

Final Project Report

Suicide Rates Overview in period 1985 - 2016

Ha Hai Vu

1. Introduction

1.1 Data Description

For this project, I used the Suicide Rates Overview 1985 to 2016 dataset from Kaggle. This dataset compares socio-economic information with suicide rates by year and country. The dataset was obtained from Kaggle. This compiled dataset was pulled from four other datasets linked by time and place and was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum. This dataset contains 27820 observations and 12 variables, 6 variables categorical/ordinal and 6 variables are continuous. However, I focused on the following variables for our analysis:

1. Country: 101 countries
2. Year: 1985-2016
3. Sex: Male, Female
4. Age: Categorized into 6 groups, beginning at age 5
5. Suicide Rate per 100k Population
6. GDP per Capita
7. Generation: G.I, Silent, Boomer, Gen X, Millennials, Gen Z
8. Economic Growth

Source: <https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016>

The age group are assigned to the original data that is 5-14, 15-24, 25-34, 35-54, 55-74, 75+. These groups are classified as life stages.

The generation groups are divided into six groups related to year born. G.I Generation is from 1901 – 1927; Silent Generation is from 1928 – 1945; Boomer Generation is from 1946 – 1964; X Generation is from 1965 – 1980; Millennials Generation is from 1981 – 1996; and Z Generation is from 1997 – 2012.

1.2 Purpose of Analysis

The goal of this project is to focus on the relationship between the suicide rates and the variables that factor into larger/smaller quantities. Since suicide is a very sensitive topic and an issue that has plagued society, visualizing general trends can lead to answers of suicide prevention.

I am going to answer the following research questions: (1) Is there a relationship between country and suicide rate? Which area(s) has the highest or lowest suicide rates? (2) Does the suicide rate depend on age for each gender? (3) How does the rate of suicide change depend on the country's economic status? What is the relationship between suicide rate and GDP per capita given age and sex? (4) For each generation, what is the general trend of suicide rate?

2. Exploratory Analysis

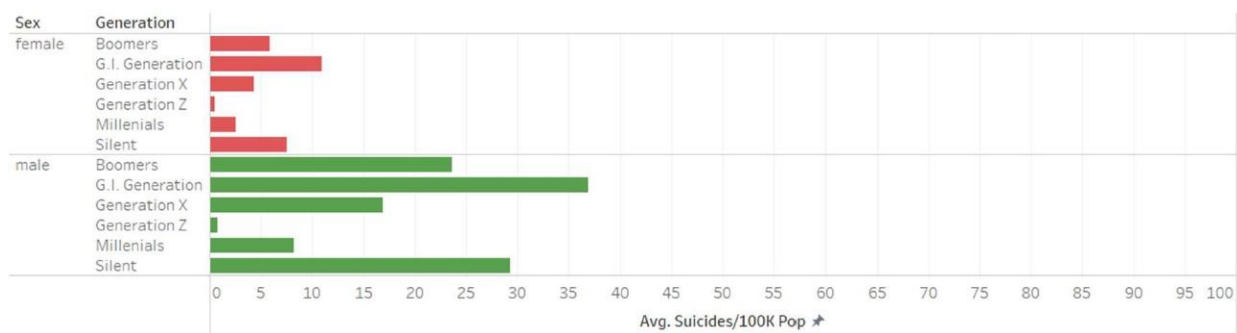
2.1 Treemap of the Countries

During our exploratory analysis, the following visual was created which looks simply at the countries with the highest suicide rate. This visual can help us focus on those countries with the highest rates to see what factors are contributing to the higher suicides.



2.2 Bar Graph of Generation

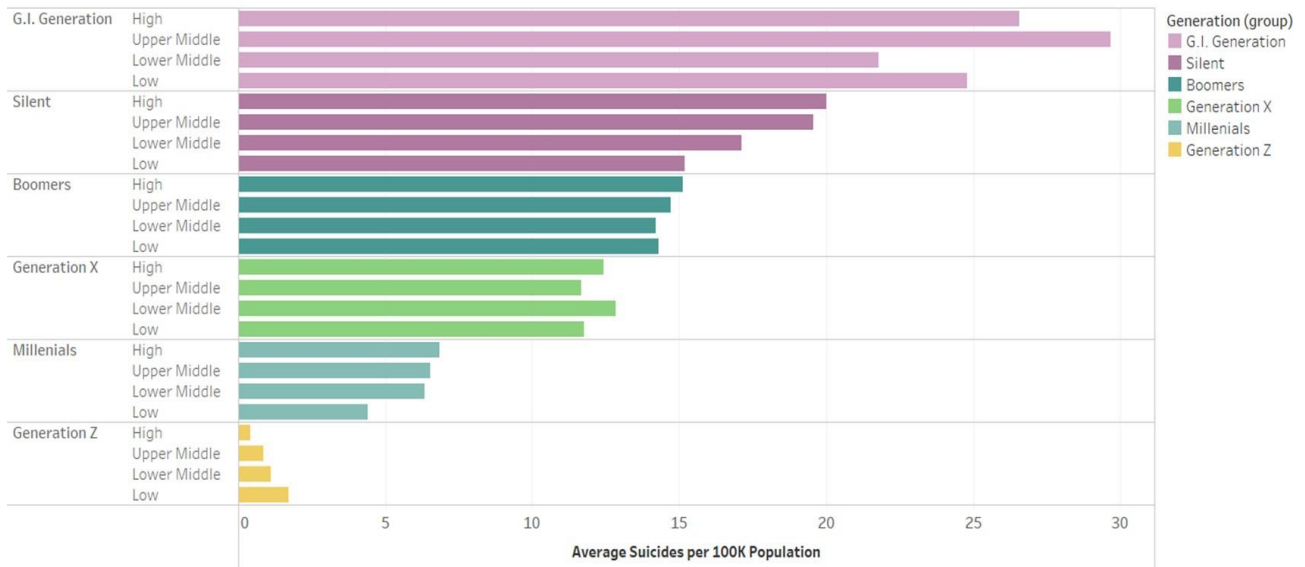
The following exploratory visual shows the sex differences between each generation in the dataset.



2.3 Bar Graph of Income Level

The following grouped bar chart displays aggregated suicide rates of the generational groups divided into subgroups based on the 2014 World Bank Income Classifications, including low-income, lower and upper middle income, and high income.

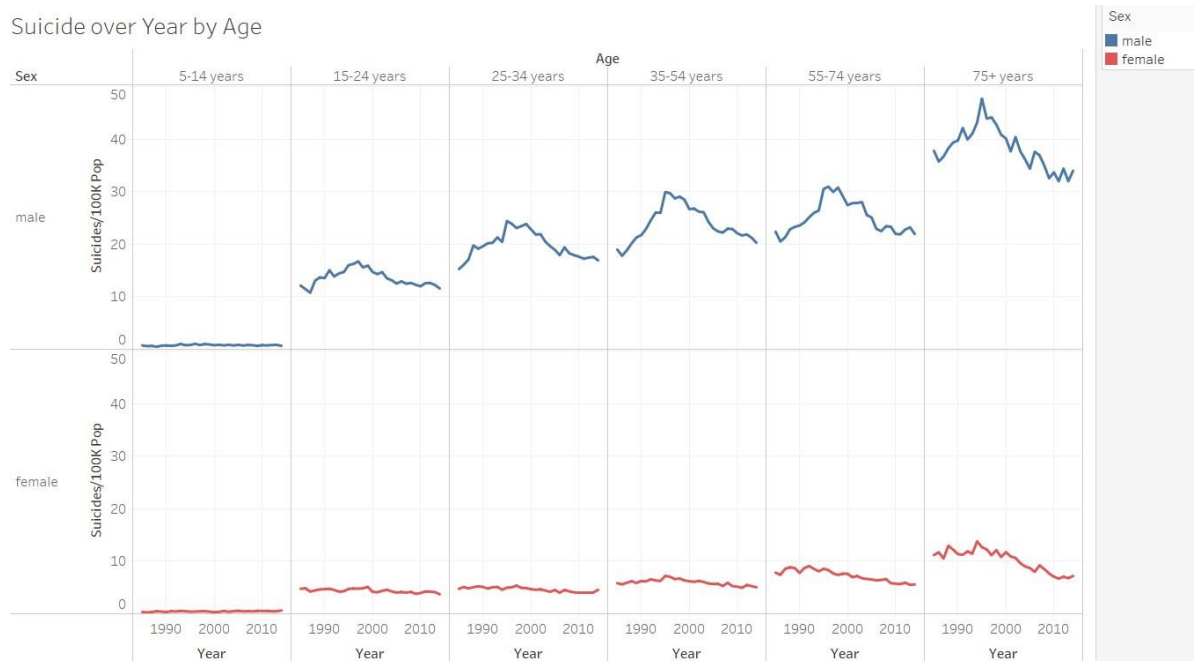
Income



2.4 Line Graph of Age Group

This visual is created the average of suicide per 100K in population by age. There are 6 age groups and divided by sex. The blue lines represent the male age group, and the red lines show the female age group. Each age group tells significant differences suicide rate so that it is a good one to make narrow to focus on.

Suicide over Year by Age



3. Explanatory Analysis Visualization

3.1 Choropleth Map of Each Country

Global Suicide Rates per 100,000 Population

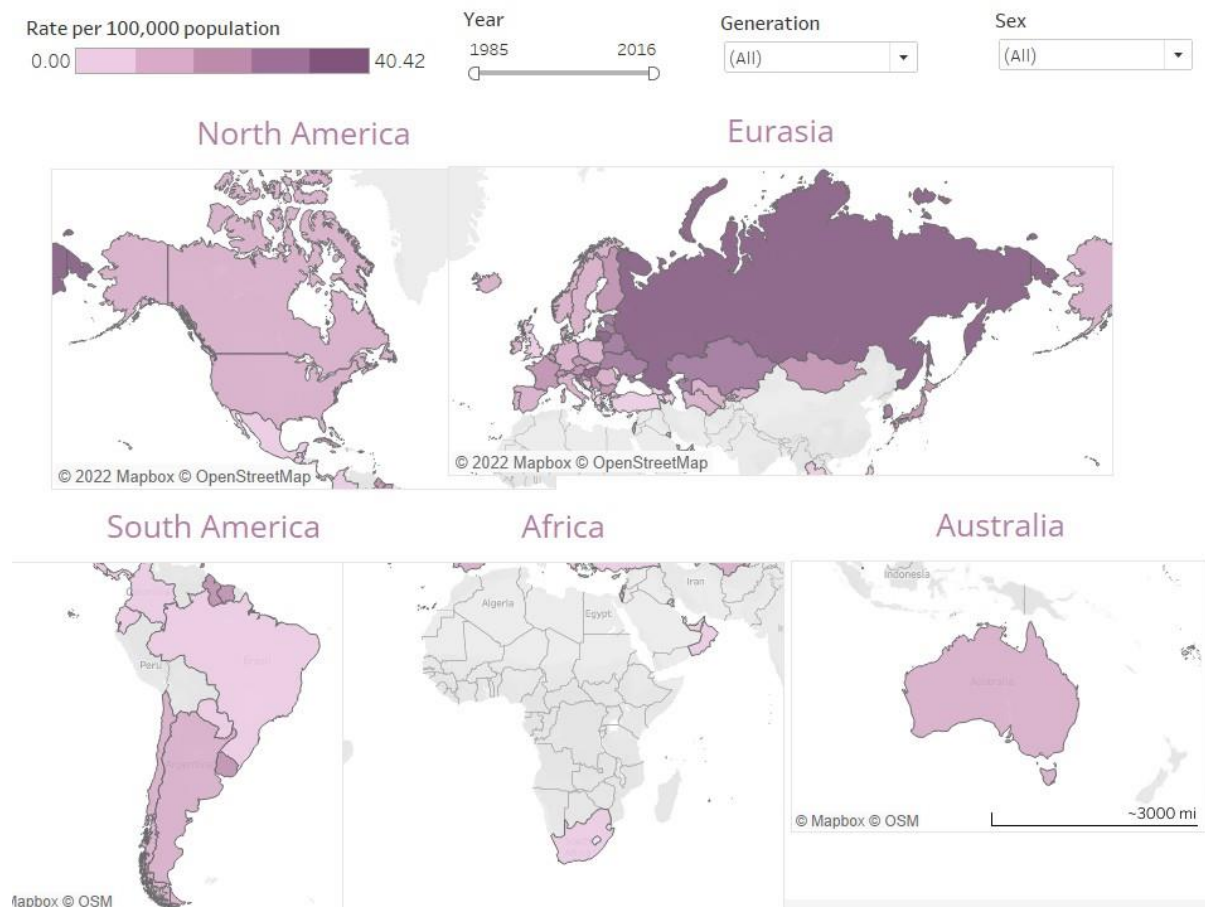


Figure 1. Choropleth Map of Global Suicide Rates per 100,000 Population for the years 1985 to 2016.

The choropleth map (Figure 1) shows the suicide rate for each country around the world where data is available in this dataset. Each country is filled with color and the intensity of the color represents the suicide rate. The map includes a color legend to explain the color intensity, the higher the intensity the higher the suicide rate is for that country. The suicide rate for each country was calculated by taking the average suicide rate of each country between the year 1985 and 2016.

The first iteration of the map in Figure 1. was created as one visual but later was created using 5 separate maps combined into one dashboard to eliminate space taken up by the oceans. This map informs the audience where suicide rate information is available around the world. For the countries with suicide information available, I can see quickly which countries have higher suicide rates and shows geographical groupings and where there are stark differences between two neighboring countries.

Figure 1. demonstrate the data availability around the globe. For African, I only have data for South Africa and not data is available for any of the other countries in the continent. I see a similar trend in South Asia, Southeast Asia, and the Middle East. I can also compare average suicide rates between countries and see patterns and differences between neighboring countries. I can also see that suicide rates are similar for North America and Australia. I would need to further divide these countries by state, counties, provinces, or territories to get more granular information for those countries.

3.2 Waffle Charts and Bar Chart of Sex and Age Group



Figure 2. Waffle and Stacked Bar Chart of Suicide Rates and Sex by Age group for the years 2000 to 2014

For this dataset, I used an Excel pivot table to calculate how many countries were included in each year. I found that the period from 2000 to 2014 had the most countries taking part in the research and could provide a good time frame to analyze the data.

In Tableau, I created six 10 x10 waffle charts to show the suicide rate ratio by sex in different charts by age group from 2000 to 2014. Within each chart, I have a hundred colored circles to indicate the represented percentage and sex in response to its relevant suicide rate ratio by sex. Each circle represents 1%. The two-color choices are understandable to indicate male (blue) and female (pink). To contrast the difference, the two colors are on the opposite side of the color wheel. To iterate the percentage per sex more clearly, I marked annotations of the ratio in the center of each chart. The waffle charts use the size of the colored area for comparison. Color is not the best property for comparison, but it could give us a general idea of ratio. If the audience is looking for a specific percentage, the annotations would be a good help. Besides, to have a deep insight about the effect of gender in suicide rate, I also created a stacked bar graph of average suicide rates and sex by age group in the research period. Using this graph, audiences can see how many there are in each one of the age groups.

Figure 2 helps audiences compare the suicide rate between males and females, and how the age group may affect this rate. From two graphs I can see clearly that female suicide rate is much lower than that of males in every age group. And the over 75 age group has the highest suicide rates. To interpret the rationale for the age groups, these are likely divided by life stages and its corresponding stress levels. For ages 5-14, the male ratio is about 60%, approximately even with the female ratio. The stress of ages 5 to 14 is probably from family and school. For ages 15-24, the ratio split is getting more obvious, with the ratio in the male group being nearly 80%. Males seem to be more stressed in school and when they find jobs. The peak of the male ratio of (81.45%) is from ages 25 to 34. The stress is understandably high in this age group because they are starting to build their new family and develop their careers. For ages 35-54, this group could be more stable in terms of not that many new changes as many as the previous group, but the ratio is the same likes ages 25 to 34. For ages over 75, the suicide ratio seemed to skew to females much more, it is almost half of the males. But it could be just because women live longer than men at the age above 75.

3.3 Stacked Bar Graph of Gender and Age Group

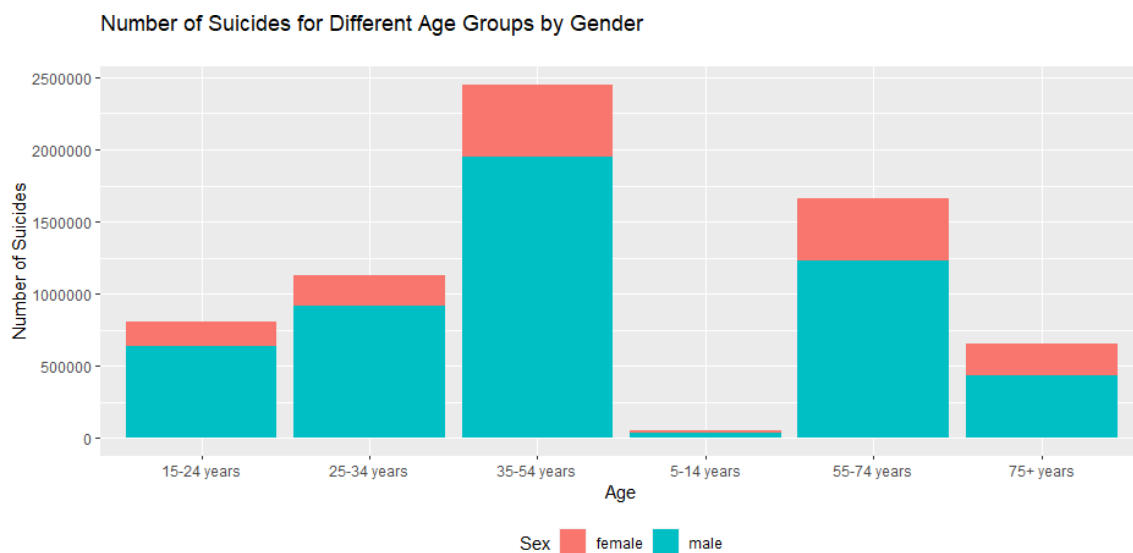


Figure 3. Stacked Bar Graph for Gender and Age group for number of suicides.

In the stacked bar graph, the rate (suicide numbers) is shown corresponding to the sex and age of the data. The graph has two genders (male and female) with 6 different age groups.

In Figure 3, I can see that the greatest number of suicides for both men and women are for the age-group 35-54 years. Also, men in that age group has the highest suicide rates overall. There should be more analysis of the factors that are causing that age-group and in particular men marking that mid-age people face a lot of problems and cannot cope up in general. Another noticeable fact is that men tend to have more suicide numbers than women.

3.4 Dot Graph of Economy Growth

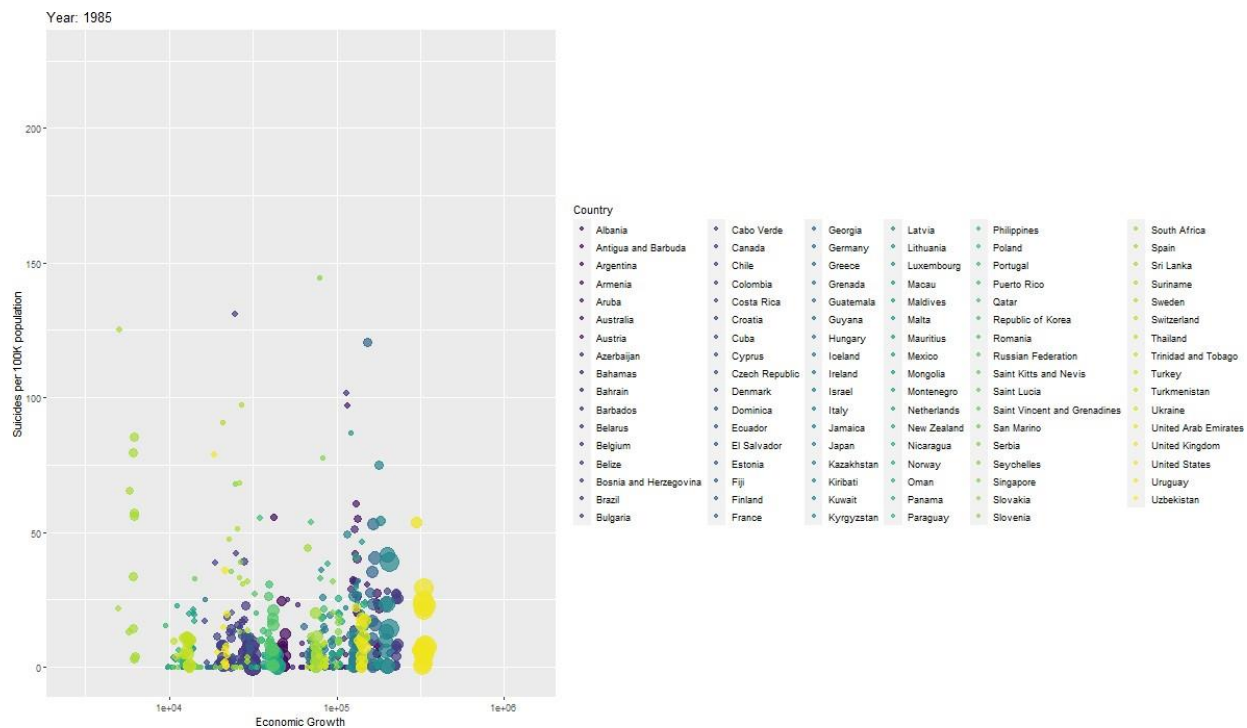


Figure 4. Change in Suicides per 100k Population as Economy Grow over the Years.

In the above animation, y-axis shows the Suicides per 100k Population, while x-axis shows the Economic growth (transformed variable – population * gdp per capita) and the change through the years: 1985 - 2016. The size of the point represents the population, and the color represents the country.

Figure 4 helps us to understand the relation between the economic growth of the country and the rate of suicides per 100k population. Also, by looking at the color I can determine the country and the size helps us determine the population for that country. In the animation, I can see that as the economy grows, suicide rates decrease over time. Although I can see that some of the countries with less population in comparison to others, despite the economic growth have high suicide rates suggesting that there are some other factors responsible in certain cases.

3.5 Mosaic Plots of Age Group

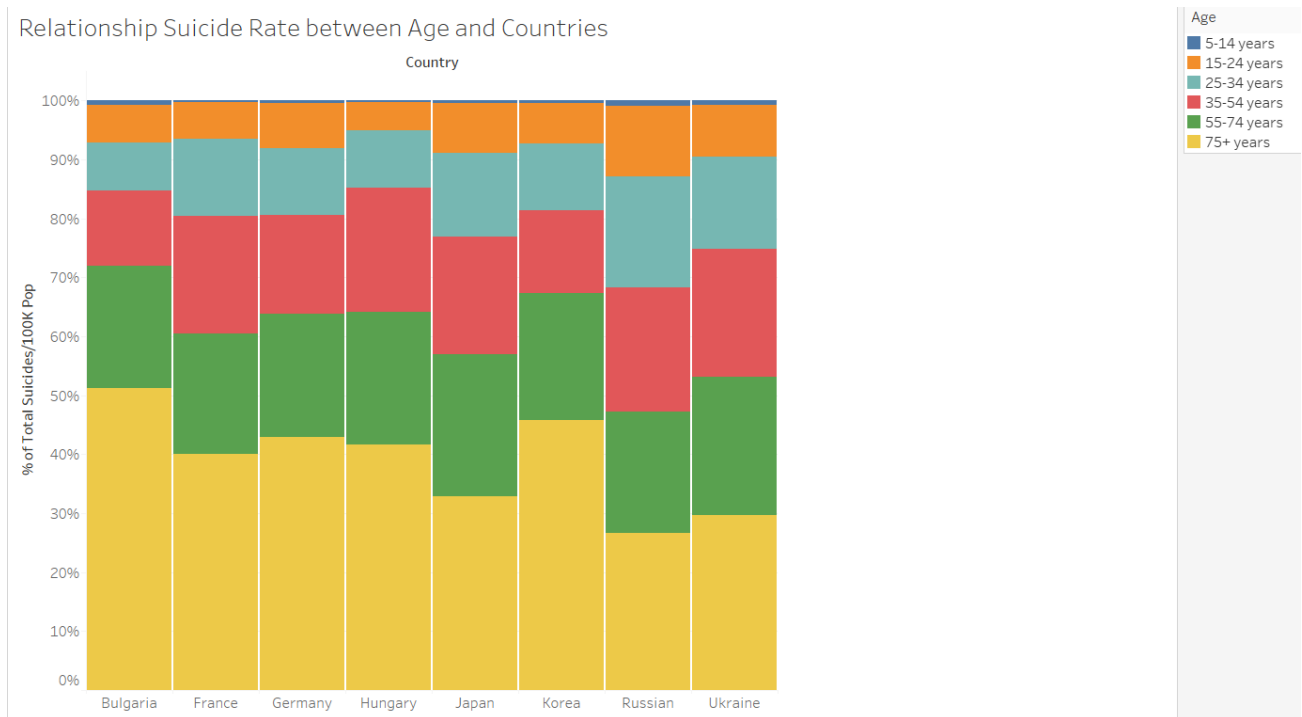


Figure 5. The Relationship of Suicide Rate between Age and Country

The mosaic plots (Figure 5) show the suicide rate by age group in different countries, Bulgaria, France, Germany, Hungary, Japan, South Korea, Russia, and Ukraine from 1985 to 2014. There are 6 age groups shown with 6 different colors, which is 5-14 for blue color, 15-24 for orange color, 25-34 for sky blue color, 35-54 for red color, 55-74 for green color, and 75+ for yellow color. These age groups of a country use percentage total so that all groups divide up proportionally. To contrast the difference between age groups, 6 colors are placed opposite side of the color bar.

Figure 5 represents the suicide rate of different age groups in different countries in Eurasia continent. In dataset, some countries have high numbers of suicide such as Russia, Japan, France, Germany, and Ukraine. Although United States and Brazil are proportionally high, there are some countries, especially Eurasia continent countries, are significantly high proportion of suicide rate. The plots show only 6 countries, Bulgaria, France, Germany, Hungary, Japan, South Korea, Russia, and Ukraine about the suicide rate in different age groups. The plots display in common thing is that there are significant high rates of people who are over 75 years old. Especially Bulgaria and South Korea are remarkably high portion of the suicide rate in all age groups.

3.6 Choropleth Map of Age Group

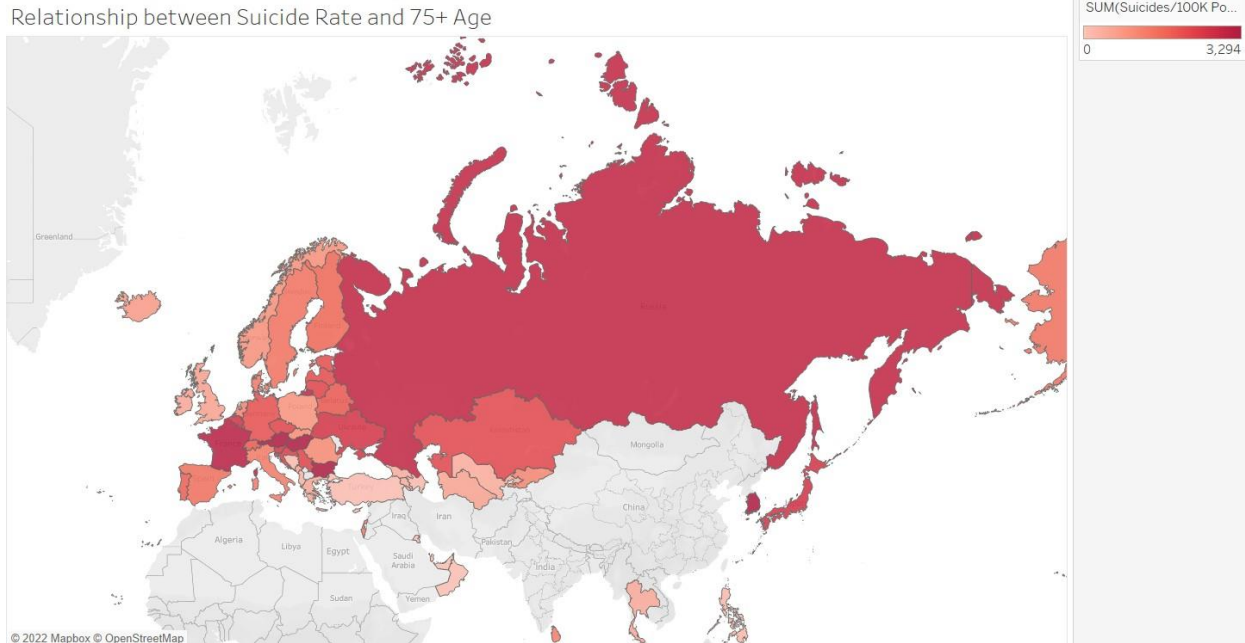


Figure 6. The Relationship between Suicide Rate and Age 75+

The choropleth map (Figure 6) represents the suicide rate per 100K population by over age, over 75 years old in Eurasia continent. The suicide rate of each country in Eurasia continent is calculated by taking the average suicide rate per 100K population of country from 1985 to 2014. Each country is filled with red color and the saturation of the color shows the suicide rate in Eurasia continent. The darker intensity represents the higher suicide rate in that country.

Figure 6 helps people understand the relationship between suicide rate and age, over 75 years old according to the choropleth map. After visualizing the suicide rate by 6 age group in the exploratory analysis, the over 75 years age group shows extreme high suicide rate in both male and female groups. The male, over 75 years age group, is the highest suicide rate all age group, but interesting point is the female, over 75 years age group. Compared to other female age groups, this is enormous increased in 1990s. I think it is a good way to check this relationship by geographically because the geographical position is included in various factors such as economics or climate. The choropleth map is easy to see the highest suicide rate all countries in Eurasia continent and see the pattern of the suicide rate in the map. The highest countries are Russia, France, Hungary, and South Korea. There is not a significant pattern of this map that these countries were different economic level, in order to there are included in both the developed countries and developing countries. And there are no clear patterns of climate factor.

4. Visualization Code

4.1 Tableau Sheet Figure 1

Choropleth Map of Global Suicide Rates per 100,000 Population for the years 1985 to 2016

Description of Region

Map based on Longitude (generated) and Latitude (generated). Color shows average of Suicides/100K Pop. Details are shown for Country. The data is filtered on Year, which ranges from 1985 to 2016. The view is filtered on Latitude (generated) and Longitude (generated). The Latitude (generated) filter ranges from -33.20 to 66.22. The Longitude (generated) filter ranges from -158.6 to 178.0.

Marks

The mark type is Map.
Stacked marks is off.

Shelves

Rows:	Latitude (generated)
Columns:	Longitude (generated)
Filters:	Latitude (generated), Longitude (generated), Year
Level of detail:	Country
Color:	Average of Suicides/100K Pop

Dimensions

Country has 100 members on this sheet

Members: Albania; Antigua and Barbuda; Brazil; Jamaica; Kiribati; ...

Year ranges from 1985 to 2016 on this sheet.

The filter associated with this field ranges from 1985 to 2016.

Measures

Geometry (generated) has 100 members on this sheet

Members: MULTIPOLYGON(((-32.3996 -3.8306, -32.3977))

Latitude (generated) ranges from -33.17 to 66.22 on this sheet.

The filter associated with this field ranges from -33.20 to 66.22.

Longitude (generated) ranges from -158.6 to 178.0 on this sheet.

The filter associated with this field ranges from -158.6 to 178.0.

Average of Gdp Per Capita (\$) ranges from \$720.73 to \$68,798.39 on this sheet.

Average of Suicides/100K Pop ranges from 0.00 to 40.42 on this sheet.

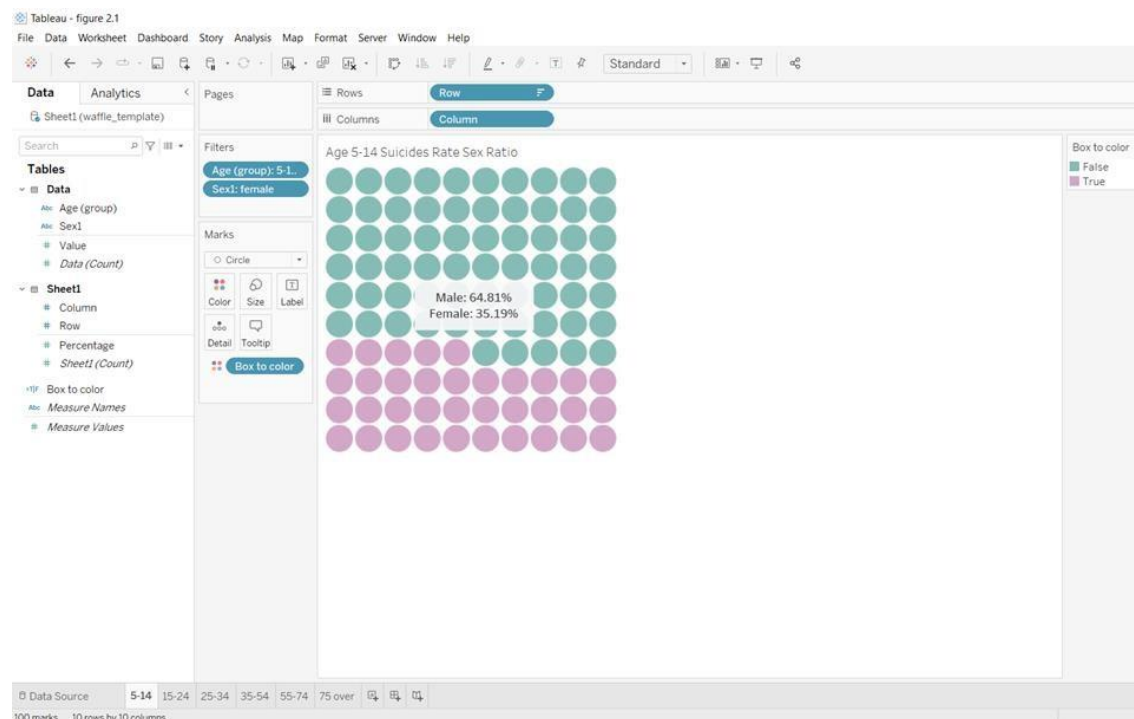
Data Source Details

Data Source:	master
Type:	Federated
Table:	master#csv

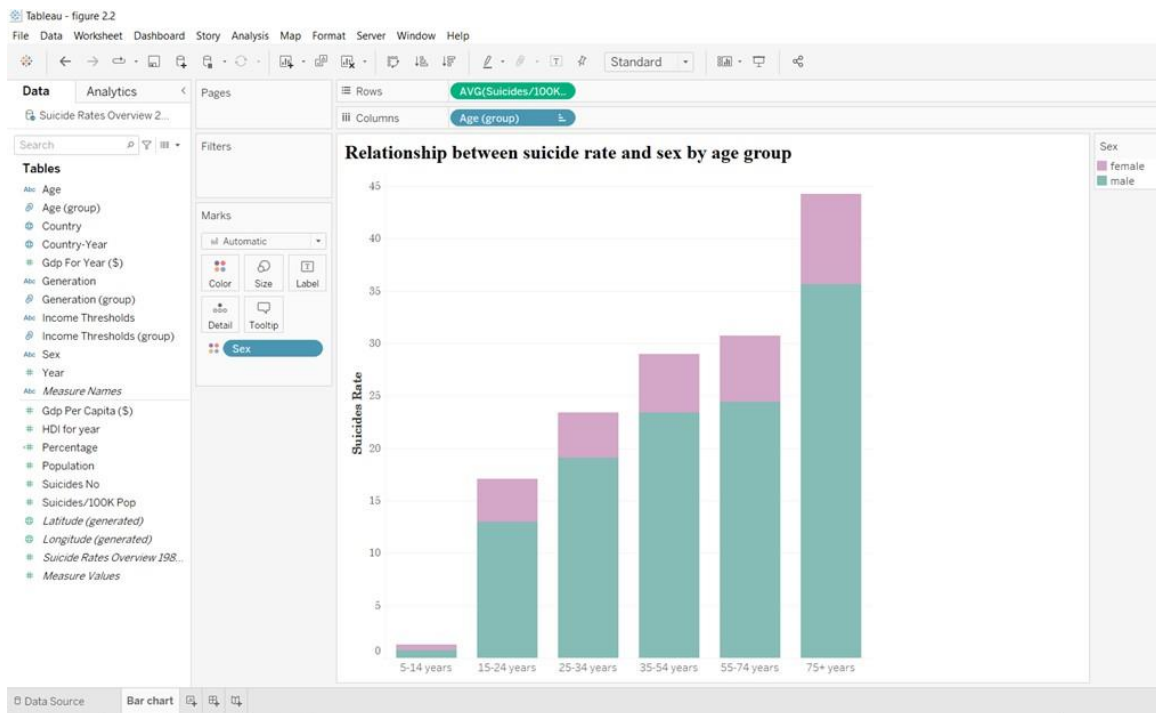
4.2 Tableau Sheet Figure 2

Waffle and Stacked Bar Chart of Suicide Rates and Sex by Age group for the years 2000 to 2014

2.1. Waffle Charts



2.2. Stacked Bar Chart



4.3 R Code Figure 3 & 4

Stacked Bar Graph for Gender and Age group for number of suicides

Change in Suicides per 100k Population as Economy Grow over the Years.

Code of Spot Matrix

```
# Spot Matrix
ggplot(data=suicide_rate_reduced, aes(generation, sex)) + geom_point(aes(size = suicides_no)) + ggtitle("Spot Matrix")

# Stacked Bar Graph
ggplot(data = suicide_rate_reduced, aes(x = age, y = suicides_no, fill = sex)) +
  geom_bar(stat = "identity") +
  labs(x = "Age", y = "Number of Suicides",
       title = "Number of Suicides for Different Age Groups by Gender \n",
       fill = "Sex") + theme(legend.position = "bottom")

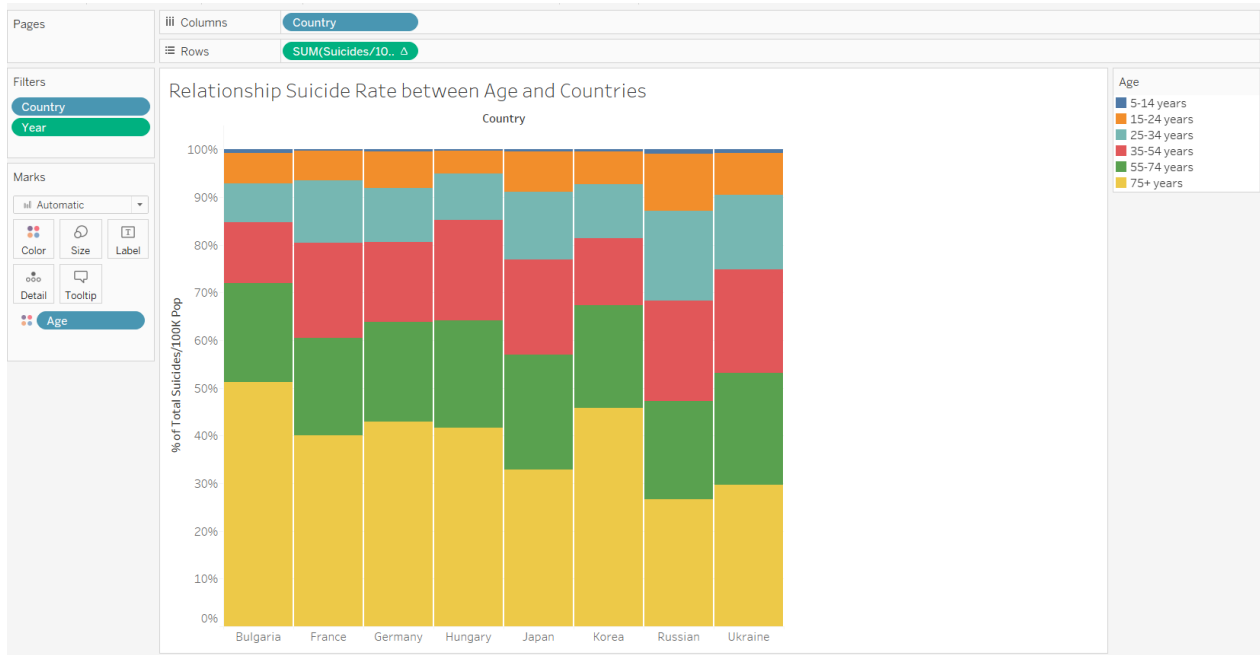
# Animation
x <- ggplot( data = suicide_rate_reduced,
             aes(x = economic_growth, y = suicides.100k.pop, size = population, colour = i..country)) +
  geom_point(show.legend = TRUE, alpha = 0.7) +
  guides(size = "none") +
  scale_color_viridis_d() +
  scale_size(range = c(2, 12)) +
  labs(x = "Economic Growth", y = "Suicides per 100K population", title = "Change in suicides per 100K population as Economy grows throughout the years",
       colour = "Country") +
  scale_x_log10() + transition_time(year) + labs(title = "Year: {frame_time}")

animate(x, height = 700, width =1200)
```

4.4 Tableau Sheet Figure 5

Mosaic Plots of Suicide Rates between Age and Country from 1985 to 2014

Figure 5. Mosaic Plots



4.5 Tableau Sheet Figure 6

Choropleth Map of Suicide Rate by 75+ Age Group from 1985 to 2014

Figure 6. Choropleth Map

