Research Topic: The pattern of suicide trends and the impact due to economy

Table of Contents

I. Research topic ……..……1
II. Data ……..……..2
III. Exploratory and explanatory data visualization ……3
IV. Conclusion and evaluation ……14

By Shiho Numakura for Data Visualization and the Web Individual Project

Research Topic:

Question

Is there a pattern in suicide rates and an impact due to the economy?

Background

Suicide is "the act or an instance of taking one's own life voluntarily and intentionally." (Merriam-Webster.com) Throughout time, people from many countries, some more than others, across the world decide to commit suicide for various reasons. From young to old, female to male, suicide is a problem for any humans living in this world. However, in recent years, there are numerous researches conducted to find preventions against such extremely unfortunate events from happening. Some research reveals the correlations between the suicide rates and the economy of the country. (Yang) https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1536-7150.1992.tb02512.x

This research incorporates suicide statistics of 1995-2015 from several countries to discover a pattern in the United Kingdom's suicide trends. It aims to find a pattern in the demographics of people who have committed suicide. Furthermore, the objective is to also discover the impact on the pattern of suicide rates due to the country's economy in the interest of seeking an explanation to its trends by comparing the trends in the United States of America and Lithuania.

Important Concepts

- Suicide rate – number of suicides per 100k population
- GDP (Gross Domestic Product) per capita ($) – an important indicator of economic performance and a useful unit derived from a straightforward division of total GDP by the population. (focus-economics.com) https://www.focus-economics.com/economic-indicator/gdp-per-capita
- United Kingdom –country where this research is conducted in; made up of England, Scotland, Wales and Norther Ireland. Ranks average in suicide rates across the world.
- United States of America – the most developed country in the world. (Newsweek) https://www.newsweek.com/ranked-most-developed-countries-world-according-un-1124311?page=3
- Lithuania – a country in Europe with the highest suicide rate in the world. (U.S. News)

Data:

Data Source

The data used for this research was obtained from the Kaggle.com website, with the subject of "Suicide Rates Overview 1985 to 2016." The version used, version 1, was updated a year ago by Rusty. It is licensed to the World Bank Dataset Terms of Use and content description states "compiled dataset pulled from four other datasets linked by time and place" (Rusty, Kaggle.com). These four datasets referenced include United Nations Development Program, World Bank., Szamil, and World Health Organization.

Data Format

The data was directly downloadable from the website in a CSV (Comma-separated values) format with 27821 lines.

Data cleaning and pre-processing

Since the data file contained the suicide rates for numerous countries, I used the software library, panda, in python to collect only the necessary variables for this research before any data was used for analysis. As I selected only the three countries needed, I noticed that Lithuania was missing the data from 1985-1994 and 2015 for United Kingdom; therefore, the research only observes and analyzes the trends of the 20 years from 1995-2015 in the three countries.

All data used for this research was cleaned in python and organized directly in the dataframes.

Exploratory and exploratory data visualization:

Variables of Interest

- Country [nominal] – the selected three countries for this research.
- Year [interval] – statistics of 1995-2015 is used.
- Sex [nominal] – the gender of the statistic represents; either female or male.
- Age [ordinal] – the age group of the statistic represents; 5-14, 15-24, 25-34, 55-74, 75+ year old age groups.
- Suicide number [ratio] – the number of suicides in the specified demographic.
- Population [ratio] – the population in the specified demographic.
- Suicides per 100k population [ratio] – the number of suicides divided by the population in the specified demographic.
- GDP per capita($) [ratio] – used for country's indicator of the economy; description given important concept.

Summary of variables

Table 1 is an example of data that is used for this research.

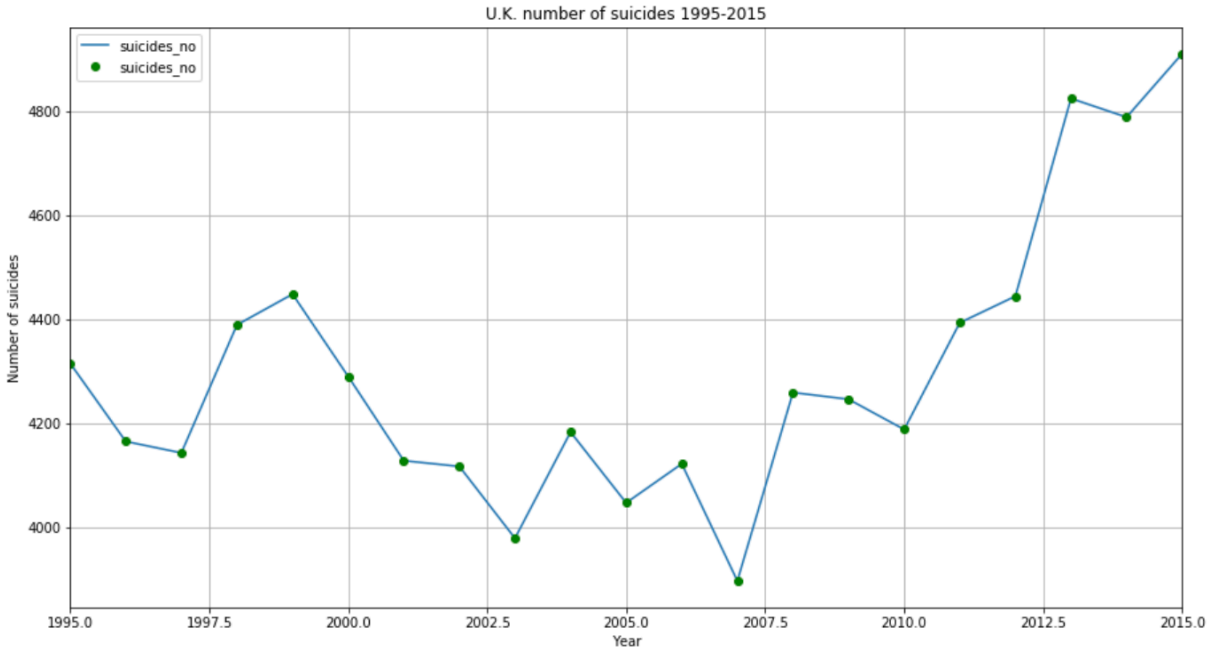
[Table 1]

	country	year	sex	age	suicides_no	population	suicides/100k pop	gdp_per_capita($)
9	United Kingdom	1995	female	15-24 years	81	3624373	2.23	24389
8	United Kingdom	1995	female	25-34 years	143	4614434	3.10	24389
6	United Kingdom	1995	female	35-54 years	344	7672199	4.48	24389
11	United Kingdom	1995	female	5-14 years	4	3664274	0.11	24389
7	United Kingdom	1995	female	55-74 years	237	5746814	4.12	24389
5	United Kingdom	1995	female	75+ years	152	2692330	5.65	24389
4	United Kingdom	1995	male	15-24 years	419	3811646	10.99	24389
0	United Kingdom	1995	male	25-34 years	858	4776564	17.96	24389

Data visualization and key findings

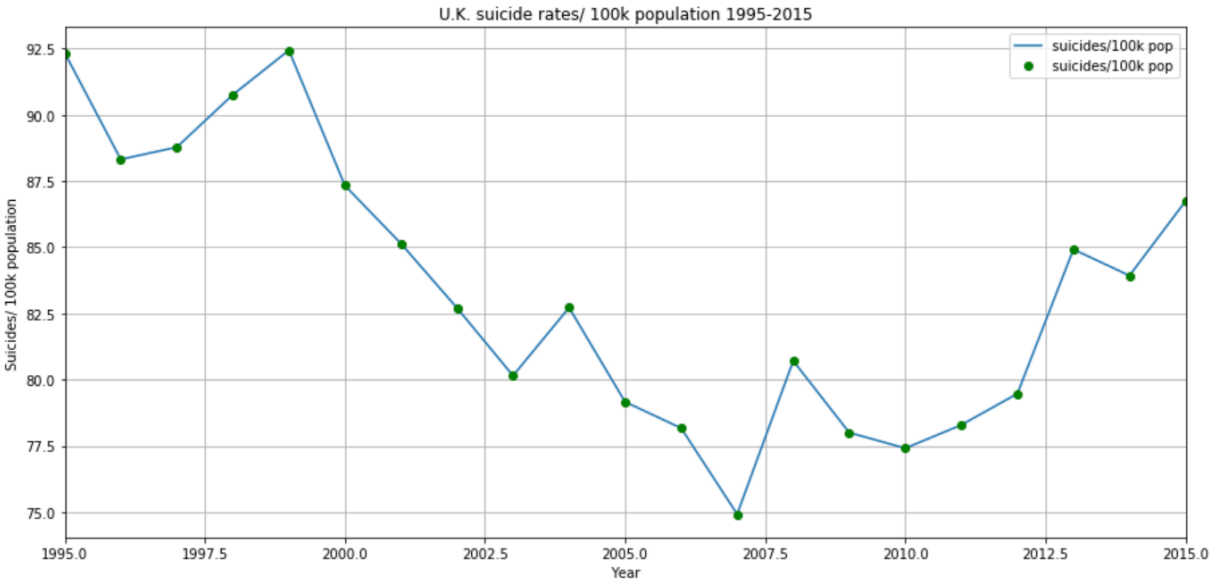
Graph 1 shows the number of suicide rates in the U.K. over time from 1995 to 2015.

[Graph 1]



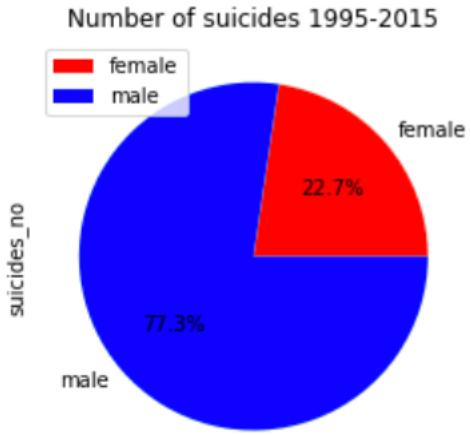
Although there is a clear indication that the number of suicides were decreasing, reaching their lowest number in 2007, the number seems to rise more after 2007. However, Graph 1 represents the raw number of suicides and lacks to incorporate the potential change in the U.K.'s population. Graph 2 shows the suicide rates in the U.K. over time from 1995 to 2015.

[Graph 2]



The suicide rate is showing rather more a drastic decrease over time until 2017 similar to the trend seen in Graph 1. However, it remains to stay lower than prior to 2007 compared to the drastic rise in Graph 1. We can also conclude that the suicide has been overall decreasing, for 1995 has the highest suicide rate in the 20 years. Nevertheless, it is important to consider the positive slope from 2007 will likely to continue its pattern after 2015. Moreover, I will be focusing on the suicide rate from now on in this research rather than the raw number of suicides to avoid bias with the differences in its populations.

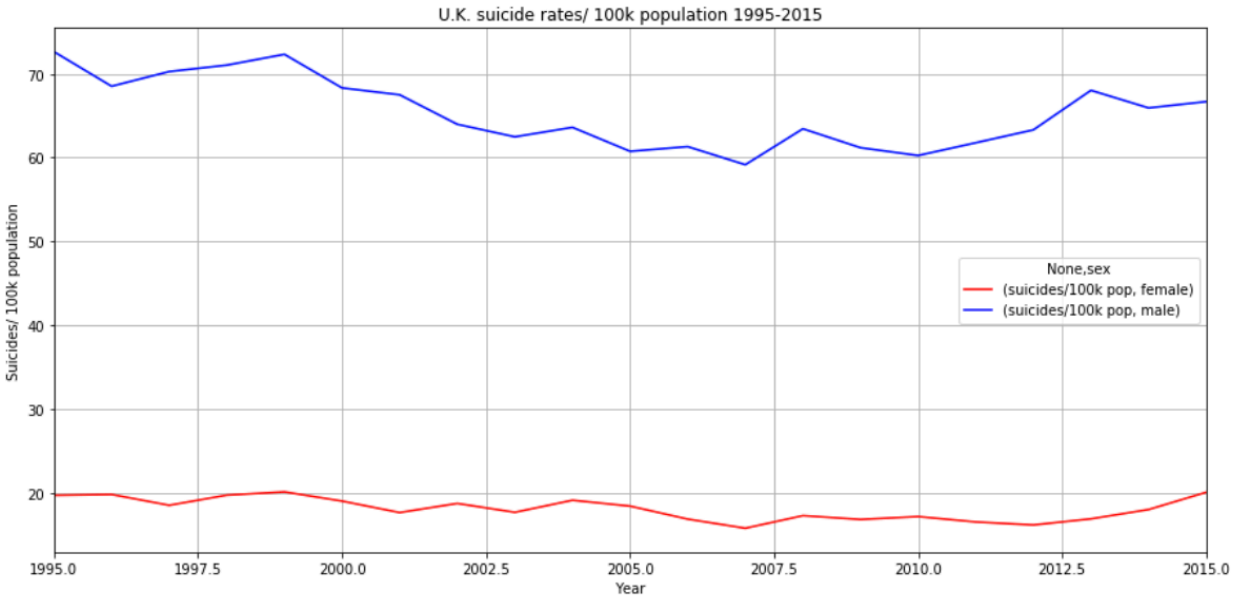
Now, we will observe further into the demographics in these trends. Graph 3 shows the ratio between the gender that is responsible for the total number of suicides in the 20 years.



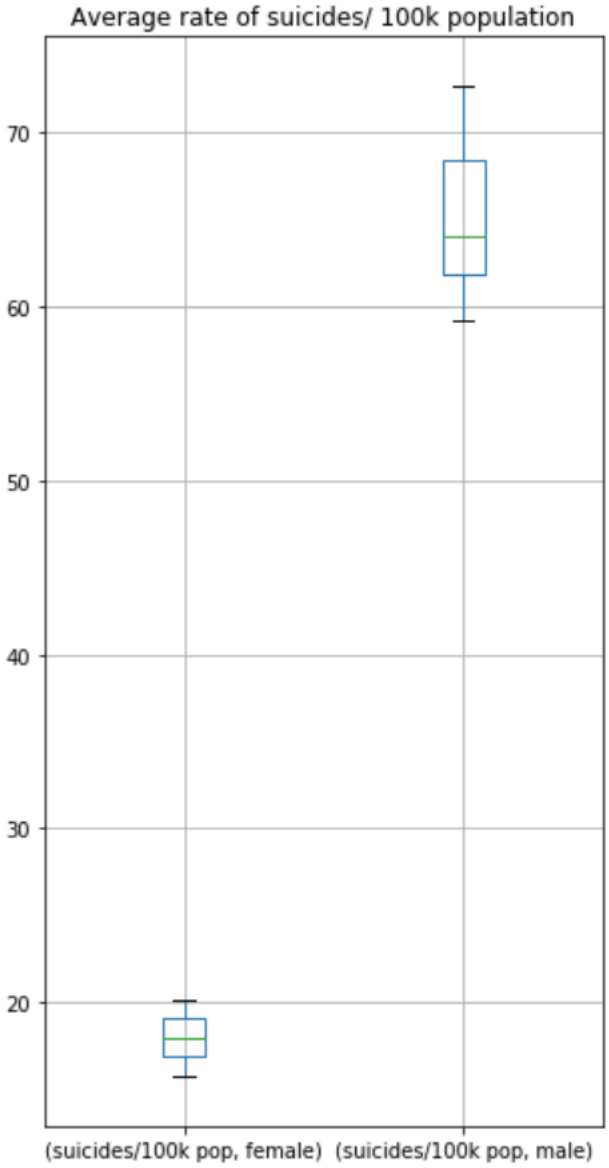
[Graph 3]

Graph 3 reveals that a significant number of suicides are committed by male. Graph 4 shows trend in the suicide rate of the genders in 1995-2015.

[Graph 4]



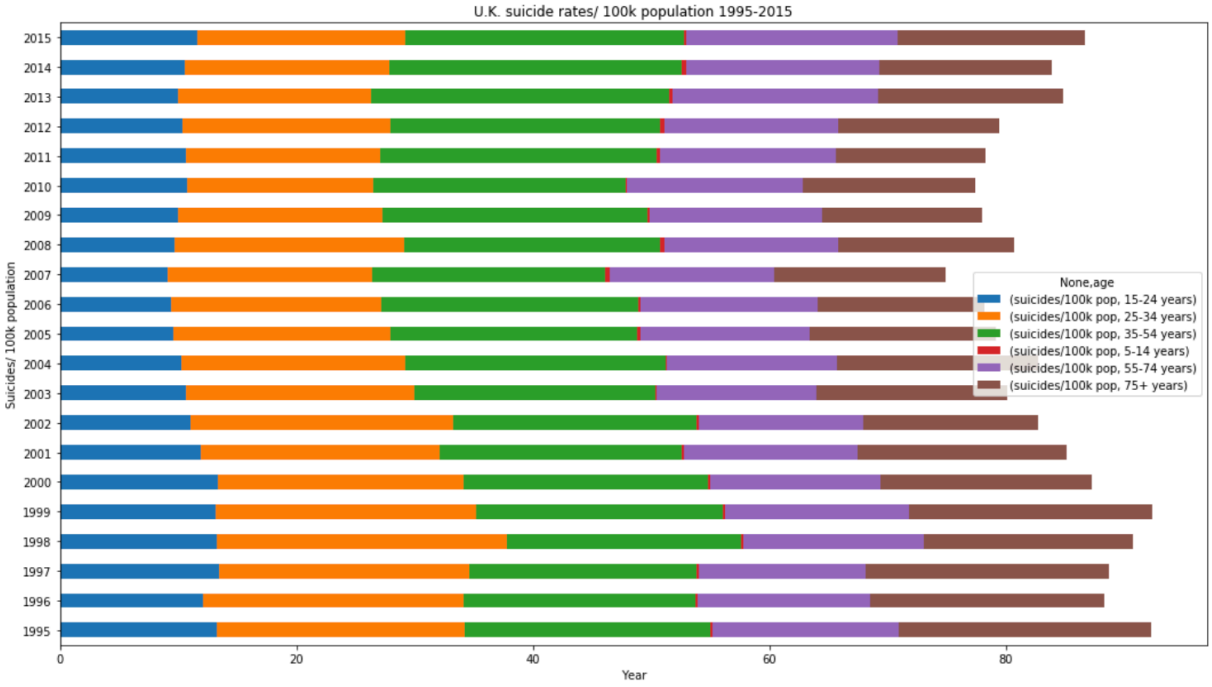
Graph 4 indicates that for over 20 years, male are constantly likelier to commit suicide. Graph 5 reveals further difference in suicide rates by the gender.



[Graph 5]

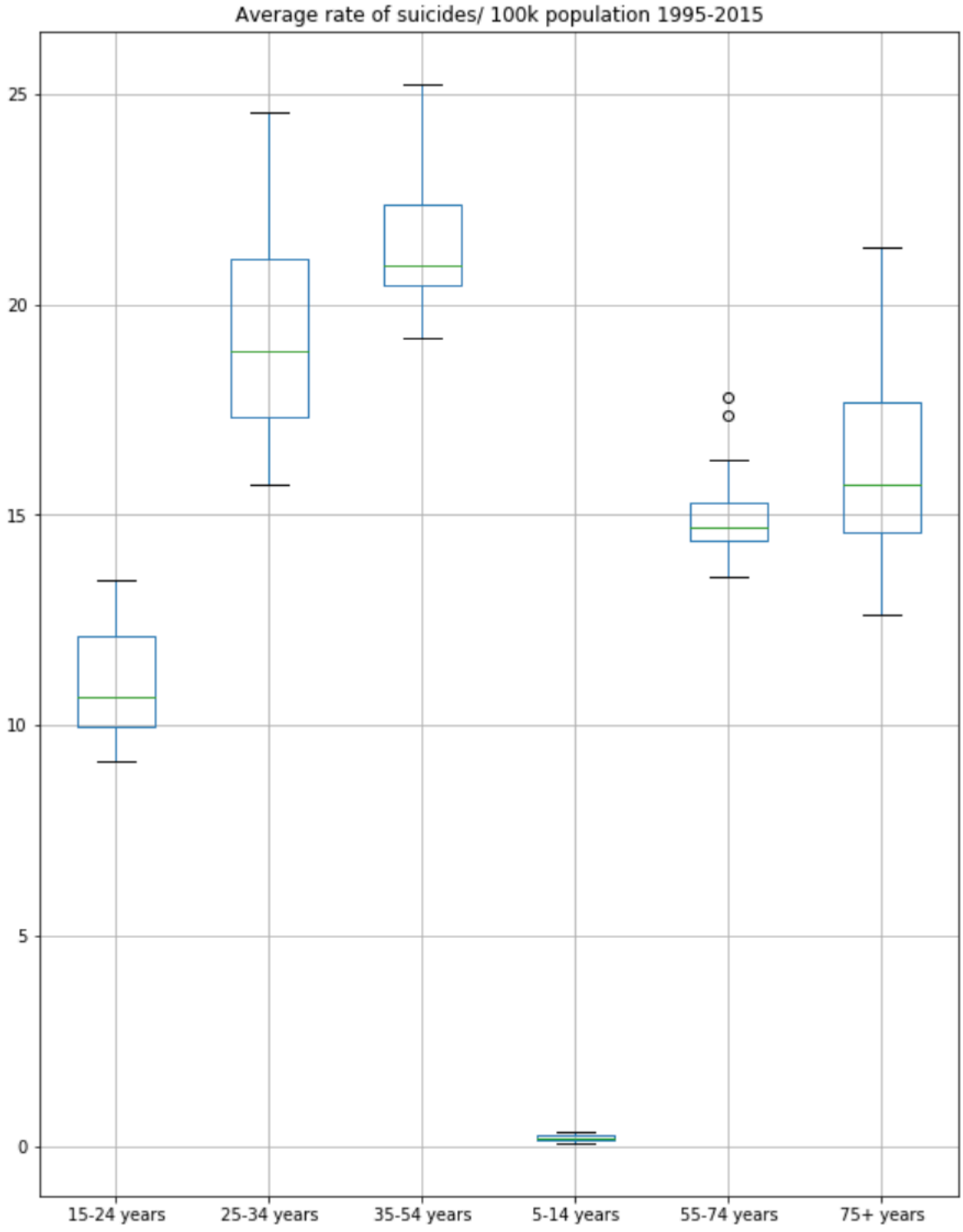
According to Graph 5, it is 3 times likelier that a male will commit suicide than females. Next, we will analyze the trends based on age groups. Graph 6 shows the trend in suicide rates between 1995-2015 by age groups.

[Graph 6]



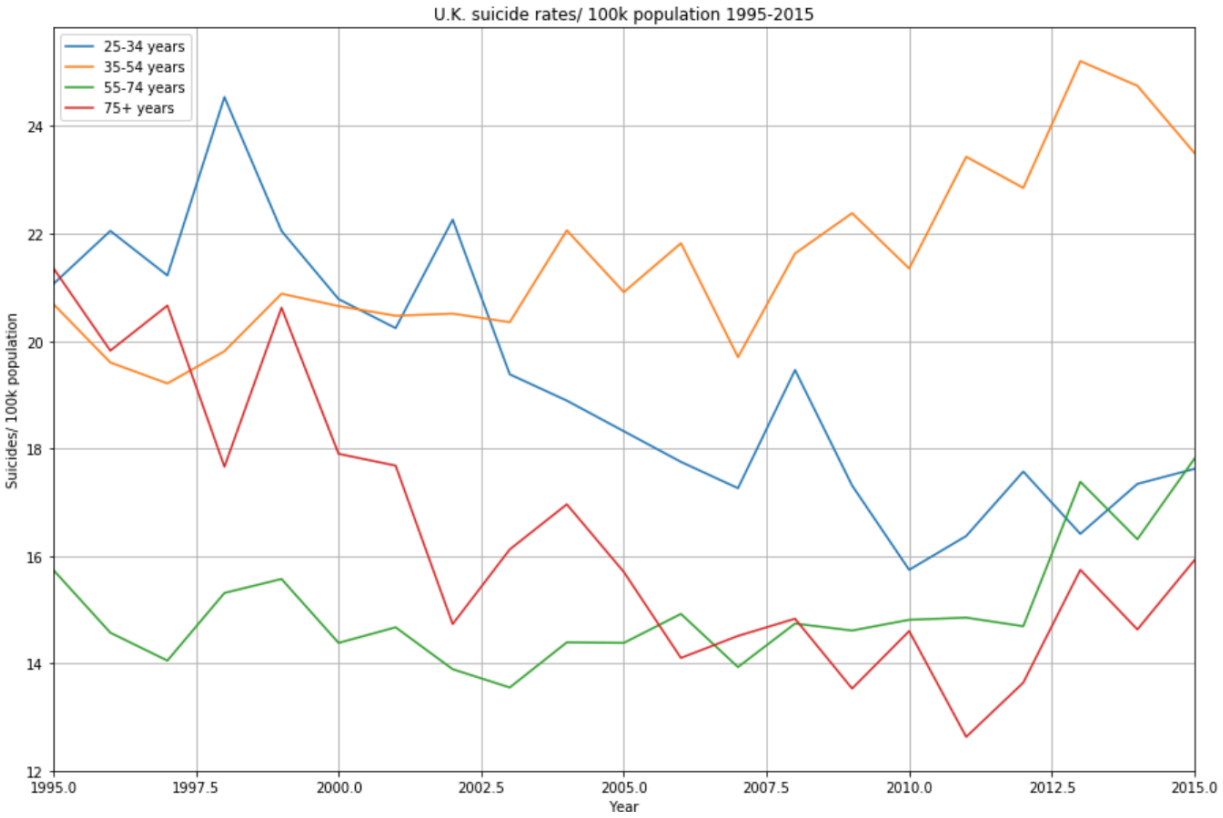
Unsurprisingly, Graph 6 shows that there is little rate of people ages 5-14 years old committing suicide. Nevertheless, aside from the 15-24 year old committing suicide being the next least popular age group to commit suicide, it is hard to conclude otherwise. Graph 6 shows the identical data in box graph for a better understanding.

[Graph 7]



Graph 7 shows that the age group 35-54 year old has the highest average suicide rate between 1995-2015. However, Graph 6 and 7 lacks to show a clear pattern of the changes in suicide rates for age groups from 24 – 75+. Graph 8 will present the pattern of these age groups.

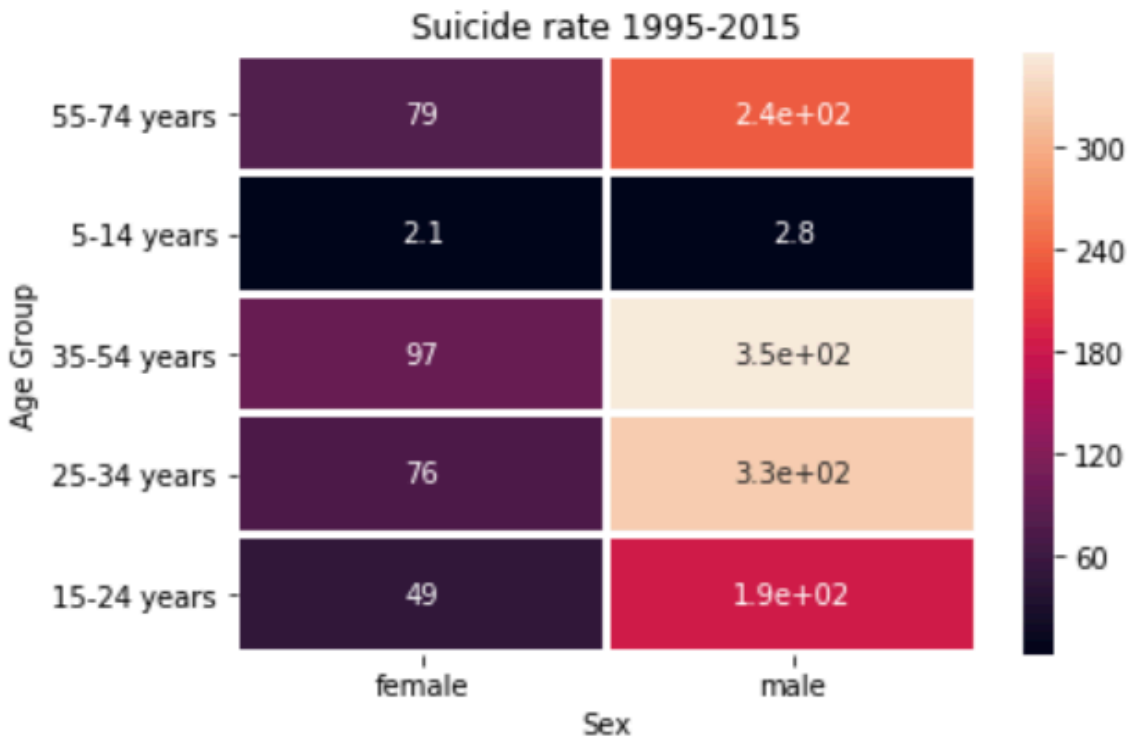
[Graph 8]



As we can see the suicide rate of age group 25-34 and 75+ year old began to decrease, age group 35-54 year old began to increase. Age group 55-74 year old remain relatively steady, showing some increase in their suicide rate.

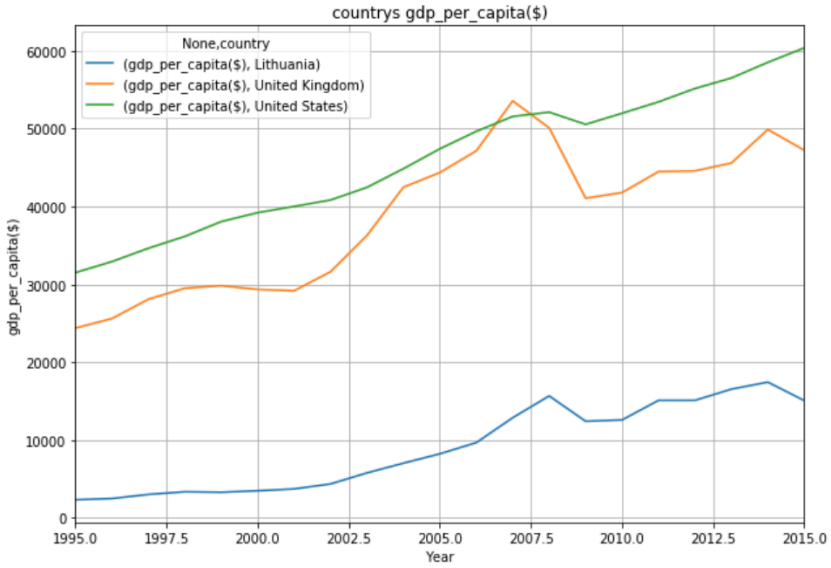
Finally, Graph 9 incorporates both gender and age group to find the demographic that has the highest suicide rate throughout the 20 years analyzed.

[Graph 9]

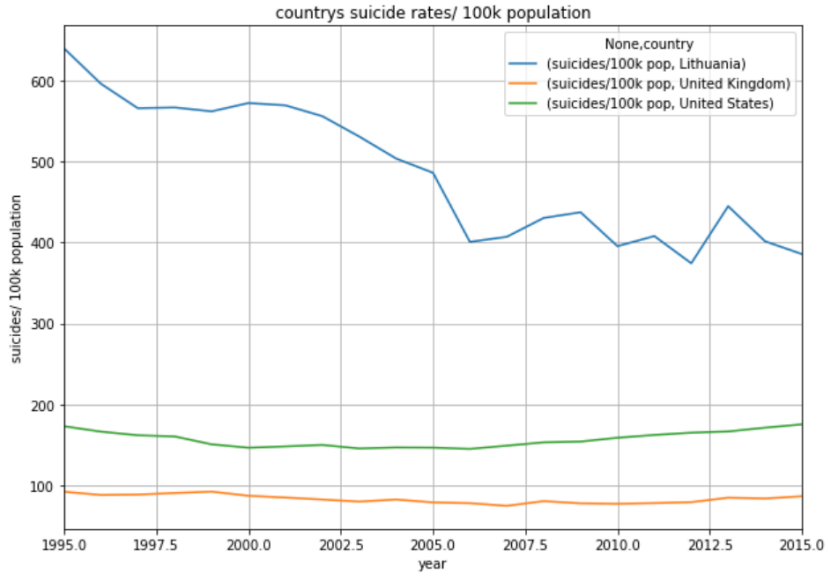


Graph 9 clarifies previous conclusions, the high suicide rate by male and age group 35-54 year olds.

Now, we will seek for the pattern's correlation to the economy. Graph 10 shows the corresponding country's economy.
[Graph 10]



As expected, U.K. and the U.S. are developed countries with well economy compared to Lithuania. Graph 11 shows the suicide rates comparing the three countries' suicide rate.
[Graph 11]



There is a significant difference between the U.K. and U.S. suicide rate to the rate in Lithuania. However, due to the great difference, it is hard to distinguish the pattern for the suicide rates in the U.S. and the U.K. Graph 11.5 is simply a graph that zooms into the two country's suicide rate trends.

[Graph 11.5]

Although there is no significant pattern that compares to the economy for the three countries, there is a general correspondence between the maximum and minimum point of the two variables found. As expected, when the economy is well, suicide rate generally goes down as well. Nevertheless, Lithuania which had the highest suicide rate had the lowest economy in the three countries used in analysis. Furthermore, age group 35-54 year old seem to be affected by the economy in the U.K. the most. This is likely due to 35-54 year old adults are often the age group who focus on work to earn money compared to other age groups.

Brief description on data type choice and human perception

As explained earlier in the research, I focused on the suicide rate rather than the actual number of suicides due to the different number of population size of the demographics. In creating visualizations on the data, I made sure to have the same variables to have the same color. Moreover, each graph was carefully chosen to communicate the goal of each analysis efficiently. This was done with close evaluation of the data pattern and the relationships of the variables.

Conclusion:

Summary of key findings

In the U.K., about 75% of the suicide were committed by male, that were ages 25-54 year old. While the impact due to the economy was not as significant as expected, there were slight correlations to the highest economy and lowest suicide rates. This was most clear for age group 35-54 year old, but no clear difference to how it affected the genders.

Process and visualization

While I faced some complications with using the tables that was organized using pivot, I was able to understand the role of each graph to efficiently decide the types of graph in each analysis. The research was carefully organized as well to create each step to be easily understood and follow by the audience.

New discoveries

During the course of this assessment, I have discovered how to represent and manipulate data in a graph using Python.

Future improvements

I wish to improve an easier way to compare the economy to the suicide rates rather than having two graphs to compare by the human eye. Moreover, three sample countries to make an indication on the impacts of economy to the suicide rate is not enough and should use more samples.