Data Analysis Project 3

SUICIDE RATES VS GDP

Introduction

In social and economic studies, the association between suicide rates and GDP (gross domestic product) of a nation is an important topic of interest and concern. The goal of this project is to investigate and analyze the relationship between suicide rates and GDP. Data analysis and statistical methods will be used in this project to investigate the relationship between GDP and suicide rates. We can find potential correlations and evaluate the statistical significance of these interactions by utilizing the data collected from the dataset chosen.

Project Two Development

We made a few changes from project two to further enhance this project:

- First, we changed our hypothesis regarding suicide rates to gain more benefit from the dataset we chose.
- Then, we decided to use a column, from our dataset, that we decided not to use in our previous project and formulated our project idea around this column.
- Moreover, we obviously changed the hypothesis test to better analyze our new hypothesis.
- Finally, we changed our visualizations to better illustrate our hypothesis.

Research Question

Is there a relationship between GDP (Gross Domestic Product) and suicide rates?

Hypothesis

Null Hypothesis: There is no relationship between GDP and suicide rates.
Alternative Hypothesis: There is a negative (opposite) relationship between GDP and suicide rates.

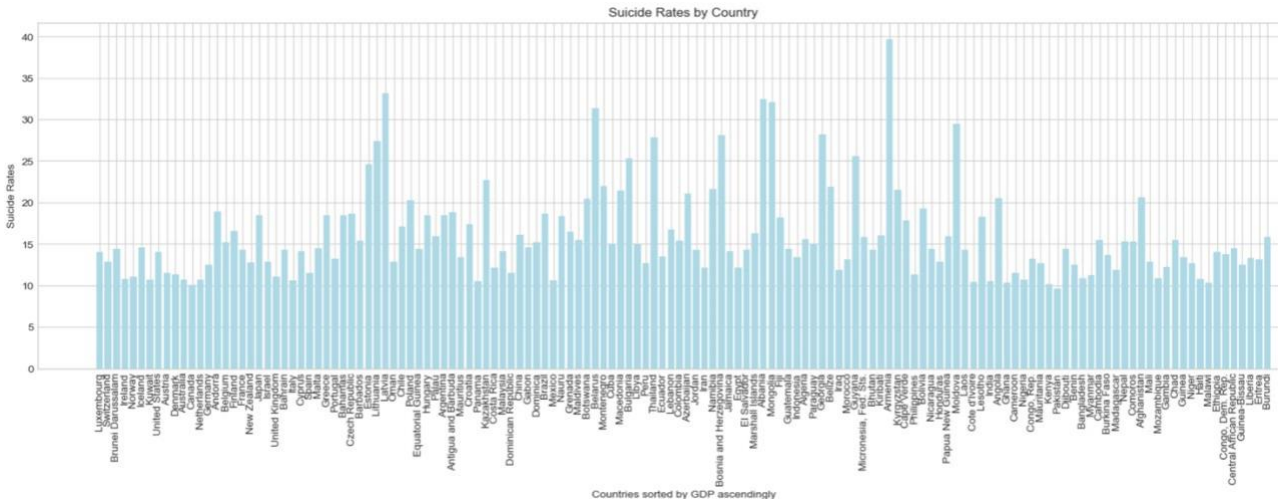
Collected Data / Dataset

The dataset contains three columns: Country, GDP per capita, Suicide rate, and 141 entries(rows)

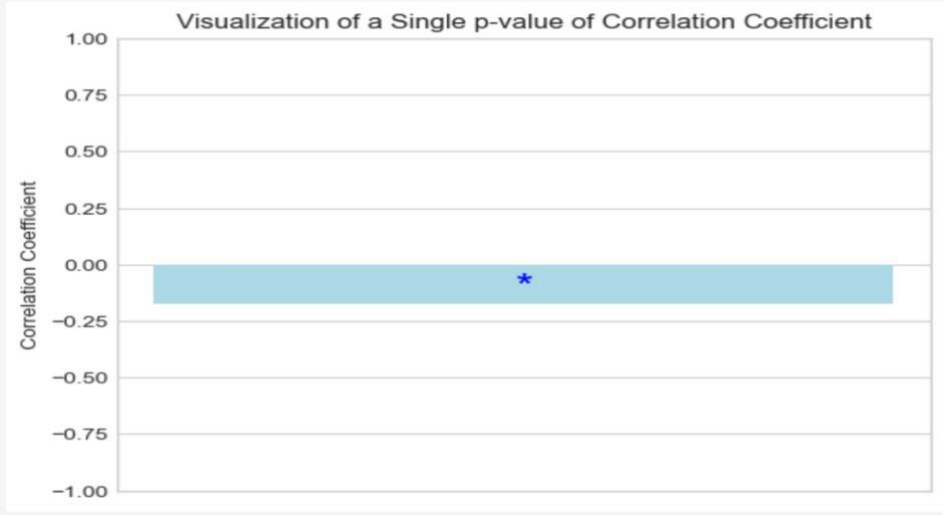
- Country: this column lists the countries from 1 to 141
- GDP: this column contains the gross domestic product of every country in the dataset
- Suicide rate: this column contains the suicide rates of every country in the dataset

Analysis

- First, we constructed a bar chart that visualizes the relationship between countries and their suicide rates. The x-axis represents the countries sorted by their GDP in descending order (from left to right) while the y-axis represents the suicide rates. The higher the bar the higher the suicide rate for that particular country. As the data tends to the right, the GDP increases, the suicide rates seem to be less than the left side of the graph, which supports our hypothesis.

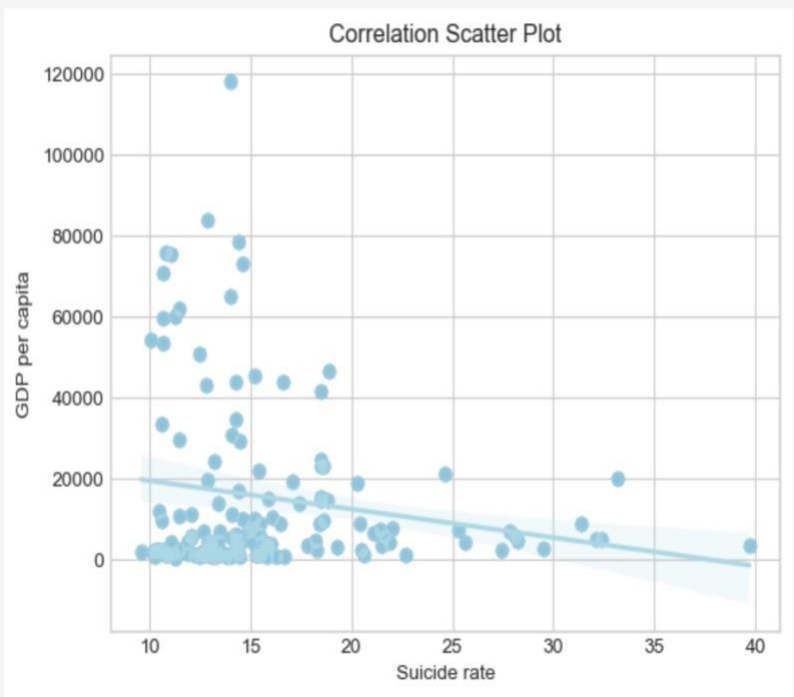


- Then, we plotted the p-value of the correlation coefficient using a bar plot, which was smaller than the alpha value (0.05). The bar represents the correlation coefficient, and the blue asterisk is only displayed if the p-value is less than the alpha value, which is true in our case.



- Lastly, we constructed a correlation scatter plot with a line of regression to visualize the relationship between GDP and suicide rates. The x-axis represents the suicide rates while the y-axis represents the GDP. As shown the data follows a downward trend from left to

right indicating a negative correlation. The line of best fit shows a negative slope which means that one variable tends to decrease as the other variable increases



Hypothesis Testing Steps

We chose correlation test as our hypothesis testing method

- **Step 1:** first, we formulated the null and alternative hypothesis (written above)
- **Step2:** then, we assumed our alpha value (significance level) would be the standard value, which is 0.05
- **Step 3:** next, we chose our dataset from Kaggle. We cleaned the data and added a few things to the dataset to ease some process while coding
- **Step 4:** last but not least, we calculated the correlation coefficient and its corresponding p-value using '**scipy.stats.pearsonr()'**
- **Step 5:** finally, we compared the p-value we generated to the standard alpha value and made notes from the correlation coefficient generated, then we conduct our conclusion from here

Conclusion

This project aims to find if there is a relation between GDP (Gross Domestic Product) and suicide rates. We tested our hypothesis using a correlation test because we have two numeric variables, GDP and suicide rates. After conducting the correlation test and analyzing the results, we generated the p-value (0.03698850127461523) and the correlation coefficient (-0.17648598878252225). For the p-value, we concluded that it's best that we reject the null hypothesis, that there is no relation between GDP and suicide rates, due to the generated pvalue being lower than the alpha value (0.05). Which also meant that we should accept our

alternative hypothesis, that there is an opposite relation between GDP and suicide rates. As for the correlation coefficient we concluded that there is a weak negative correlation between GDP and suicide rates, which supports the conclusion from the p-value. Finally, after analyzing our conclusion we could say that there is a negative (opposite) relation between the GDP and suicide rates.