# Analysis of Global Suicide Trends 2000-2010

Stephen Babcock [stbabcoc@syr.edu]

Sarah Bolton [sbolton@syr.edu]

Kekona Soon [klsoon@syr.edu]

Bibhuti Timalsina [bitimals@syr.edu]

Paul Tuck [patuck@syr.edu]

### Table of Contents

1. [Project Abstract](#_avc5bsiwrje3)
2. [Contribution Statement](#_57ua8j8t3uqe)
3. [Final Report](#_g6id5r52n9xy)
   1. [Introduction](#_da5yv2jg2rx1)
   2. [The Dataset](#_se9xcxmvakhj)
   3. [Methodology](#_bkvyqbe6mg90)
   4. [Analysis](#_n4xlewhnel30)
   5. [Results](#_2g7vp4tc6ijw)
4. [Conclusions](#_tb4hsi8a2abh)
5. [References](#_uzhu5z6y25mv)
6. [Appendix](#_hn5qmg7z1mfu)

### 

### Project Abstract

### This project makes use of a data set compiled by the World Bank. The World Bank compiled four data sets into the data set we used. Thes combined data includes world suicide statistics from 1985 through 2016, GDP (in US$), and Human Development Index(HDI) statistics. We were unable to use the HDI due to how incomplete this part of the data set is.

Our group looked at the time frame 2000 through 2010 because this period had the most complete data across the most countries. The intent of the analysis is to identify which groups are most at risk, if any groups are in urgent need of improvement, and how suicide rates are trending in general. We are also interested in identifying if economic factors (measured by GDP) play a role in suicide rates in this period. Our hypothesis is that GDP does play a role, but it may be based on changes in GDP, or based on how well developed the country is.

We used “ggplot” to look at trends across gender, age groups, and generations to identify any risk groups. We then used linear modeling to analyze the impact of the economy on suicide rates. The linear models were mostly flat and indicated no correlation between the GDP or GDP per capita and suicide rates.

### Contribution Statements

**Stephen Babcock:** Developed the data set to look at trends and annual change in GDP and suicide rates. Applied linear models and built visual representations of these data points. Created conclusions and takeaway slides after entirety of team had given results of their respective tasks.

**Sarah Bolton**: Analyzed changes rates, in regards to gender, annually to help answer one of our data questions through data modeling and explained the analysis summary. Created the information on PowerPoint giving background of the IASP. Pitched how our dataset related to the stakeholders for the presentation.

**Kekona Soon**: Proposed overarching research question that we investigated and found organization that could benefit from our project. Analyzed data to find which year range had the most data for all countries across variables to aid in giving optimal conclusions. Developed Linear Regression Modes to see correlation between GDP/GDP per capita and Suicides and Suicide Rates. Investigated differences in “developed” and “developing” countries suicide rates.

**Bibhuti Timalsina**: With the help of “dplyr”, grouped data to specific requirements. Utilized “ggplot” for visualizations and explained data analysis method. Presented introduction for the International Association of Suicide Prevention.

**Paul Tuck**: Provided means and standard deviations for the number of suicides and suicides per 100k for each country. Conducted exploratory analysis throughout project. Compared the mean suicides for each age group in each country with the 95th percentile of the average mean for the world.

### Introduction

The onset of this project was motivated to see if the data supported our initial hypothesis that lower performing countries, in respect to GDP and GDP per Capita, were more likely to have an increase in suicide rates versus high performing countries. We believed that if we saw strong correlation between the two, we would be able to share this with the IASP to help them cater their involvement in suicide prevention where it was most needed. Furthermore, we understood that this was the overarching question and would additionally require looking at trends to give an optimal solution. Therefore, we analyzed if periods of economic fluctuations in the aforementioned metrics were related to fluctuations of suicide rates. This question could further allow the IASP to create custom action plans when seeing certain economic upturns or downturns in a respective country. For example, if the data showed that decreases in GDP per capita were bringing increased suicide rates, then we could recommend to the IASP that countries expecting economic downturns will need to have resources in place before to try to minimize suicide rates rising. Furthermore, we understood that it was possible that our initial approach could indicate weak predictive power so we also investigated age and gender as critical factors in predicting suicide rates across cultures.

Applying Liner Regression models to our first question we found:

1. GDP and GDP per capita, when viewed independently, had very slight predictive power.
2. Higher GDP countries do have a larger average of total suicides. However, we also saw that the number could be misleading due to population also having correlation to average # of suicides.

Using ggplot visualizations:

1. Observed oldest male group has the highest suicide rate across countries.
2. Changes in GDP and GDP per capita did not have much influence on the change in suicides and suicide rates.
3. The Republic of Korea began as a country of concern, and was identified as the leader in having a positive trend in suicide rates following the analysis of the time period.

Recommendations:

1. Based our findings, we recommend to not use changes in GDP or GDP per capita as being indicative of a countries suicide rate.
2. Focus efforts on Republic of Korea, Luxembourg, Albania, and Aruba as these countries have an increasing rate of suicides.
3. Conduct further studies that involve economic metrics such as Human Development Index to see effect on suicide rates. This metric encapsulates a more clear picture of how a countries economic position affects the population and takes more things into account than GDP does.

### The Dataset

The data set compares suicide rates to multiple variables. The data is tracked by country per year. There are 101 unique countries and 32 years (1985-2016). Within each country, there are various variables considered: sex (male vs female), six age groups, and six generational groups. The data set was downloaded from: [Suicide Rates Overview 1985 to 2016 | Kaggle](https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016)

### Data Analysis Methods

First, we did a basic summary of the data, calculating the mean and standard deviations for the number of suicides and suicides per 100k for each country.

One of the ways we compared the suicide numbers for the age groups is to compare the means of each age group and each country with the yearly mean for the entire world. To do this, we created a sample mean distribution by generating 10,000 sample means of yearly suicides of 1,000 observations from the original dataset (with replacement). The means for each age group in each country were then compared to the sample means to see if the age group means fell within the 95th percentile of the sample means. If they did, they were considered high risk. This was done for both the suicide number and suicides per 100k. An age group for a country was considered at highest risk if the average suicide number and suicides per 100k were above the 95th percentile for the sample means.

The data set we downloaded was from Kaggle on Suicide Rates Overview 1985 to 2016. The first step we performed was to clean the data. Since there were a lot of data present, we wanted to include the recent statistics, we filtered out data from year 2000-2010 covering all 101 countries. After the filtering of data was performed, we looked at the data set again and figured, we could get rid of HDI column as that column made no difference to the research we were working on. Later, descriptive statistics was performed to calculate the mean and standard deviations for the number of suicides, suicides per 100K and GDP for each country.

We also used library functions like “sqldf” and “dplyr” to group the data set based on countries to find out the top 10 and bottom 10 countries for suicides happening per 100K population. Also, the data set were grouped based on year, age and gender to figure out what age group of population and what gender were committing more suicide so that data could be presented to our audience to control the rate of suicide in certain countries and provide awareness in that area. After the descriptive statistics were performed, the next step was visualization. They were performed using the library function ”ggplot” and data we cleaned from our previous step so that the people viewing our data would be more clear looking at it visually rather than just seeing the numbers. With the cleaning of data and building visualizations, we saw that older age group male were at higher risk as they were the ones attempting suicides. The final step we performed was the modeling portion where we could predict and conclude what could be the main reason for suicide rates to climb up high. Linear model was done and with this we concluded that GDP was not one of the factor the people were attempting suicides. We built a total of three linear models: two of the models had suicides per 100k as the dependent variable, with both gdp and gdp per capita as the independent variables; and the third model had total suicides as the dependent variable and total gdp as the independent variable. The quality of these models were determined through the adjusted R squared values. We also used sen’s slop to calculate at what rate aspects were changing.

### Analysis Summary

Figures 1 and 2 show the average yearly number of suicides and suicides per 100 for each country analyzed, respectively. In terms of raw numbers, the Russian Federation, United States, and Japan had the largest average number of suicides per year. However, when looking at suicides per 100k, the Republic of Korea (i.e. South Korea) and The Russian Federation were the highest. This seems to indicate that the people of the Russian Federation are at particularly high risk of suicide.

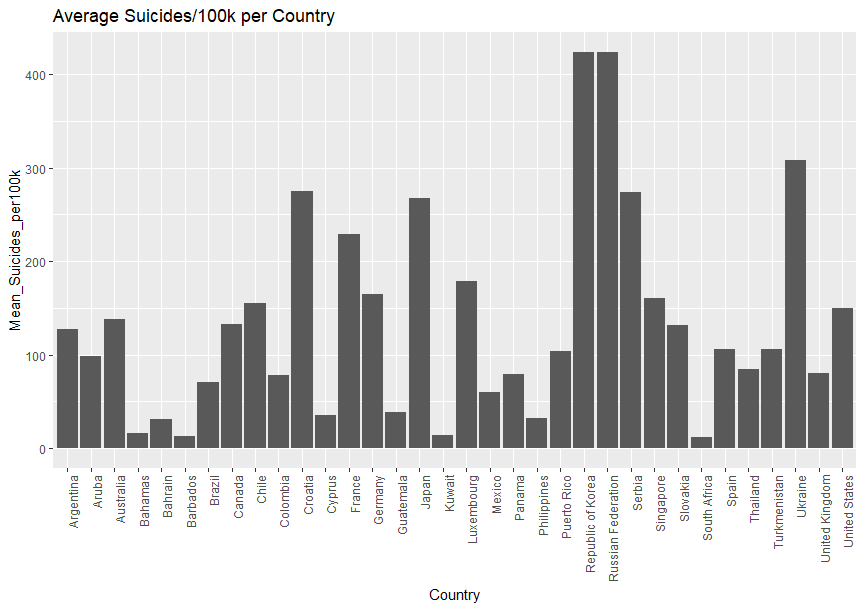
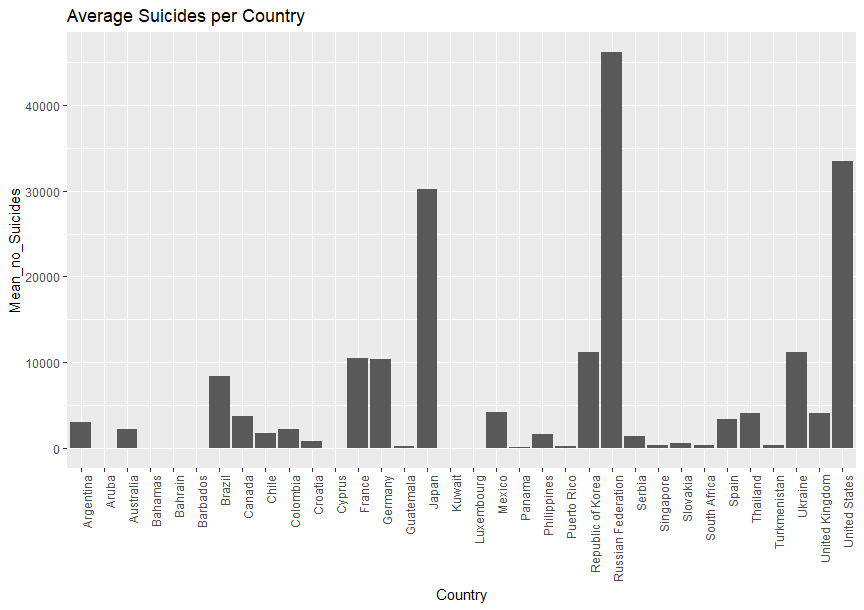


Figure 1 Figure 2

The question we wanted to answer from our dataset was does the GDP per capita effect suicide rates, and if so, to what extent? We began by looking at the various factors of the data including the different vectors of information we had such as countries, suicide rates, and suicide rates per 100k of the population, and started with exploratory analysis of the whole dataset. This helped us understand our data and from there pinpoint which areas were worth exploring. For instance, from our exploratory data and analysis we were able to discover that for many countries, there was not data for all years and many were missing which helped us narrow our dataset down to the years 2000-2010, which made the data more contemporary as well as more evenly distributed by country. Figure 3 shows the trends in average suicide rates per 100k of the population for Males versus Females, while figure 4 shows the same for various age groups. Both show us that suicide rates were decreasing annually across the board between 2000 and 2010.

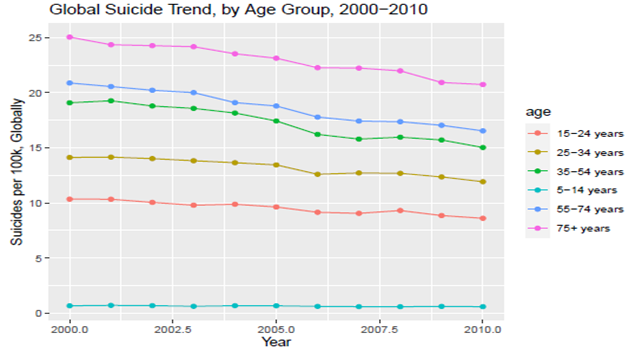
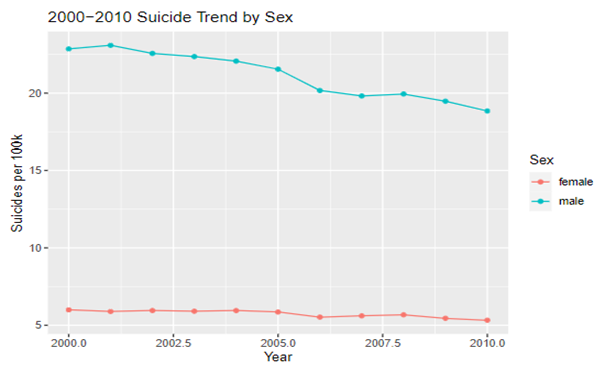


Figure 3 Figure 4

Once we finished with the exploratory data, we began examining what effect different sections of the population have on suicide rates. Our data contained information on different age ranges as well as genders for each country for each year they had data on. Our analysis showed that between males and females, males were committing suicide at a higher rate, and of the age groups the age range 35-54 were most at risk. Also through our analysis, it was determined that there is no clear correlation or way to project these outcomes from the dataset we were utilizing in a significant way.

Figures 5 demonstrates the average rate of change in suicide rates from 2000-2010 where the Male rate of change is on top and the Female rate of change is on the bottom. Figure 6 shows the rate of change for both genders averaged together from 2000-2010, showing there is no significant slope or pattern for Gender as a factor as a whole. This negative slope for each gender supports our original findings that suicide rates for males and females are going down, and we can see that men have a steeper rate of change than females which is more gradual. Figure 7 shows each age group’s average yearly numbers of suicides and suicides per 100k that are greater than the 95 percentile of the world average, grouped by country. For all countries, those aged 35-54 were more likely to commit suicide than any other age groups for both suicide numbers and per 100k of the population, however when only looking at per 100k in our exploratory data, those over 75 were more at risk of being likely to commit suicide.

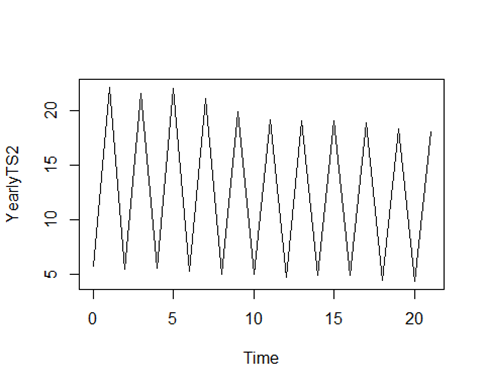
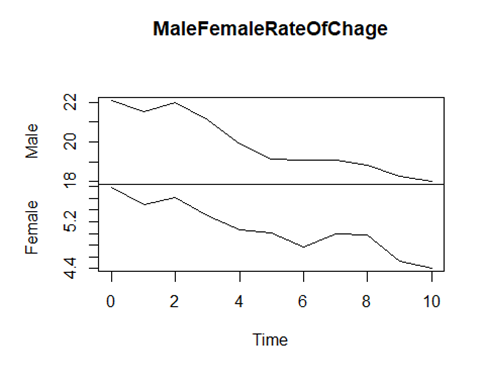


Figure 5 Figure 6

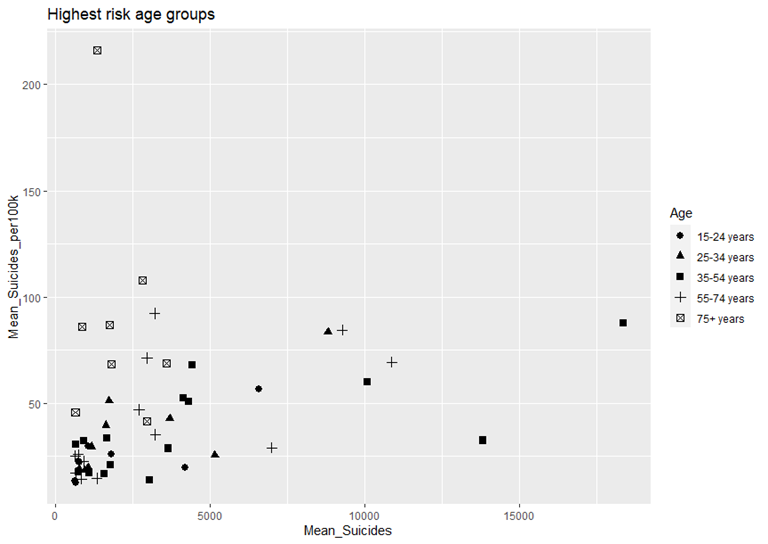


Figure 7

Next, we looked at the economic impact of the GDP on suicide rates. This is broken down into 3 questions—do economics effect suicide rates? What is the difference between the most and least developed countries and the economic impact on suicide rates for these? Do changes in economic production influence suicide rates?

Figures 8-10 show that there is not a real correlation between the GDP or suicide rates through linear data modeling. Figure 8 represents the correlation of the average GDP per capita on suicide rates per 100k of the population. Since this data is scattered randomly (R-squared = -0.01773661), we conclude for these variables that suicide rates are not well explained by GDP per capita. Figure 9 represents the correlation of the average total GDP on the average suicide rate. This model is slightly more predictable but still has minimal effect of the GDP on suicide rates, though the positive relation does exist (R-squared = 0.05923996). Finally, figure 10 shows the correlation of average total GDP on the average number of suicides, which shows a possible connection of higher GDP countries being connected to more suicides on average (R-squared = 0.6661958). This could be explained by a number of factors, but we acknowledge that the higher GDP populations are more likely to have a higher population and therefore higher changes of thier population committing suicide. Overall our findings conclude there is not a clear and relevant link between the GDP and suicide rates right out, however we conducted further analysis to see if looking at the data through a different lens helped.

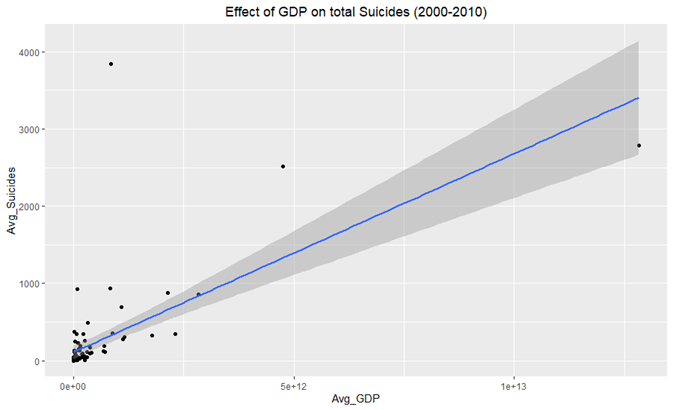
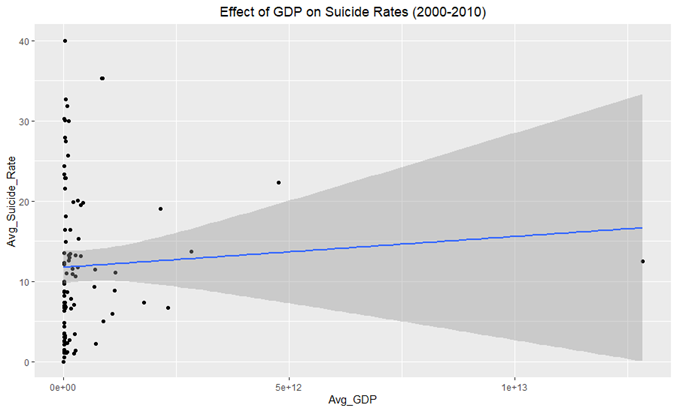
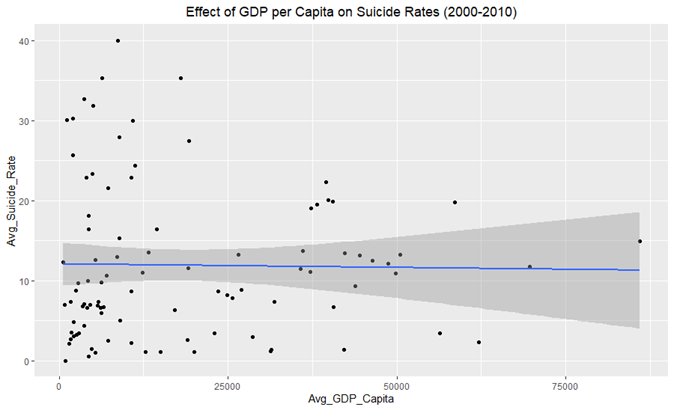
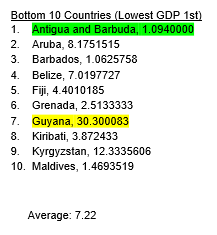
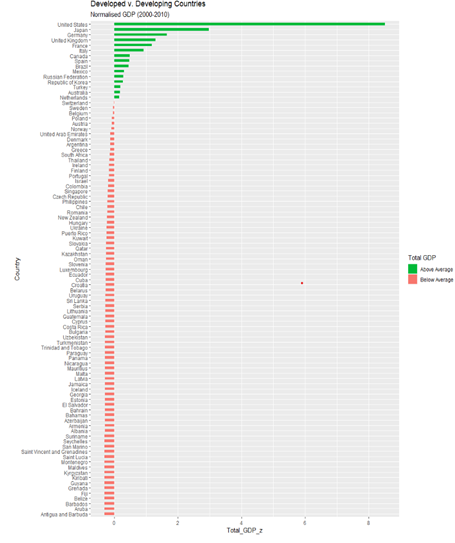


Figure 8 Figure 9 Figure 10

The next question asked is, is there a difference between developed countries and developing countries and suicide rates? For this, we had to create our own definition of developed vs developing countries as some of the countries did not have sufficient data. For this, our developed countries were the ones with the highest GDP and countries that were only -.2 away from a zero Z-score within the dataset. Figure 11 shows countries as developed as opposed to developing through the standardization of the GDP average over 10 years within the dataset. The results of this show that there is not a real correlation between developing countries versus developed ones. We also did consider that given how inconsistent some of the data was especially for lower GDP countries, it may be likely there are more similarities or even higher suicide rates that were simply not in our dataset.



### Figure 11 Figure 12 Figure 13

For our last question, we wanted to see if a change in GDP could predict a change in suicide rate. Each country’s data was calculated to provide the change in suicide per 100k population, and the change in GDP per capita The first step measured the trend in change by finding the mean in both metrics. Figure 14 shows a linear model derived from the mean of changes in suicide as a function of the mean of annual change in GDP per capita for each country. The abline formed by this linear model is nearly flat with an adjusted R-square of –0.01. We then looked at the individual years. Figure 15 demonstrates an abline for the linear model based on the annual change in suicide rate as a function of the annual change in GDP per capita. This is another very flat abline; the linear model provides an adjusted R-square value near zero. Figure 16 is a zoomed in version of Figure 15, removing outliers to get a better look at the majority of the data points. In this graph, suicide rate is on the y-axis; above the x-axis is an increase in suicide rate and bellow it is a decrease in suicide rate. The x-axis is GDP per capita; to the right of the Y-axis demonstrates an increase in GDP per capita, while a decrease is shown to the left of the Y-axis. In Figure 16, if there were a correlation between suicide rate and GDP per capita, we would see a noticeable diagonal line. Instead, we see a flat line showing that suicide has been decreasing agnostic to changes in GDP. As such, we find that, with the data from our dataset, there is no clear indication that GDP has an effect on suicide rates.

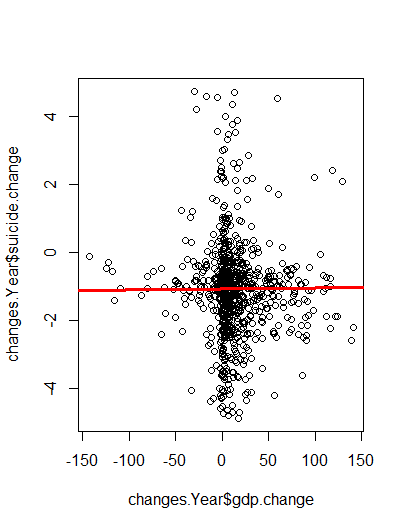
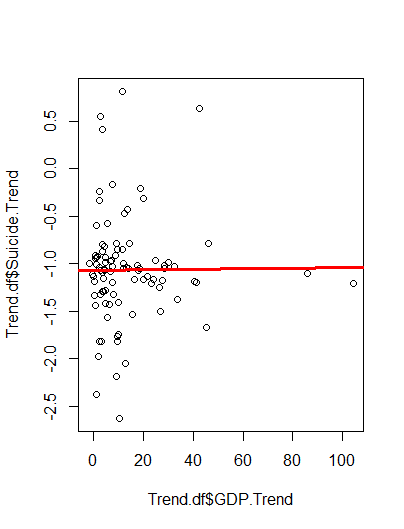


Figure 14 Figure 15

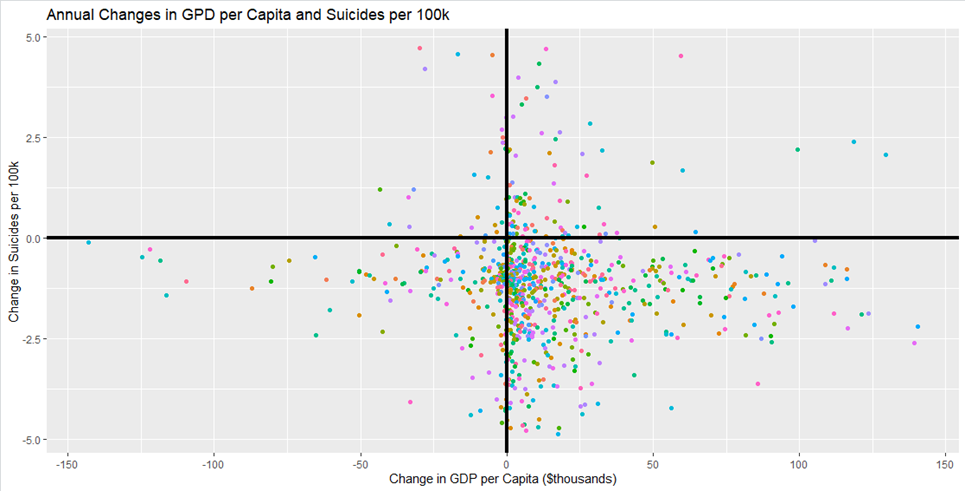


Figure 16

### Conclusions

In conclusion, we initially set out in this project to better understand how GDP and GDP per capita affected countries suicide rates. Additionally, we set out to understand what other people groups may be at risk (e.g. older/younger or male/female) for increased suicides. This was all conducted in an effort to aid the International Association of Suicide Prevention to understand which countries might require more resources to decrease suicides. Our utilization of linear modeling showed that the direct correlation of GDP and GDP per Capita on Suicide Rates is miniscule at best. Furthermore, our investigation in to the change of GDP/GDP per capita year over year displayed a small effect on suicide rates.

From our analysis we recommend that the IASP not consider GDP or GDP per capita, in isolation, when considering which countries are at higher risk for suicides. We believe this is an important conclusion to highlight because it can dissuade readers of a common notion that just because a country is “developing” that they must be committing suicides at an increased rate. We also implore the IASP to work with country’s local health authorities in some “developing” countries where data appeared inconsistent. An increase in data collection efforts have the possibility for this same type make of analysis to make conclusions with higher certainty.

We also identified some countries that are trending in a positive manner when it came to suicide rates. Republic of Korea, Luxembourg, Albania, and Aruba should be countries that the IASP gives significant consideration towards in attempt to reverse trend. The data also indicated that the age/sex combination of 55+ males being the highest risk group for suicides globally. Due to this, we would recommend the IASP make these type of statistics known on their website and provide generic mental health information tailored towards this demographic.

### References

The World Bank. (2018, December 01). *Suicide Rates Overview 1985 to 2016*. Retrieved from www.kaggle.com: <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/metadata>

### Appendix

You should display the appropriate code here. If you have placed everything in a markdown file please submit it with the final report and not it here. If you have a code repository link to it here.