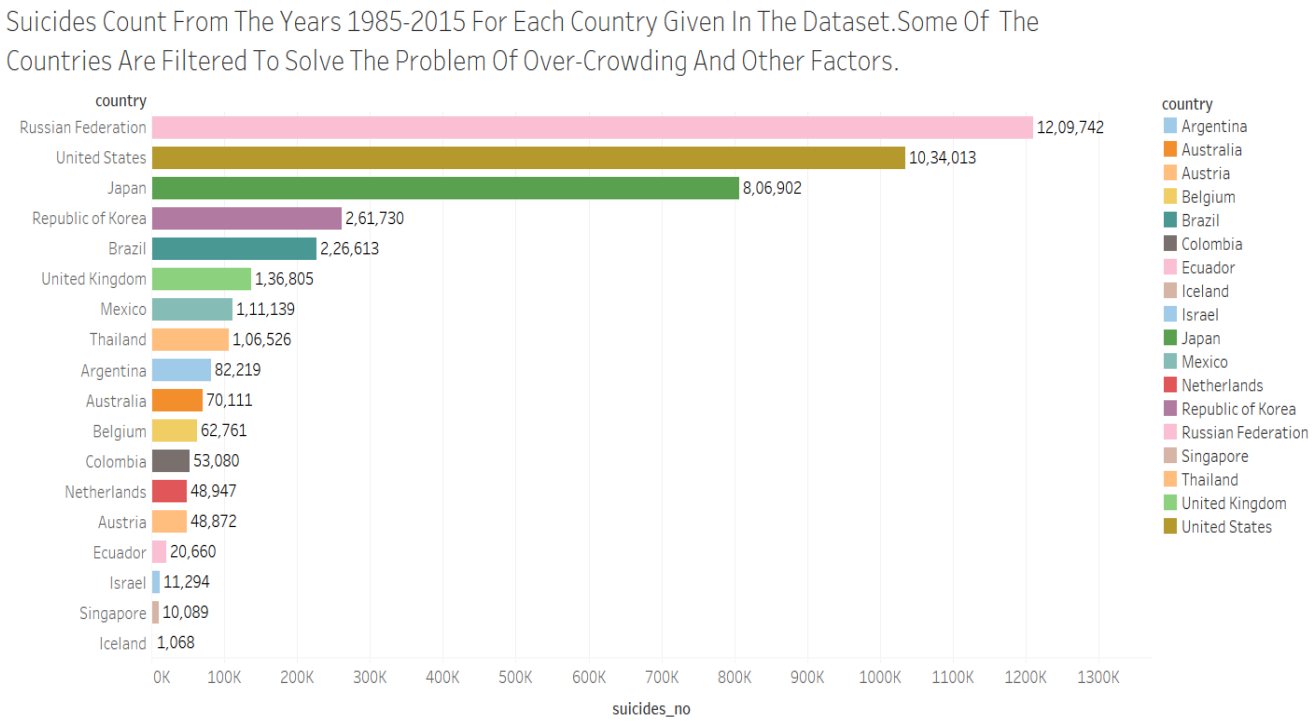

Figure 1 Total suicides count for 18 selected countries from the years 1985-2015

From Figure 1 it can be noticed that the Russia has the highest total suicides count that is 12, 09,742 suicides and Iceland has the lowest total suicides count that is 1,068 suicides.

2. Visualizing the Average Suicide rate for the 18 selected countries from the years 1985-2015 in Figure 2 and identifying the countries that have highest Average suicide rate and lowest Average suicide rate

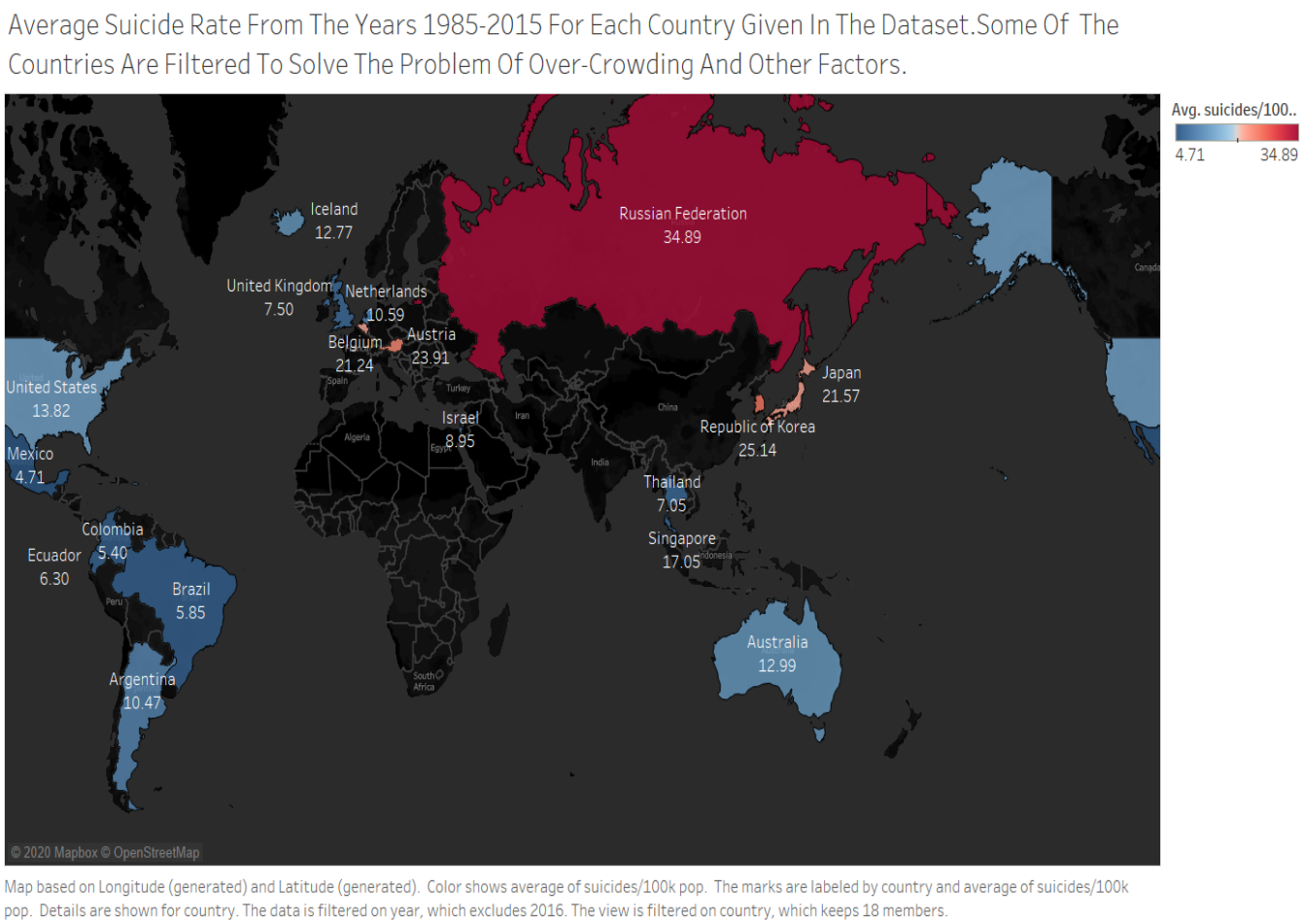

Figure 2 Average Suicide Rate for the 18 selected Countries from the years 1985-2015

From Figure 2 it can be noticed that Russian Federation has the Highest Average Suicide rate that is 34.89 suicides per 100,000 population and Mexico has the lowest Average Suicide rate that is 4.71 suicides per 100,000 population.

3. Visualizing the Average gdp_per_capita for the 18 selected countries from the years 1985-2015 in Figure 3 and identifying the countries that have highest Average gdp_per_capita and lowest Average gdp_per_capita

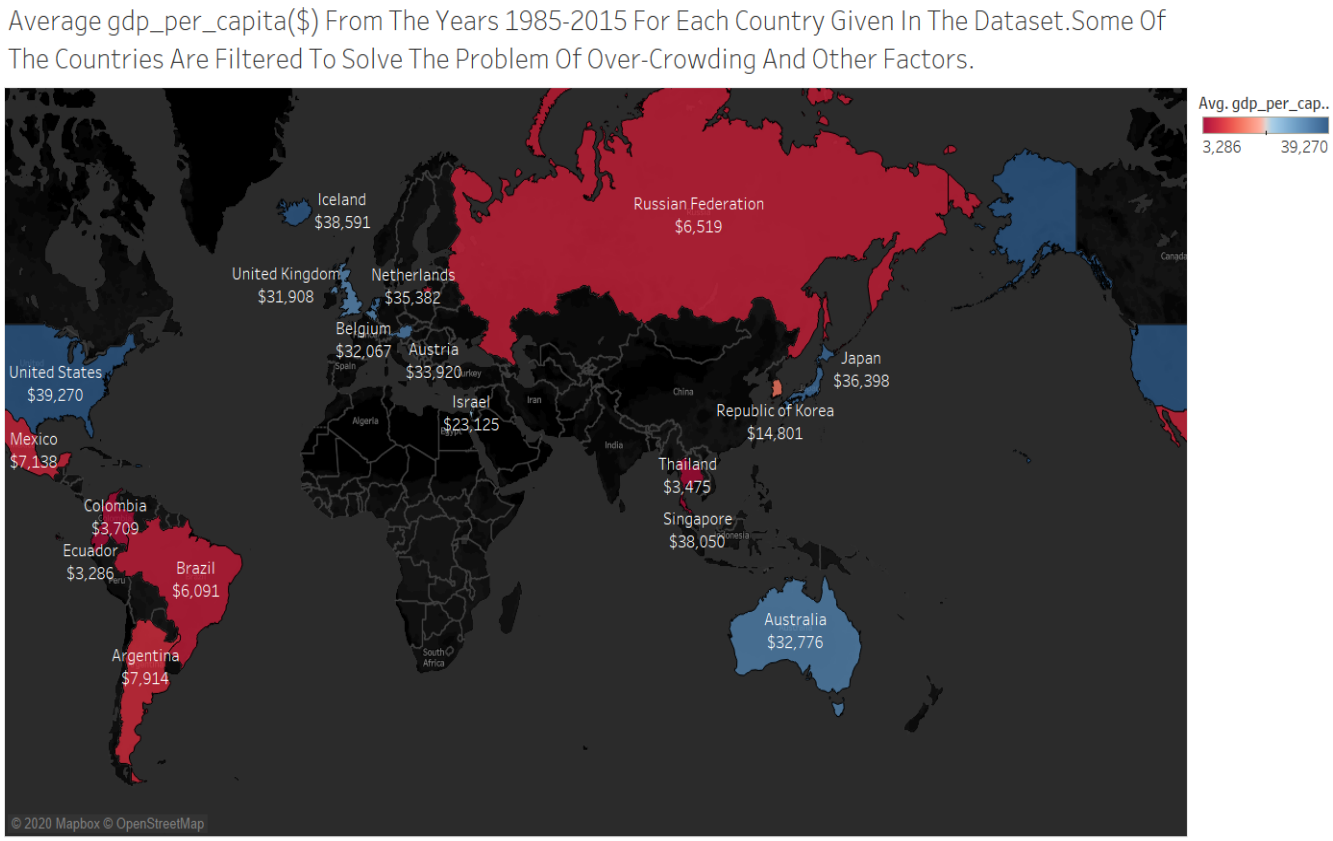

Figure 3 Average gdp_per_capita for the 18 selected countries from the years 1985-2015

From Figure 3 we can say that U.S.A has the highest average gdp_per_capita that is $39,270 and Ecuador has the lowest average gdp_per_capita that is $3,286.

I now want to see if there is any relation between average gdp_per_capita and average suicide rate as I see from Figure 2 and Figure 3 for some countries such as Russia, Republic Of Korea the average suicide rate/100,000 population is high and average gdp_per_capita is low and for U.S.A, Netherlands the Average suicide rate is low and Average gdp_per_capita is high.

4. I now want to test a hypothesis that is Countries having high average gdp_per_capita have low average suicide rate and Countries having low average gdp_per_capita have high average suicide rate. In Figure 4 I visualize the data regarding average gdp_per_capita and average suicide rate and in Figure 5 I analyse the R-squared and p-value by drawing a regression line and then decide whether to accept or reject null hypothesis.

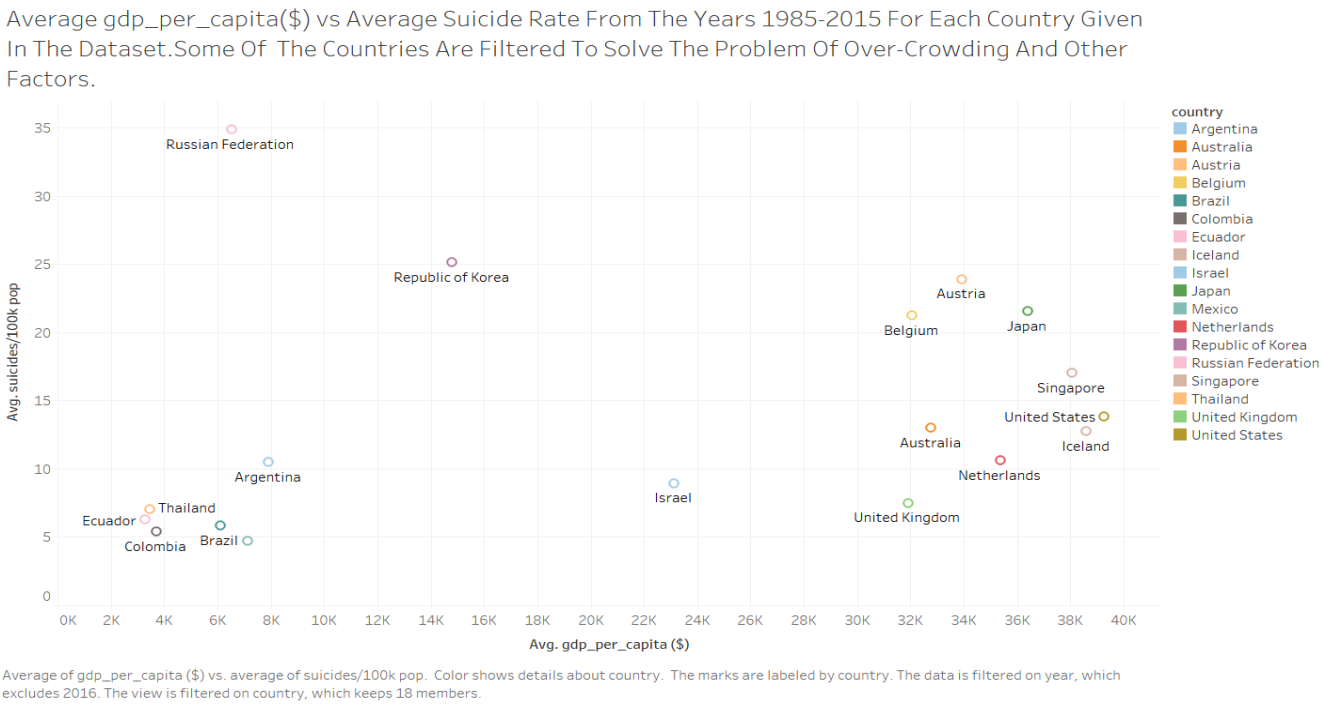

Figure 4 Average gdp_per_capita and Average suicide rate for the 18 selected countries from the years 1985-2015

By seeing Figure 4 we can say that there is no relation between average gdp_per_capita and average suicides/100k population but it would be better if we draw a regression line to see if there is any relation between average gdp_per_capita and average suicides/100k population based on the R-squared value and p-value we decide whether to reject or accept null hypothesis.

Null Hypothesis: There is no relationship between average gdp_per_capita and average suicide rate that is the slope of linear relationship between average gdp_per_capita and average suicide rate is 0 that is H_0: $B_1 = 0$ where B_1 is the slope.

Alternative hypothesis: There is relationship between average gdp_per_capita and average suicide rate that is the slope of linear relationship between average gdp_per_capita and average suicide rate is not equal to 0 that is H_a: $B_1 \neq 0$ where B_1 is the slope.

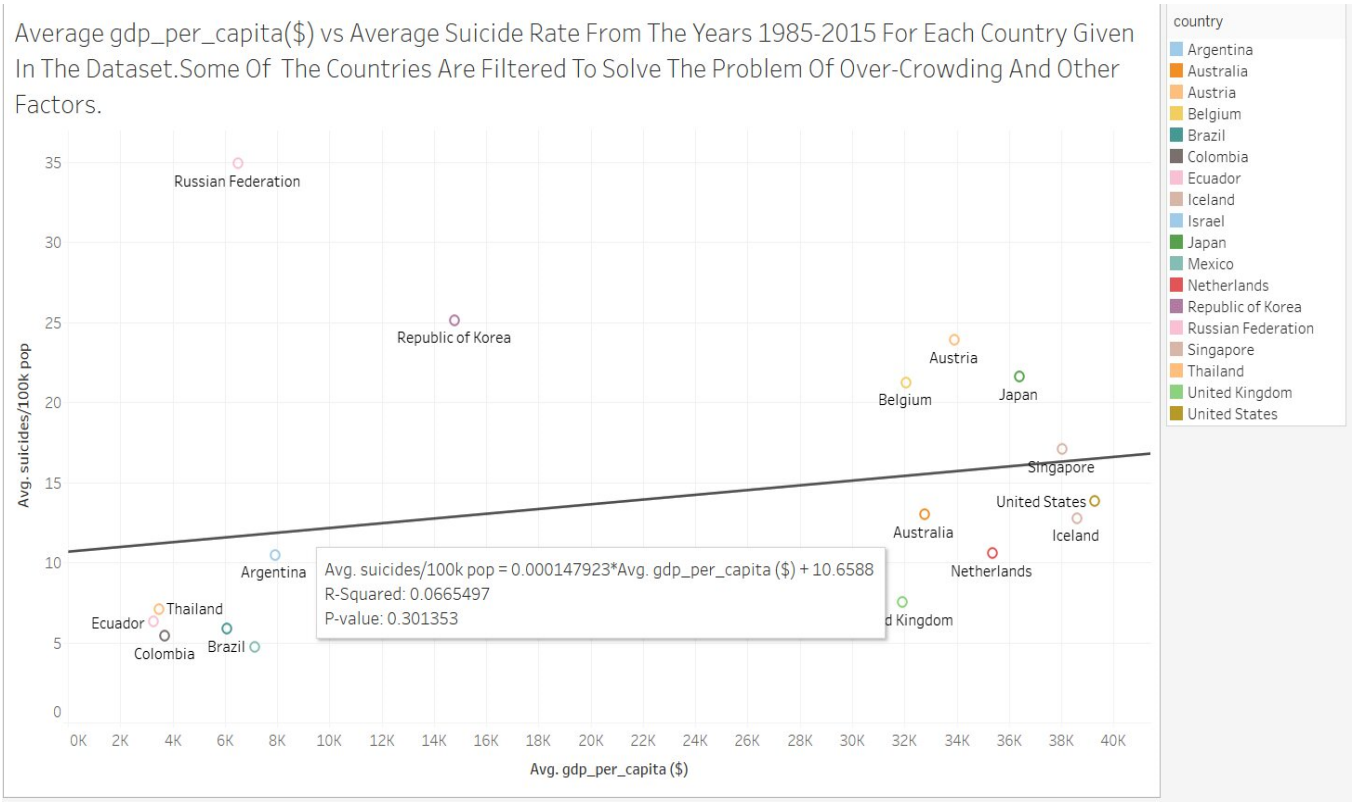

Figure 5 Testing the hypothesis for average gdp_per_capita and average suicide rate for 18 selected countries.

We accept the Null Hypothesis as we can see from Figure 5 the R-squared value is 0.00665497 that is 6.65497% which is very low that is the average gdp_per_capita/average suicide rate accounts for 6.65497% of the variation and coming to the probability value it is 0.301353 that is 30.1353% which is very high, the p value should be less than 5%.

From here on I want to see the trends of suicides, suicide rate in U.S.A from the years 1985-2015 based on gender, Age, gdp_per_capita

5. Visualizing the data for population of U.S.A for each year since the year 1985-2015 and also visualizing the total male population and total female population in the total population for each year since the year 1985-2015 in Figure 6

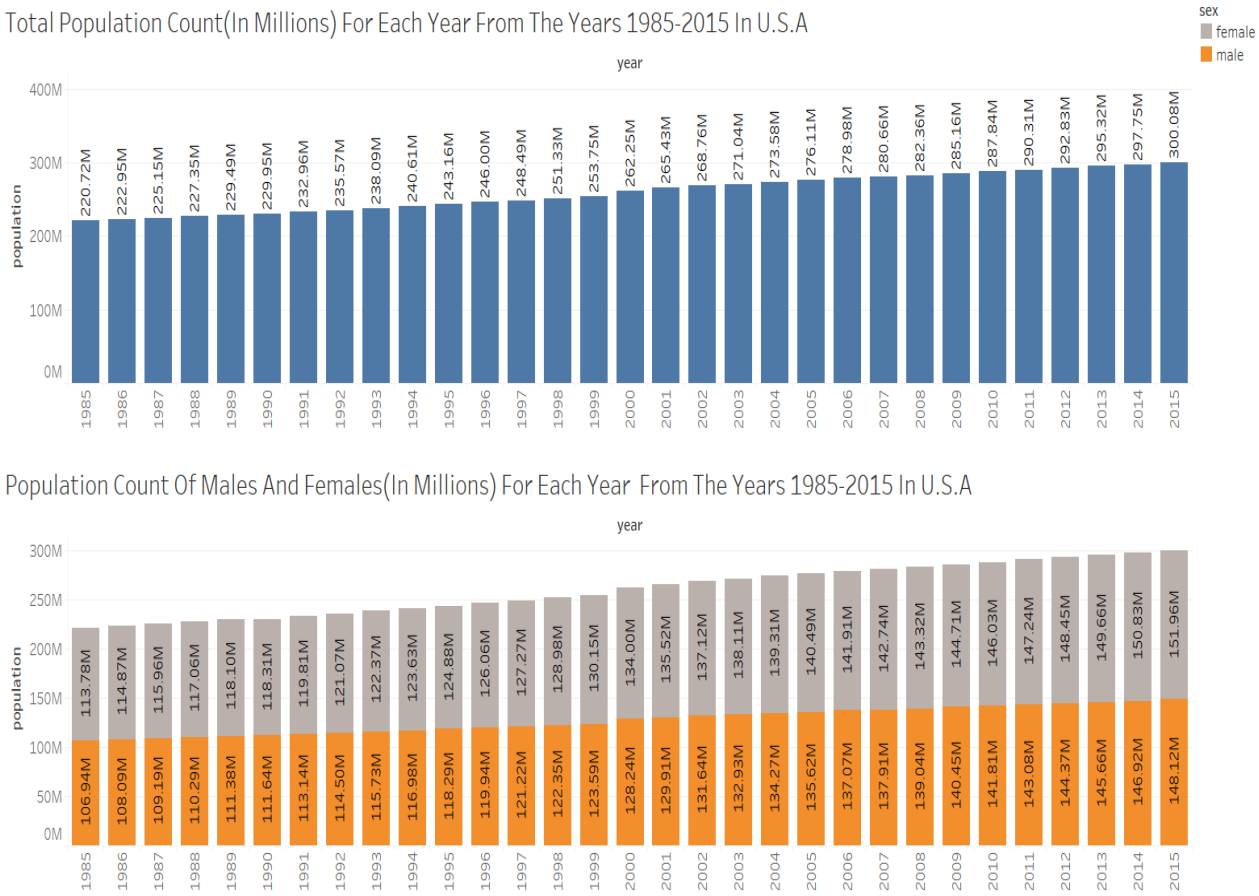

Figure 6 Total Population, Total Male and Female Population in the Total Population for Each year since year 1985-2015

From Figure 6 it can be inferred that the population in U.S.A kept on increasing from the years 1985-2015 and for each year the female population was greater than male population.

Many people say that the number of male suicides are more than female suicides due to various reasons such as males use more lethal means of committing suicide so the chance of survival after committing suicide are low, some say that females go and talk to psychiatrists and counsellors if they have suicidal thoughts whereas males don't go to psychiatrists and counsellors if they have any suicidal thoughts which results in more suicides for males.

6. I now want to see the total suicide count for each year from the year 1985-2015 and see the total suicide count of males and females, see if male suicides are more than female suicides in U.S.A for each year since years 1985-2015.

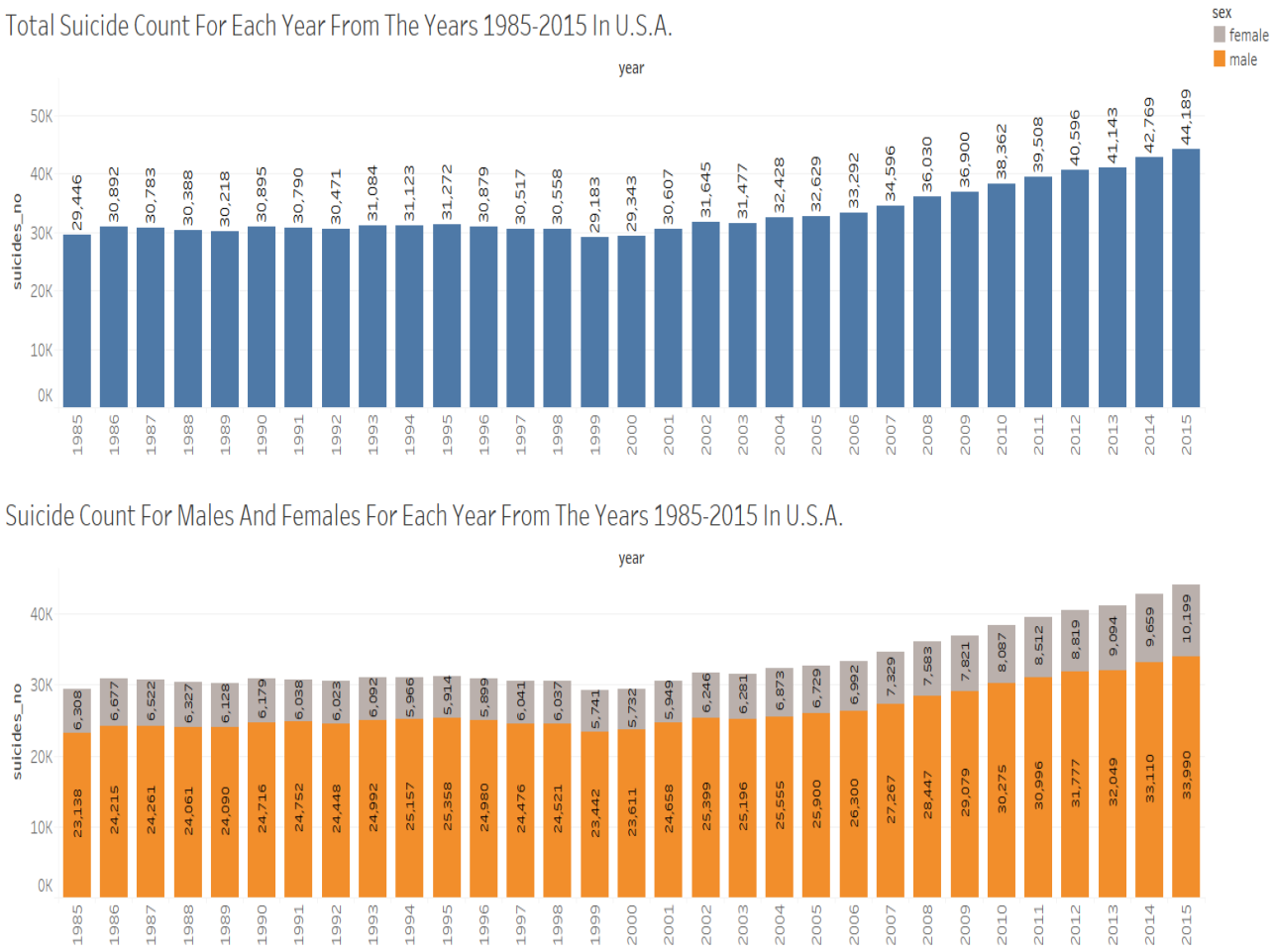

Figure 7 Total suicides for each year since 1985-2015 and total male suicides, total female suicides for each year since 1985-2015

As we can see from Figure 7 for every year since 1985-2015 the total suicides count of males is greater than that of females, the number of suicides have been increasing since the year 2003 to 2015. From Figure 6 and Figure 7 it can also be inferred that even though majority of the population are females for every year since 1985-2015 the total suicide count of males is greater than that of females.

7. What is the total suicides count and what is the total suicide count of males and females till 2015 since 1985 In U.S.A?

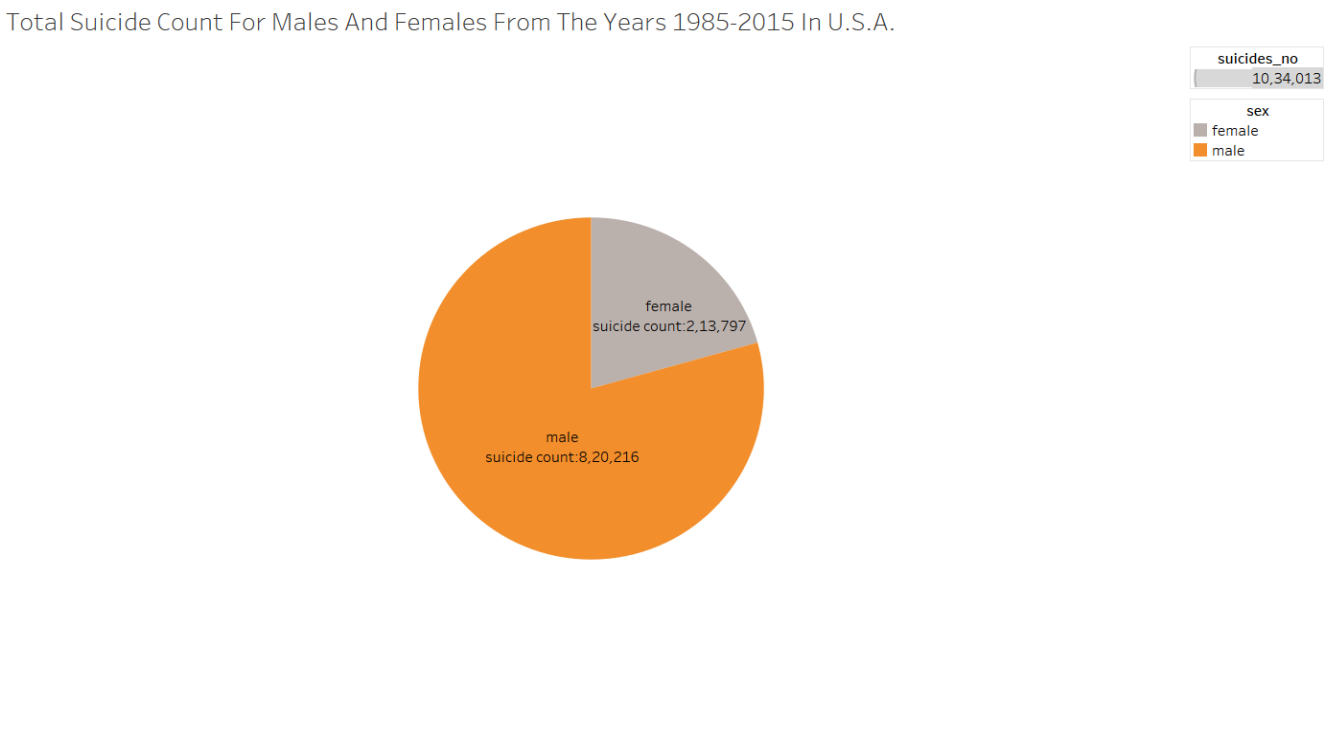

Figure 8 Total Suicide Count of Males and Females from the years since 1985-2015

From Figure 8 we can say that The total suicides since year 1985-2015 are 10, 34,013 in which total suicide count of males is 8, 20,216 and total suicide count of females is 2,13,797.

8. I now want to test the hypothesis that is male suicide rate is higher than female suicide rate in U.S.A for each year in U.S.A since years 1985-2015. To test this hypothesis we visualize the suicide rate for each year since 1985-2015 in Figure 9, analyse it and then decide whether to reject or accept the null hypothesis.

Null Hypothesis: There is no difference between male suicide rates and female suicide rates

Alternative Hypothesis: The suicide rates for males is greater than females.

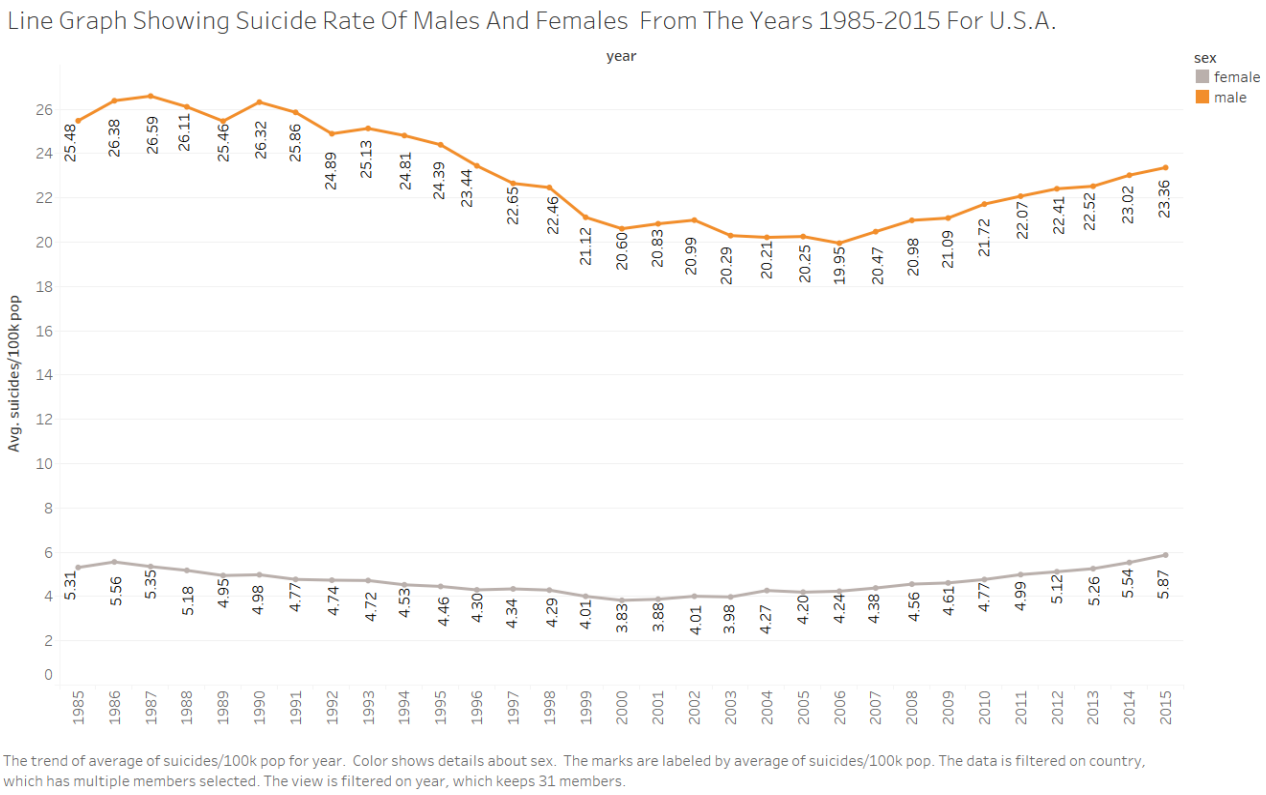

Figure 9 Male and Female suicide rates for each year from 1985-2015

We reject the Null Hypothesis as we can see From Figure 9 that for every year the suicide rate for males is higher when compared to females, there were fluctuations in male suicide rate from the year 1985 to 2005 and from the years 2006-2015 the male suicide rate kept on increasing for every year compared to the previous year. For females the suicide rate kept on fluctuating from the years 1985 to 2004 and the suicide rate increased from 2005 to 2015.

9. Visualizing the total suicide count based on age since 1985 to 2015 and also visualizing the total suicide count of males and females by age in Figure 10.

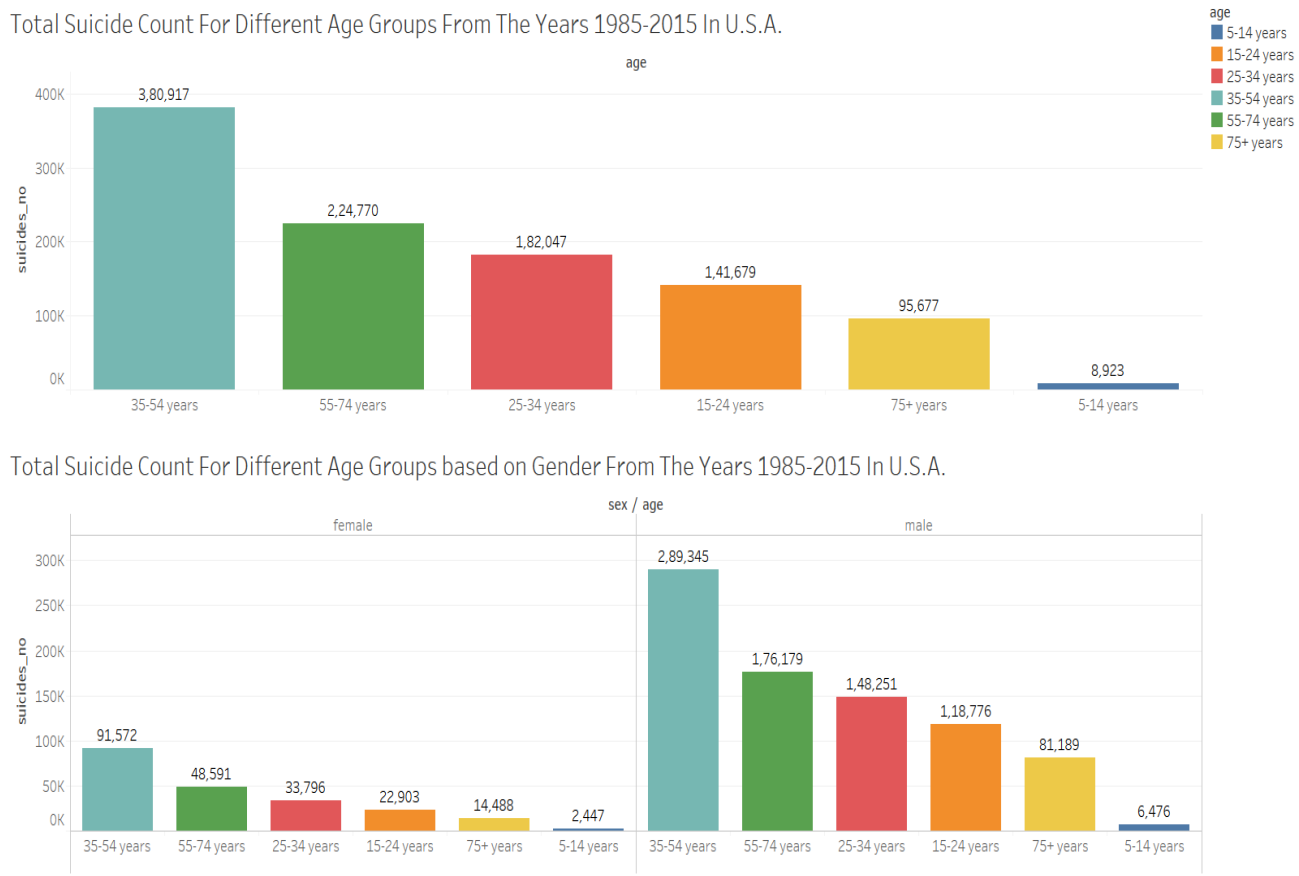

Figure 10 Total suicide count for different age groups, Total suicide count of males, Females for different age groups from the years 1985-2015

From Figure 10 we can say that the total suicide count of Age Group 35-54 is the highest and the suicide count of Age Group 5-14 is the lowest. When the total suicide count of Males and Females are compared for different age groups it can be inferred that the Total suicide count for males of age group 35-54 is the highest that is 2,89,345 suicides from the year 1985 to 2015 and Total suicide count for females of age group 5-14 is lowest that is 2,447 suicides.

According to (Durkheim & Suicide, 1952) it was stated in his book that the suicide rate for people who are old is the highest as socializing decreases and socializing has an impact on suicide rate.

10. I want to see if old people have higher suicide rate when compared to other age groups. I assume that people whose age is greater than 75 are old.

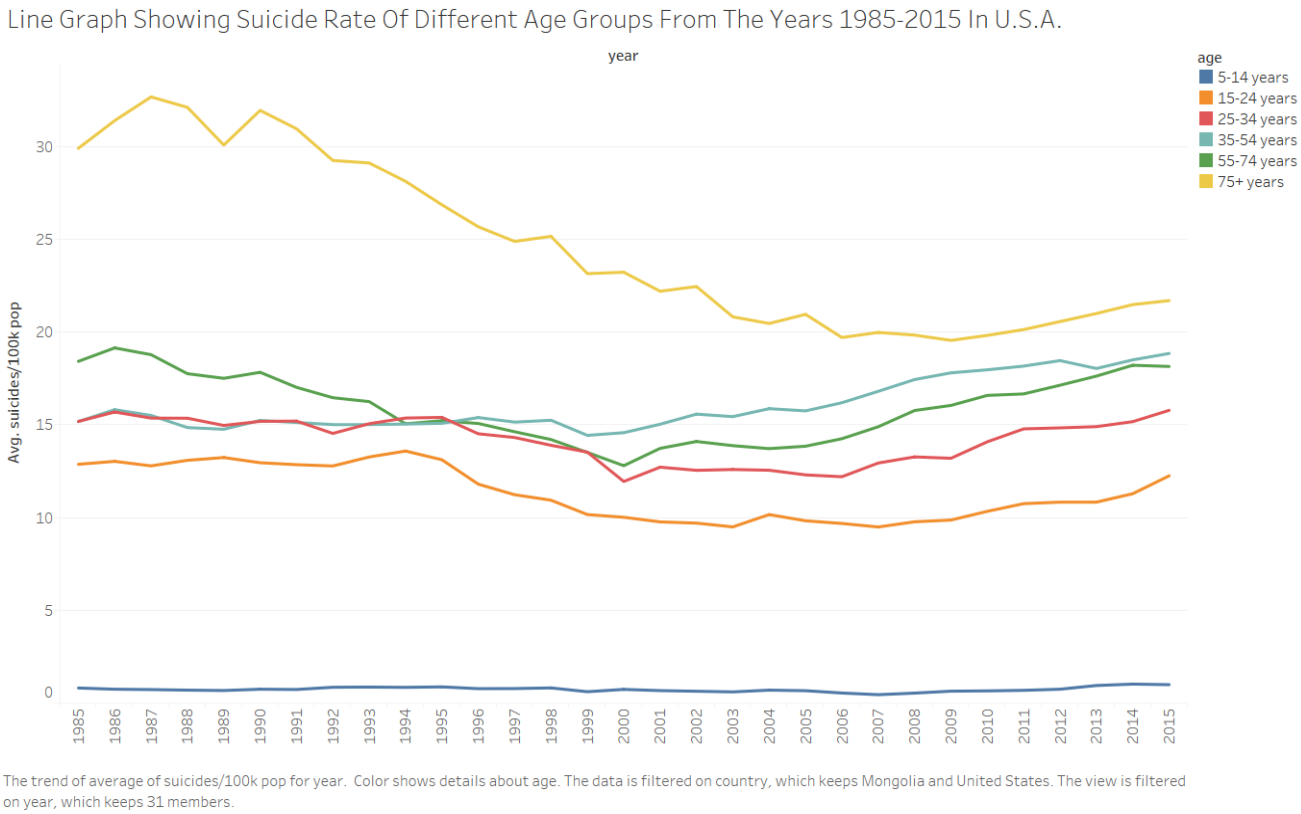

Figure 11 Suicide rate for different age groups for each year from the years 1985-2015

As we can see from Figure 11 the suicide rate for people whose age is above 75 is the highest for every year when compared with other groups.

11. I want to see how the gdp_per_capita and suicide rate are for each year from the years 1985-2015 and is there any relationship between gdp_per_capita and suicide rate.

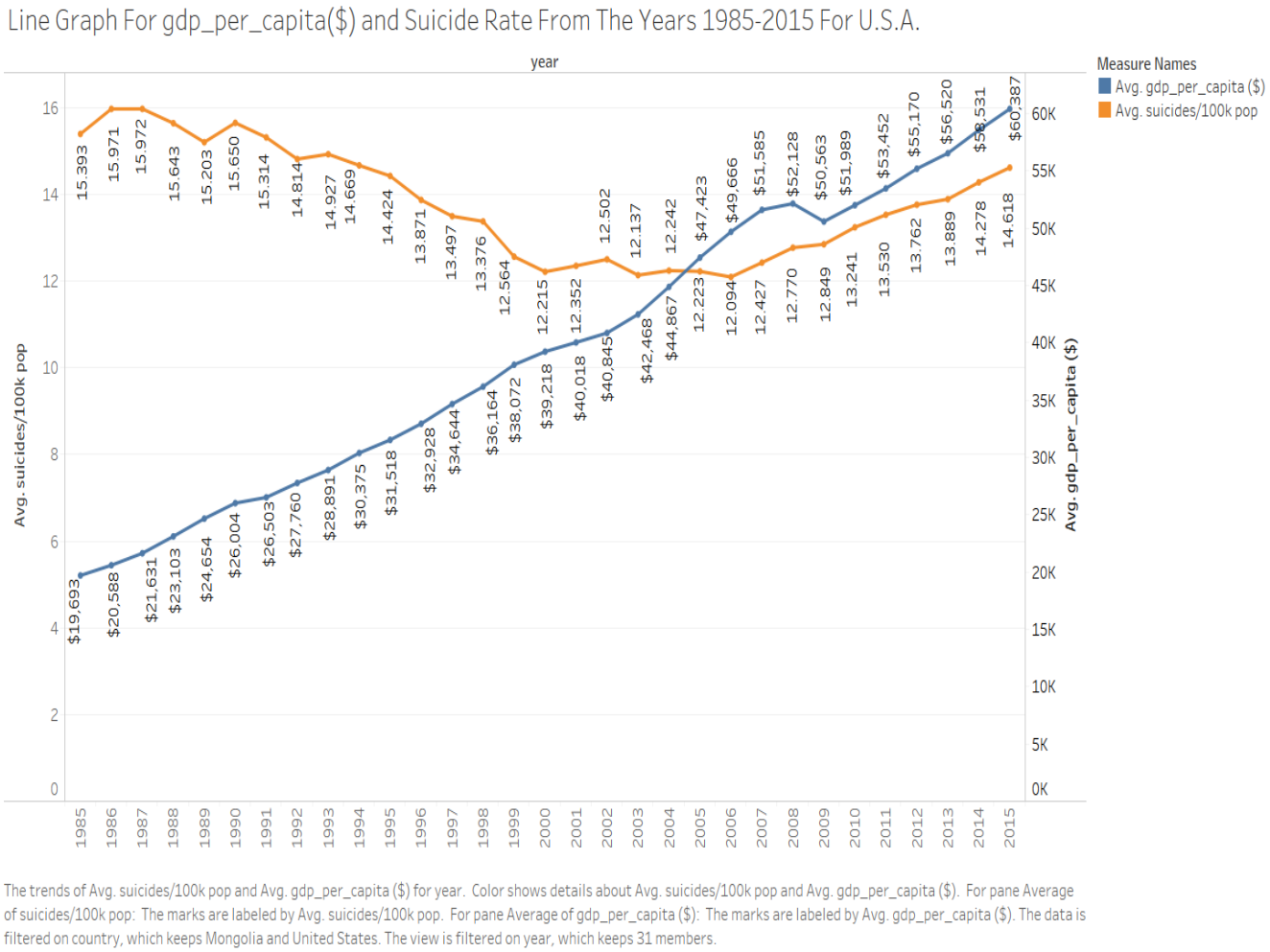

Figure 12 Line graph for gdp_per_capita and suicide rate for each year from the years 1985-2015

By analysing Figure 12 Line graph for gdp_per_capita and suicide rate for each year from the years 1985-2015 we can say that gdp_per_capita and suicide rate have no relation as we can see for year 1986 even though the gdp_per_capita increased when compared to the previous year the suicide rate increased and from the years 2009-2015 even though the gdp_per_capita increased kept on increasing for each year the suicide rate kept on increasing but it would be better if we draw a regression line to see if there is any relation between gdp_per_capita and suicide rate, based on the R-squared value and p-value we decide whether to reject or accept null hypothesis.

12. I want to test a hypothesis that is as gdp_per_capita increases the suicide rate decreases and as the gdp_per_capita decreases the suicide rate increases. To test this hypothesis we visualize the suicide rate and gdp_per_capita for every year from 1985-2015 in Figure 13, analyse Figure 13 and then decide whether to accept or reject the hypothesis by drawing a regression line and analysing the R-square and p values.

Null Hypothesis: There is no relation between suicide rate and gdp_per_capita that is the slope of linear relationship between gdp_per_capita and suicide rate is 0 that is H_0: $B_1 = 0$ where B_1 is the slope.

Alternative Hypothesis: There is relation between suicide rate and alternative hypothesis rate that is the slope of linear relationship between average gdp_per_capita and average suicide rate is not equal to 0 that is H_a: $B_1 \neq 0$ where B_1 is the slope.

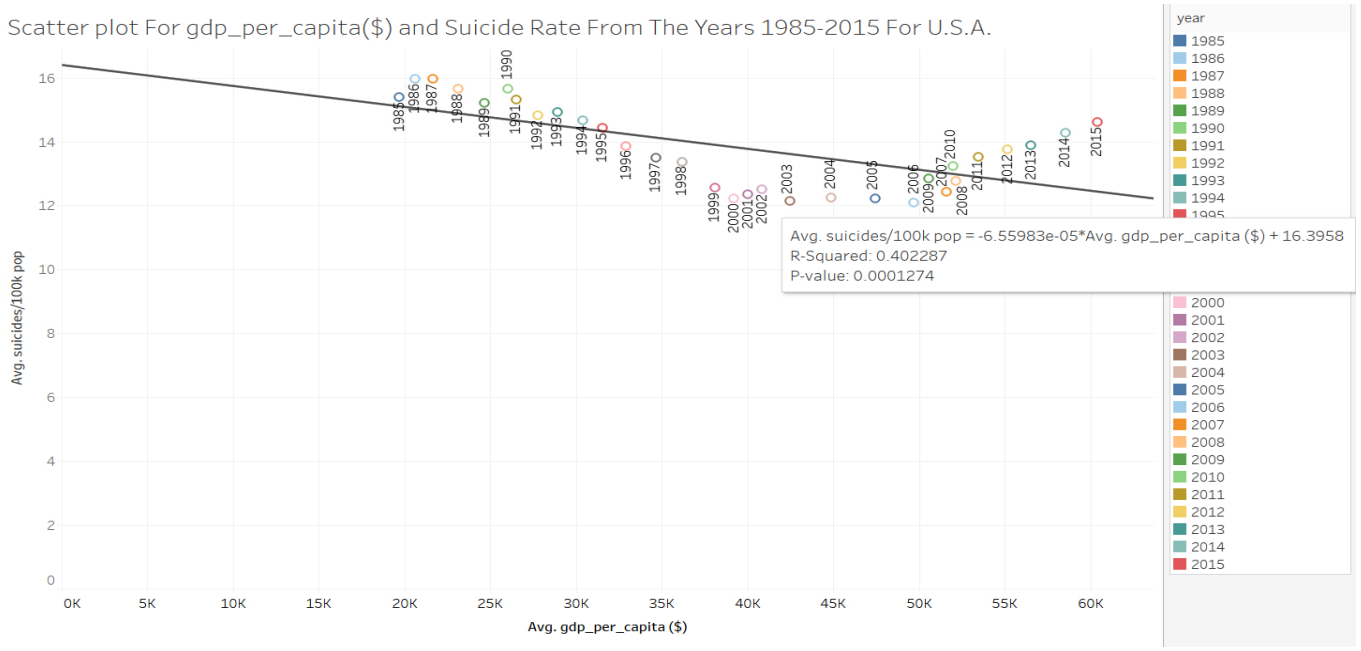

Figure 13 Scatter plot of gdp_per_capita and suicide rate for every year from the years 1985-2015

We accept the Null Hypothesis as we can see from Figure 13 that the R-square value is 0.402287 that is 40.2287% which is low that is the gdp_per_capita/suicide rate account for 40.2287% of the variation and coming to the probability value it is 0.0001274 but due to the value of R-squared we accept the Null Hypothesis.

DISCUSSION

I found out that average gdp_per_capita and average suicide rate have no relation for the 18 selected countries. I analysed the total suicides count of males and females in U.S.A and found out that the total suicides count of males is greater than that of females for every year from the years 1985-2015 even though majority of the population were females for every year. I tested the hypothesis to see if male suicide rate are higher than female suicide rates In U.S.A from the years 1985-2015 and found out that male suicide rates are greater than female suicide rates and I analysed the total suicides count for different age groups from the years 1985-2015 in U.S.A and found that age group 35-54 have the highest suicide count and I analysed the total suicides count for different age groups based on gender and found out that Total suicide count for males of age group 35-54 is the highest suicides and Total suicide count for females of age group 5-14 is lowest from the year 1985 to 2015. I analysed the suicide rates for different age groups and found out that the suicide rate for people whose age is greater than 75 have the highest suicide rate and people whose age is between 5-14 have the lowest suicide rate from the year 1985-2015 in U.S.A. I have also analysed if there is any relation between gdp_per_capita and suicide rate in U.S.A from years 1985-2015 and found out that there is no relation between gdp_per_capita and suicide rate.

CONCLUSION

Average gdp_per_capita and average suicide rate have no relation for the selected 18 countries from the years 1985-2015, there may be reasons other than gdp_per_capita for increase or decrease in suicides count. The suicide count of males is greater than females for every year since 1985-2015 in U.S.A and the suicide count of age group 35-54 is the highest and 5-14 is the lowest, suicide rate for people whose age is greater than 75 is the highest in U.S.A from years 1985-2015, gdp_per_capita and suicide rate have no relation with each other in U.S.A & the suicide rate for U.S.A has been continuously increasing since 2006-2015.

Literature Cited

Agerbo, E., Sterne, J., & Gunnell, D. (2007). Combining individual and ecological data to determine compositional and contextual socio-economic risk factors for suicide. *Social science & medicine, 64*(2), 451-461.

Bondy, B., Buettner, A., & Zill, P. (2006). Genetics of suicide. *Molecular psychiatry, 11*(4), 336-351.

Durkheim, E., & Suicide, A. (1952). *A study in sociology.* London: Routledge & K. Paul.

Girard, C. (1993). Age, gender, and suicide: A cross-national analysis. *American sociological review*, 553-574.

Hedegaard, H., Curtin, S., & Warner, M. (2018). Suicide mortality in the United States.

Mościck, E. (n.d.). Gender differences in completed and attempted suicides. *Annals of epidemiology, 4*(2), 152-158.

Van Hal, G. (2015). The true cost of the economic crisis on psychological well-being: a review. *Psychology research and behavior management, 8*(17).

Weir, K. (2019). Worrying Trends in US Suicide Rates. *Monitor on Psychology, 50*(3), 24-39.

World Health Organization. (n.d.). *Preventing suicide: A global imperative.* World Health Organization.

Ying, Y., & Chang, K. (n.d.). A study of suicide and socioeconomic factors. *Suicide and life-threatening behavior, 39*(2), 214-226.