Trend of Suicide cases over the years and its relationship between economy and sex

Yue Chen Shen 1005350790

December 14, 2020

Abstract

In this report, the data based on the suicides rates from 1986 to 2016 found on Kaggle website. Kaggle is a large online data science community, where users will post useful datasets and other resources for the community to share and explore. This report will be using suicide cases value from different years to explore the trend and pattern, also using variables such as GDP and sex to explore the relationship they have with suicide cases.

Introduction

Mental health is a common and serious topic that is important to everyone, many factors contribute to mental health and can even result in suicide in serious situations. Factors such as bullying, stress, depression and many more can all lead to mental health issues and suicide. Nowadays, technology can also contribute to cyber bullying, online threat, fraud, which are also factors contributing to suicide. Throughout this analysis, the target country is Albania, a diverse and religious country located in Southeast Europe, I am interested in finding out that since 1986, does suicide cases keep increasing as year increases, what are the trend and pattern of suicide cases over the years. Besides, I will also be interpreting the relationship between GDP and suicide cases, to see if there are more suicide cases during times of economic recession and vice versa. For the relationship between years and suicide cases, histogram will be constructed to give a precise and general trend and shape of the relationship. As for the relationship between GDP and suicide cases, a scatterplot will be used to observe the relationship between the two. A linear regression model will also be used to predict how much suicide cases will go down as GDP increases. Lastly, a bar graph will illustrate the volume of suicide cases according to male and female. By going through these plots and regression, we will be able to find out the trend of suicide over the years, how greatly is economy contributed to suicide death and how different sex is associated with different number of suicide cases.

Data

The data csv file is downloaded from the website Kaggle, https:\\www.kaggle.com (https:\\www.kaggle.com) The first scatterplot, figure 1, I want to explore the relationship between the country's GDP and its number of suicides. Using the plot function, I generated a scatterplot with GDP on the x axis and Suicide No on the y axis, I also used the abline function to construct a linear line through the points to see the trend. I want to find out if increasing in GDP value will leads to a decrease in suicide numbers, or there is no clear indication of relationship between the two. Both suicide no and GDP are numeric variables, suicide no has values of suicide cases from 1986 to 2015 and consists of all age group. The other variable GDP represents Gross domestic product per capita, also can be described as how much value Albania has created within that certain year. From the scatterplot, we clear see that there are more dots, more suicide cases when the GDP is at a very low value, and as GDP value moves up, the

suicide number dots are gathered at a low value area. Since there are repeated values in the GDP column, the linear line does not indicate a very clear trend and the dots are bit overcrowded. However, the pattern of the scatterplot and the linear model can still provide details on the relationship.

For Figure 2, I am analyzing the trend of suicide cases over the years to see how number of suicide cases are varying from year to year. I have used a histogram to build several bars to give a straightforward view of the trend, as well as the max and min value. The variable suicide is again about the number of suicide cases over the years. The variable year is still a numeric variable, the year variable is from 1985 to 2015. The year variable is a very simple and straightforward variable, and we use ggplot to relate year and suicide number together and observe the trend.

For figure 3, I make a new plot called trend of suicide cases over the years, which is similarly to the histogram in Figure 2, the variables are the same as the variables used in Figure 2. The reason for constructing this extra plot is that I want to further investigate and confirm the trend of suicide over the years, I also want to compare the results between the two type of graphs and come up with a more accurate and balanced results by using the results from the two figures.

For Figure 5, I choose to explore the gender difference when it comes to suicide numbers, I want to know whether male tends to commit suicide more than female does or vice versa. The variable sex is a categorical variable which contains female and male, again from the sex variable column, there are many repeated values, so I use the group by function to get a cleaner data. Then I combine sex variable and suicide numbers into a bar graph, by looking at the two different bars from the graph, I can know and compare the difference in suicides numbers between male and female.

Model

In this report, there are three focus that I mainly want to explore, which is the relationship between suicide and economy, the overall trend over the years and the difference between sex in suicide numbers. In order to explore these relationships, I have constructed two different models and plots using r studio.

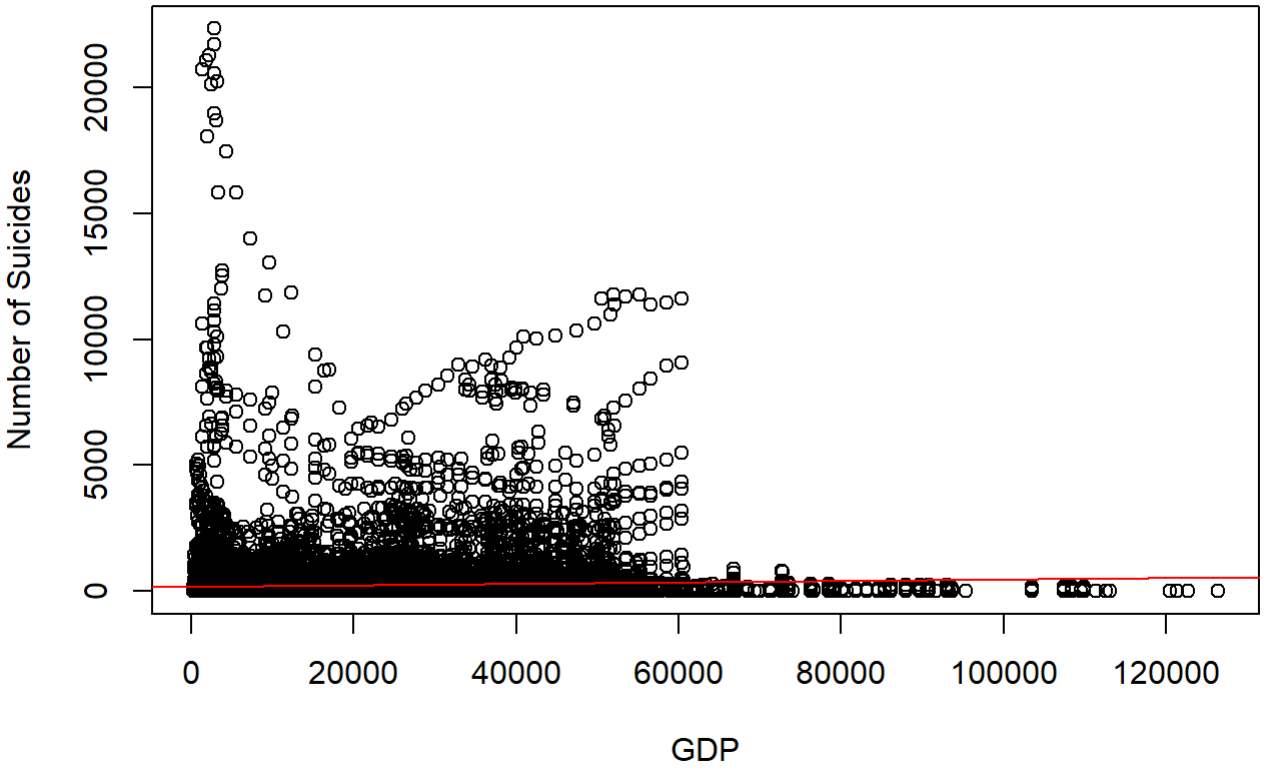
The first model is that I used r code to construct scatterplots to explore the relationship between GDP and suicide numbers, and on the trend of suicide cases over the years. The first scatterplot from figure 1 shows how suicide numbers are varying as GDP value increases. The second scatterplot from figure 3 shows the change in number of suicides given different year input. I chose to use scatterplot because I want to show the distribution between the two variables, scatterplot is more detailed compare to histograms and scatterplots can help identify a general trend and the overall distribution.

In order to further prove the trend of suicide cases over the years, I decide to use ggplot model to create a histogram and a bar graph, Figure 2 is a histogram which contains several bars ranging from 1986 to 2015, each bar represents different corresponding suicide numbers, together they form a shape and gives off a representation of distribution of suicide numbers across the year. The reason why I chose to include this extra histogram to support the trend of suicide numbers is because histogram can give off different characteristics of distribution, such as left-skewed, right skewed or normal distribution.

I believe that there exists a difference between male suicide rates and female suicide rates and it is hard to predict the result without using a graphic model. I use bar graph to find the gender difference in suicide, figure 4 is a bar graph that shows the number of suicide numbers for male and female. This bar graph can give a very straightforward comparison between suicide numbers for each gender.

Results

Figure 1. Scatterplot of GDP and Number of Suicides




```
##
## Call:
## lm(formula = suicides_no ~ gdp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -563.3  -222.9  -193.5  -105.5 22136.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.932e+02  7.237e+00   26.69   <2e-16 ***
## gdp         2.929e-03  2.858e-04   10.25   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 900.4 on 27818 degrees of freedom
## Multiple R-squared:  0.003761,   Adjusted R-squared:  0.003726
## F-statistic:   105 on 1 and 27818 DF,  p-value: < 2.2e-16
```

Figure 2, Suicide cases over the year

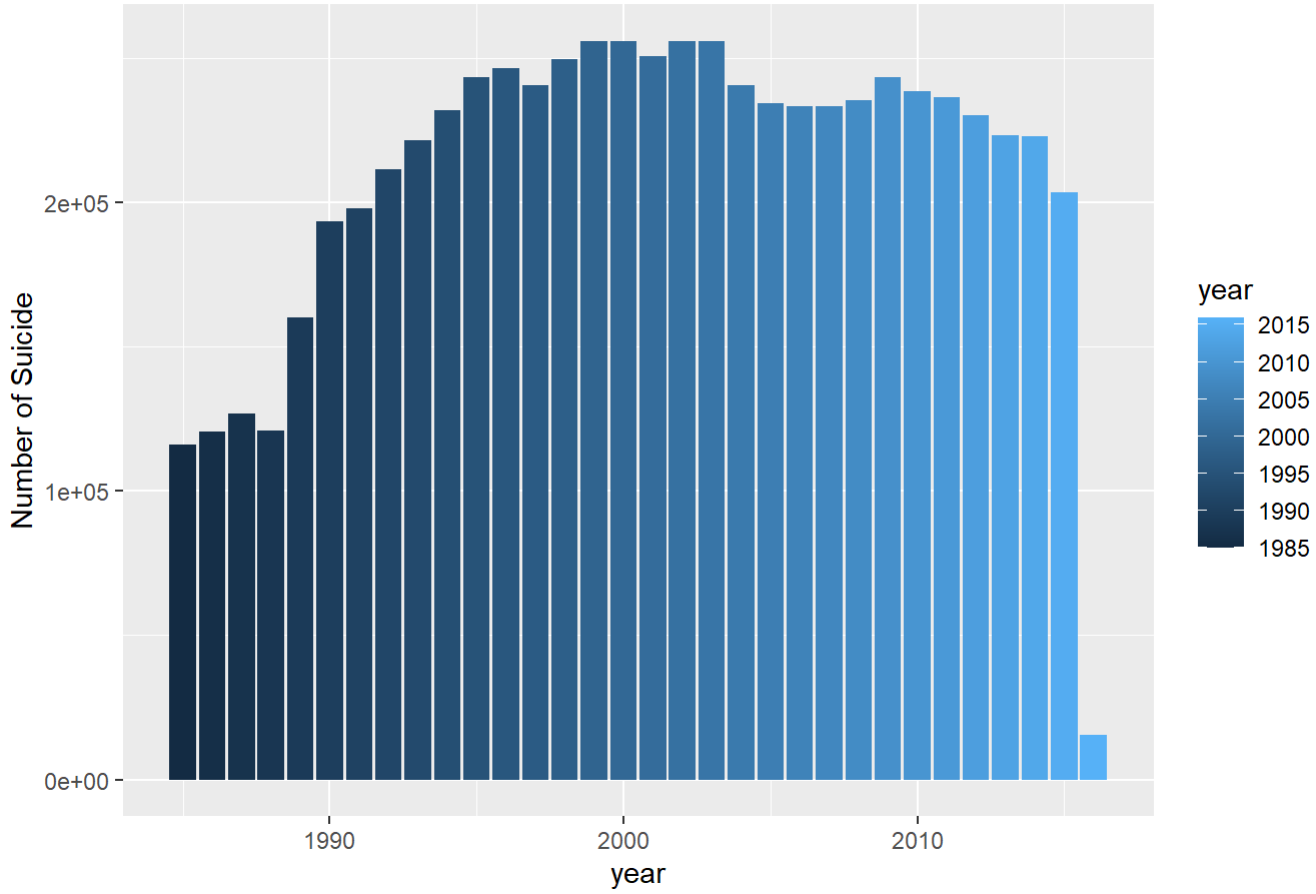


Figure 3. Trend of Suicide cases over the year

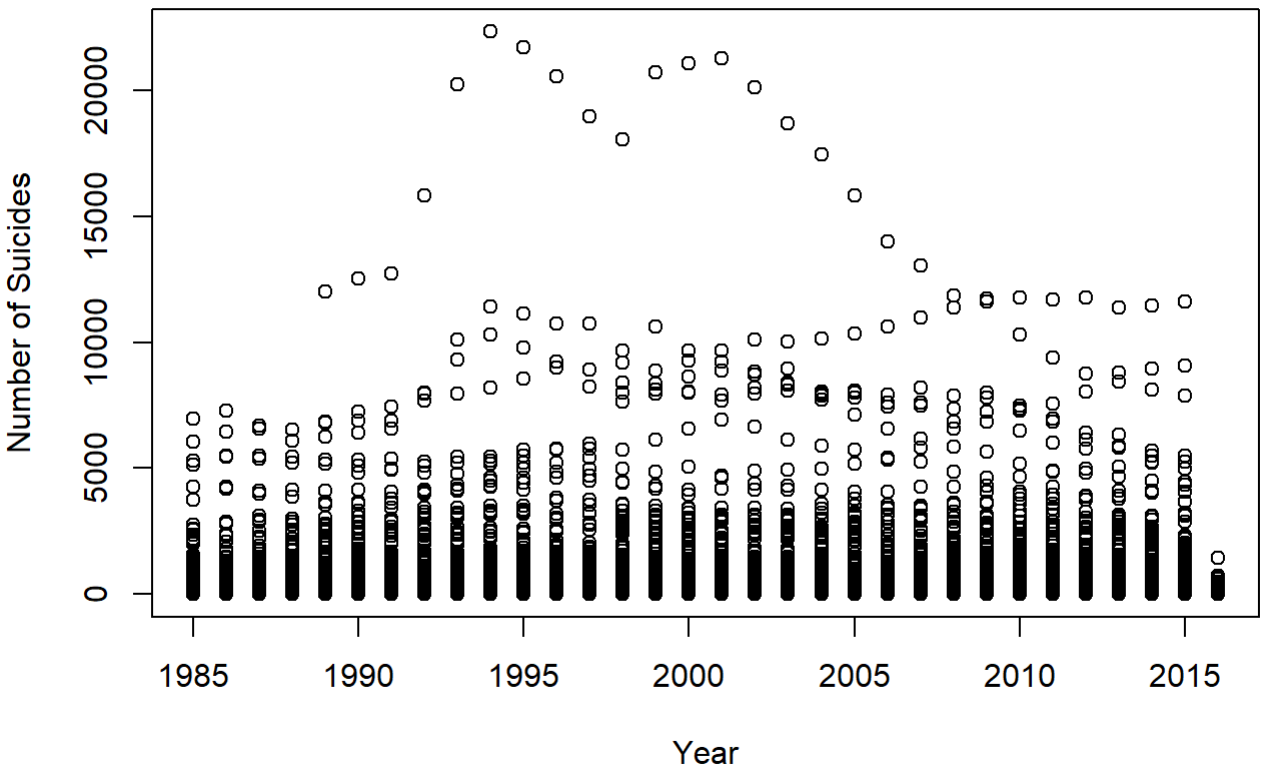
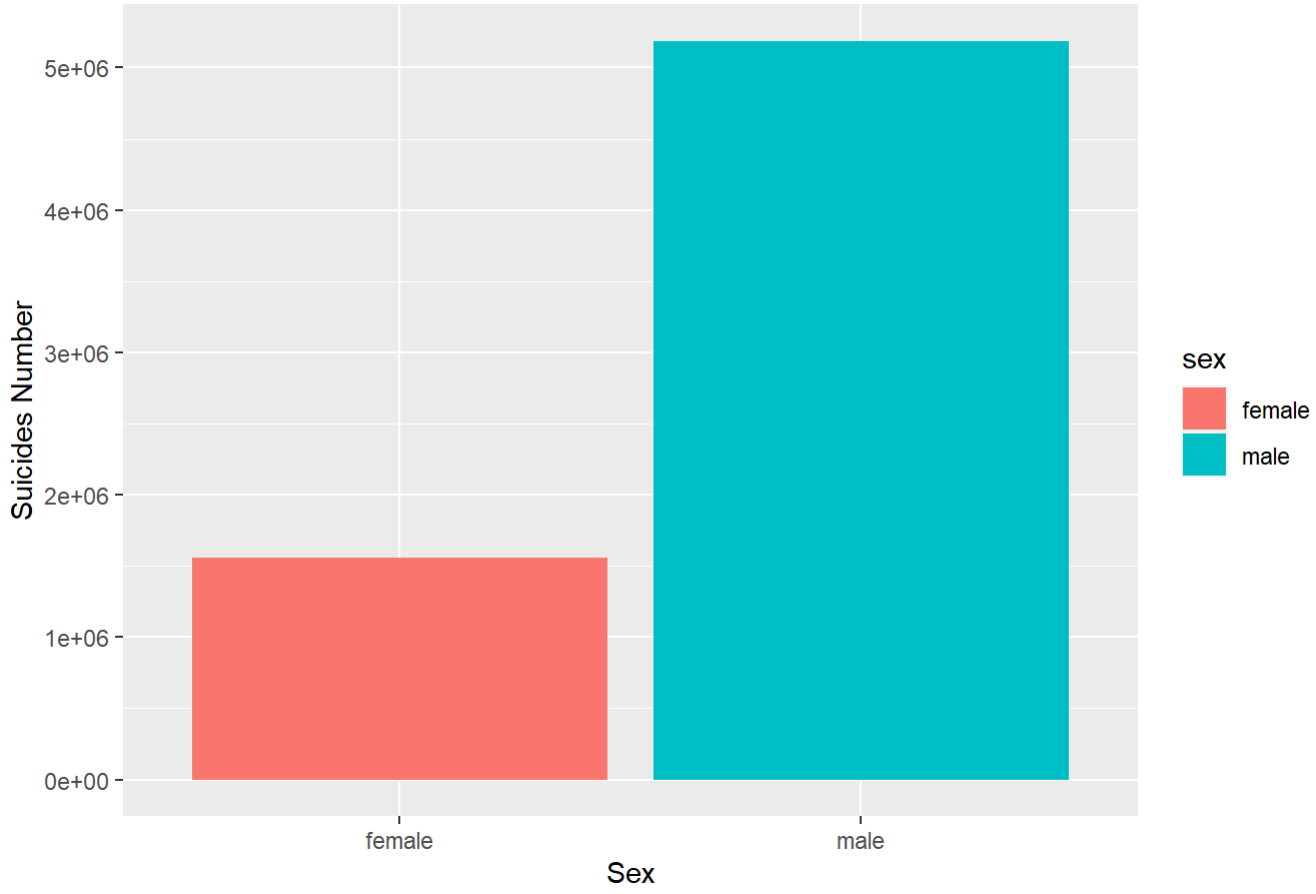


Figure 4, Number of Suicides by Sex



Discussion

Summary The final goal of this report is to observe the trend of suicide cases over the year, determine whether increasing GDP will cause a decrease in Suicide numbers and observe the difference between male and female suicide rates. The target country in this report is Albania, the reason why I chose the independent variable GDP is because I think financial crisis or economic issue can cause serious stress for adults and majority of the people, many people that own large amount of debt or face a serious business failure is likely to take the extreme and end their life. Since the data for this report is from 1986 to 2015, it becomes essential and necessary to take this opportunity to explore the trend of suicide numbers in Albania, to see whether it is overall decreasing or increasing. In terms of the last variable, I think sex is also an important topic in analyzing suicide, male and female have completely different styles and capability, which is going to create a difference when describing number of suicides. After deciding my variables, I have used and group the data to create graphic models to analyze and observe the pattern and relationship that these figures present.

According to the analysis and Figure 1, we can clearly see that the left portion of the graph contains higher suicide numbers whereas the middle and right portion of the graph have data points that are a lot lower. What this indicates is that there are more suicide cases during times of recession where GDP is at a low output, when GDP is growing or staying at a higher value, we see that the number of suicides is not as high. When GDP is at a very high value, which is what the right portion of the graph is showing, we see that suicide numbers are gathering near almost zero. However, according to the red line on the graph and the linear regression model associate with figure 1, the red line is horizontally flat and does not indicate an obvious slope, therefore we cannot conclude that increase in GDP leads to a direct decrease in suicide numbers. According to an article published by BMJ journals, economic growth is not the dominant factor in contributing to suicide rates, even if the economic is growing and trending up, without proper mental health services and corresponding welfare care, suicide rate can still increase

as GDP increase. Overall, we can see that during the times when Albania's economy and domestic output is a very low value, also known as recession, we see high suicide numbers. However, we cannot conclude that there is a strong negative correlation between GDP and suicide numbers since increase in GDP is not strongly correlated with decrease in suicide numbers.

In terms of the trend of suicide cases in Albania, the histogram from figure 2 shows a left-skewed distribution. We see that starting from 1986, the suicide numbers are relatively low, then gradually increases and reaches the peak around the year of 2000. After the peak in 2000, it later starts to decrease gradually with a slight increase in 2008, but the overall trend is declining after 2000. Comparing these conclusions with the scatterplot, we see that the scatterplot has two peaks, one around 1994 and the other one around 2000. Each graph shows the fact that suicide numbers in Albania is relatively low in 1986, reaches the peak in 2000, and then gradually declining until 2015. The reason behind the peak around the year of 2000 is very likely due to the Albanian Civil War. According to gloablsecurity.org, lots of people were killed during the civil war, government losing control and riots have spread over the country. During such unpleasant and dark times, social security and economic performance are both at their lowest, which could explain the peak of suicide rates that both figures have shown.

Through the analysis of gender and suicide numbers, the bar graph clearly shows a huge gap between male and female suicide numbers. The results indicate that in Albania, male suicide rates are approximately three times as high when compare with female suicide rates. In an article from BBC, the author mentions that one important risk factor is that "women are willing to share their problems and men tend to bottle them up" (Schumacher, 2019). This important natural characteristic suggests that men is less willing to share their feelings and their struggles, they tend to keep everything to themselves, which causes men not being able to get appropriate help when needed. This is an important factor across the world that cause the gender gap, men are less likely to seek for mental health help, which explains why men suicide rate is way higher than female suicide rate in Albania.

Weaknesses

The weakness of this analysis is the lack of mathematical support, the variables that I have pick for this topic is better when using graphs to analyze, which causes the majority of the analyze is based on graphic models and are being completed using histograms or scatterplot. The lack of mathematical model or regression analysis makes this report not capable of predicting the future outcome and future trend of the variable. The lack of mathematical model and regression also makes it harder to convince and discuss.

Next Steps

There are two things that can be improved upon, the first improvement can be made base on the fact of lacking mathematical model, building a linear regression model and analyzing the model to provide more findings on the relationship between GDP and suicide numbers. The second one is to use more variety of graphs, I will try to use other varieties of graphs other than scatterplot and histogram, to provide more detail and make the comparison easier.

References

"Why More Men than Women Die by Suicide." BBC Future, BBC, www.bbc.com/future/article/20190313-why-more-men-kill-themselves-than-women.

Pike, John. "Military." Albanian Civil War (1997), www.globalsecurity.org/military/world/war/albania.htm.

Blasco-Fontecilla, Hilario, et al. “Worldwide Impact of Economic Cycles on Suicide Trends over 3 Decades: Differences According to Level of Development. A Mixed Effect Model Study.” BMJ Open, British Medical Journal Publishing Group, 1 Jan. 2012, bmjopen.bmj.com/content/2/3/e000785.

Appendix

GitHub Link: