PROJECT PROPOSAL- GROUP 2
BUAN 6356 - BA With R

Aditi Dutt, Avijeet Bose, Christian Candler, Sangeeta Singh

INTRODUCTION

Generally, people decide to attempt suicide impulsively rather than planning it out extensively. There can be many factors that can influence a person to commit suicide, and one of them is depression.

According to the World Health Organization (WHO), one of the significant health issues worldwide is a suicide, a leading cause of death. Close to **800,000 people** die every year due to suicide, which is one person every **40** seconds. However, suicides can be prevented when adequate measures are implemented at the individual, municipal, and national level. Suicide is a global phenomenon that occurs in all regions of the world.**The focus here is to identify the factors which can help to predict the suicide rate.**

NULL HYPOTHESIS

The project proposal's goal is to observe effects of various factors like year, age groups, countries, gender, generations, population, and GDP on the count of suicides from 1985 to 2016. The data consists of the countries around the world. The population of interest are the people from ages 5 and above who have committed suicide.

The report also measures the trends in suicides in different age groups, generations and countries.

The null hypothesis is:

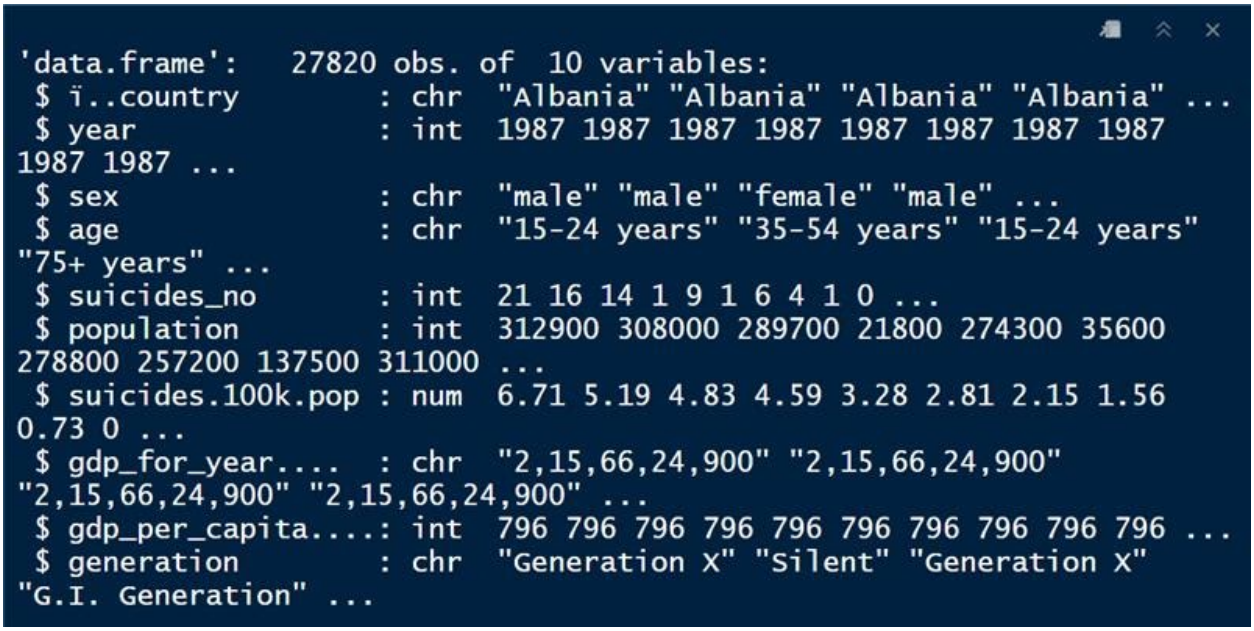
H0: There is no impact of factors like year, age groups, countries, gender, generations, population, and GDP on the suicide count.

DATA DESCRIPTION

The data to be used in this project was collected by extracting information from an online database (Kaggle.com).The majority of the data used in this analysis was obtained from the World Health Organisation.

The dataset contains 27,820 observations recorded between 1985-2016. There are 10 variables (stated below).

The target variable is “suicides_no” which is to be predicted by country, year, sex, age group, population, gdpforyear, gdppercapita, generation (based on age grouping average).

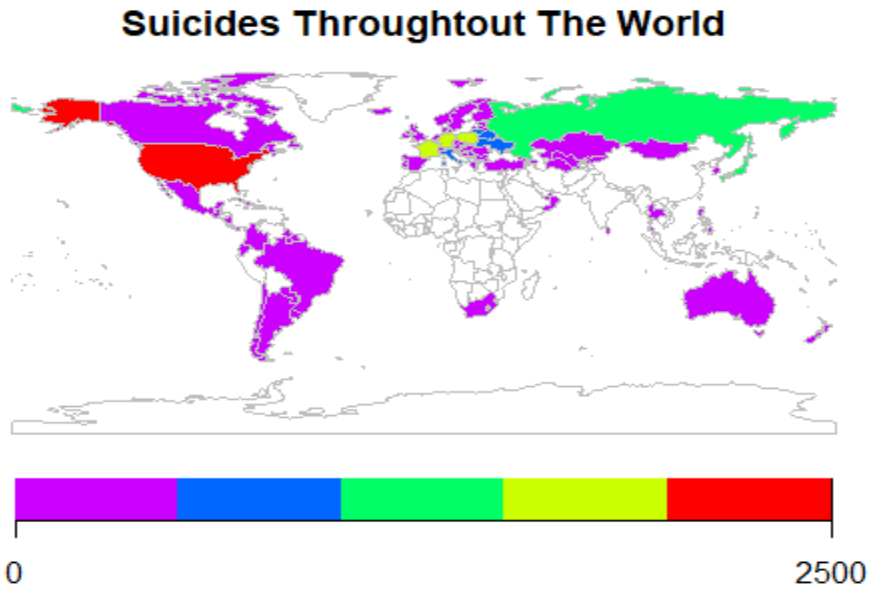


```
'data.frame':   27820 obs. of  10 variables:
 $ ï..country        : chr  "Albania" "Albania" "Albania" "Albania" ...
 $ year              : int  1987 1987 1987 1987 1987 1987 1987 1987
1987 1987 ...
 $ sex               : chr  "male" "male" "female" "male" ...
 $ age               : chr  "15-24 years" "35-54 years" "15-24 years"
"75+ years" ...
 $ suicides_no       : int  21 16 14 1 9 1 6 4 1 0 ...
 $ population        : int  312900 308000 289700 21800 274300 35600
278800 257200 137500 311000 ...
 $ suicides.100k.pop : num  6.71 5.19 4.83 4.59 3.28 2.81 2.15 1.56
0.73 0 ...
 $ gdp_for_year....  : chr  "2,15,66,24,900" "2,15,66,24,900"
"2,15,66,24,900" "2,15,66,24,900" ...
 $ gdp_per_capita....: int  796 796 796 796 796 796 796 796 796 796 ...
 $ generation        : chr  "Generation X" "Silent" "Generation X"
"G.I. Generation" ...
```

EXPLORATORY DATA ANALYSIS

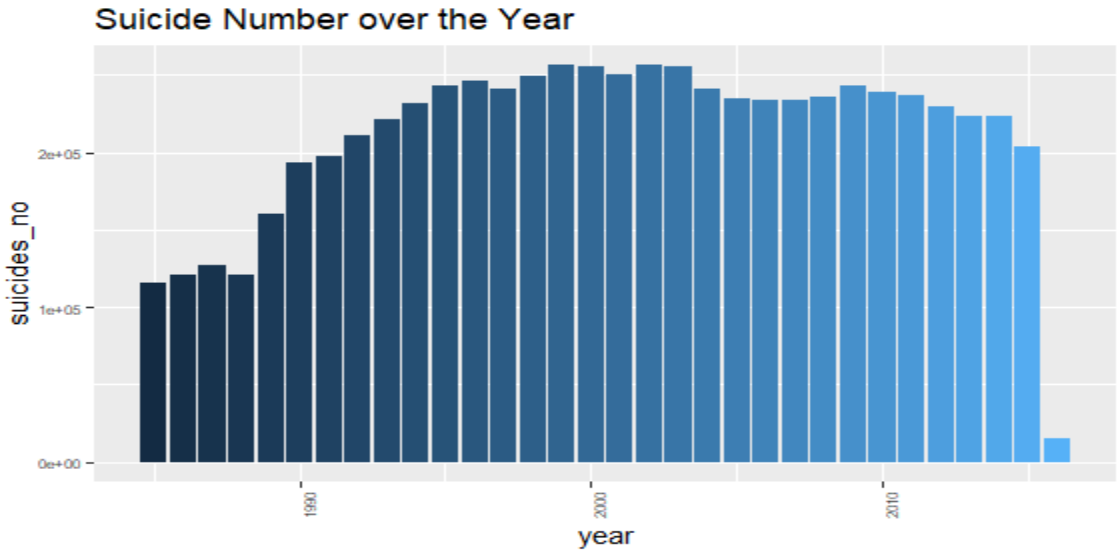
Number of suicides vs. Country

- The below graph shows the number of suicides across the world between 1985-2016.



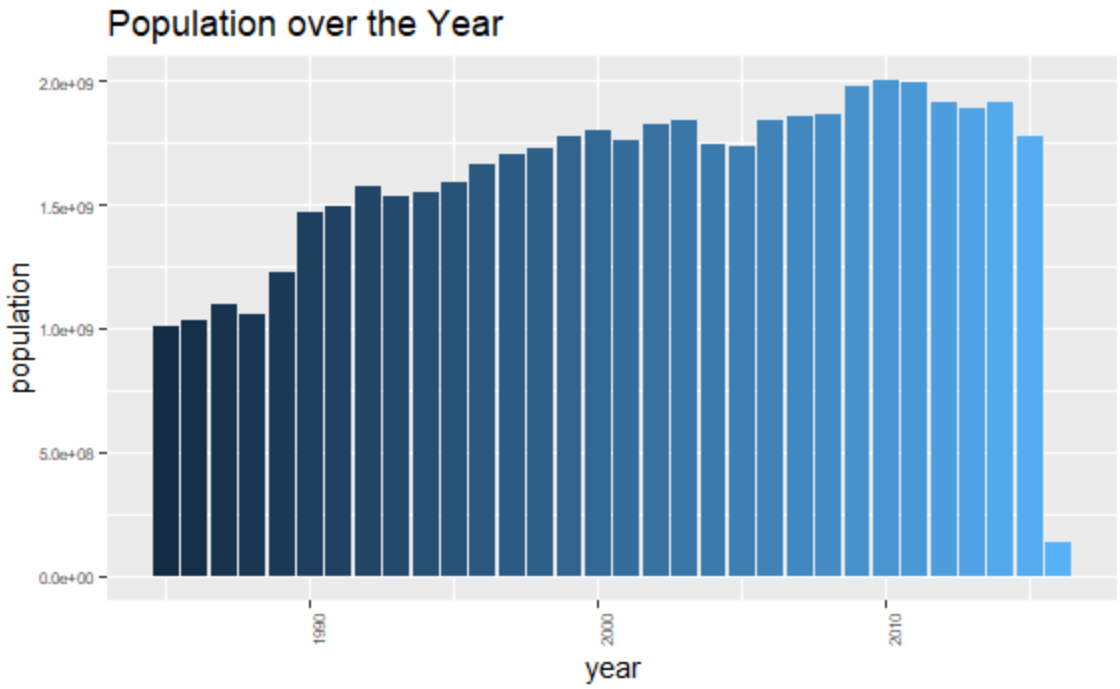
Number of suicides overall from 1985-2016

- The below bar chart shows the trend in the suicide rate from 1985-2016.



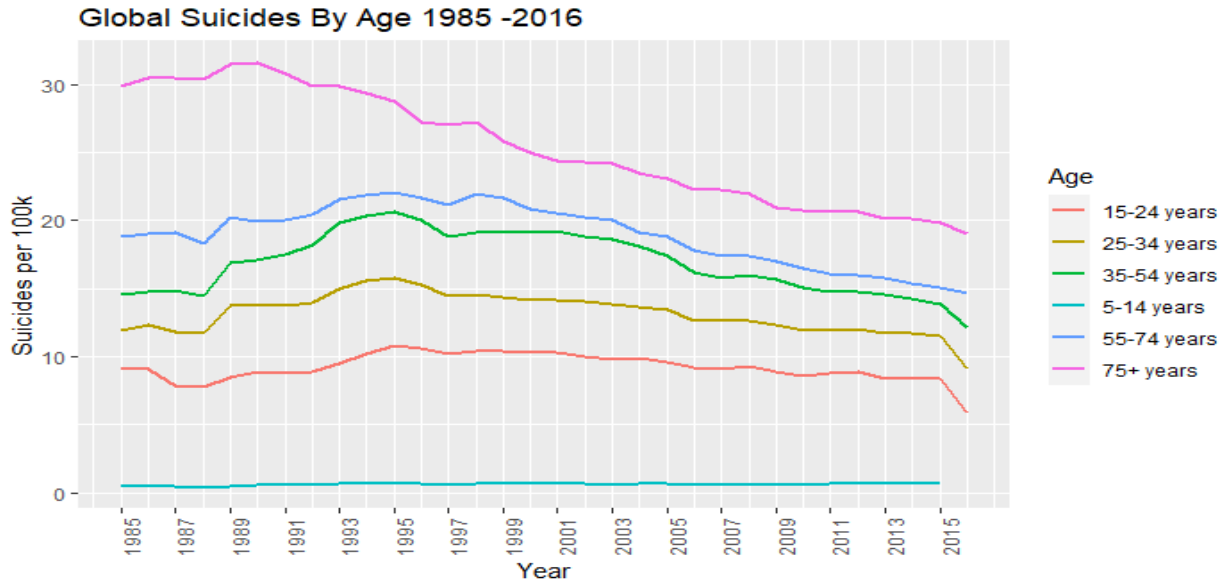
Population Trend between 1985-2016

- The below bar chart shows the population trend between 1985-2016. The graph helps in establishing the relationship between the number of suicides and the population.



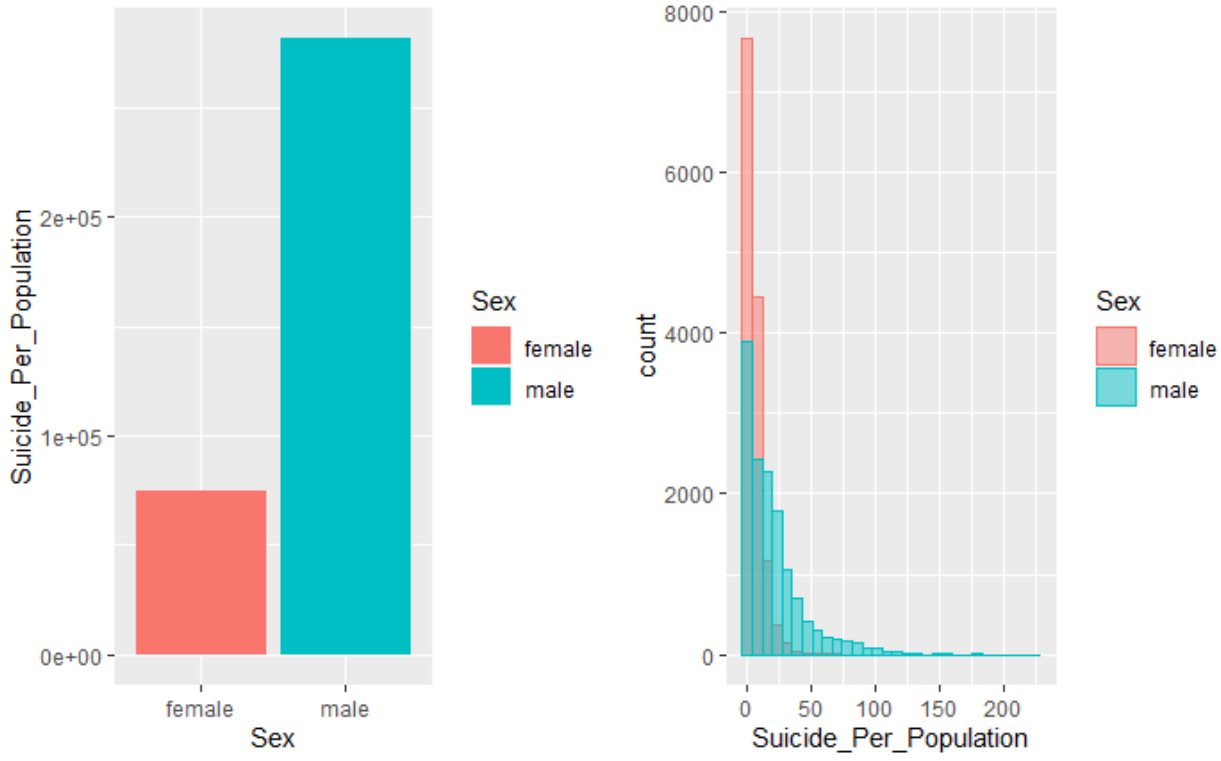
Number of suicides over the years based on Age Groups

- The below graph shows the trend in the suicide rate among different age groups over the years.



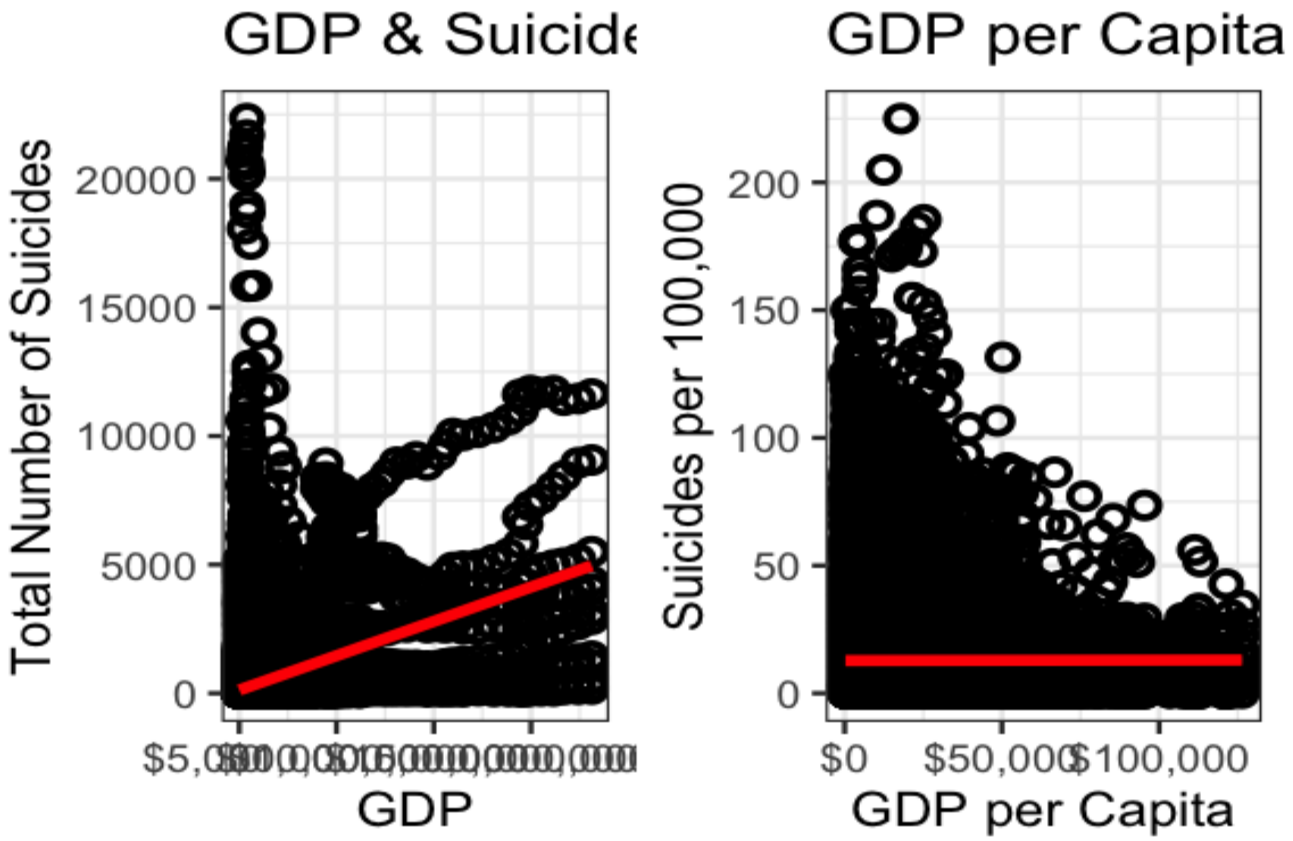
Number of Suicides based on sex

- The below graph shows the trend in the number of suicides among males and females



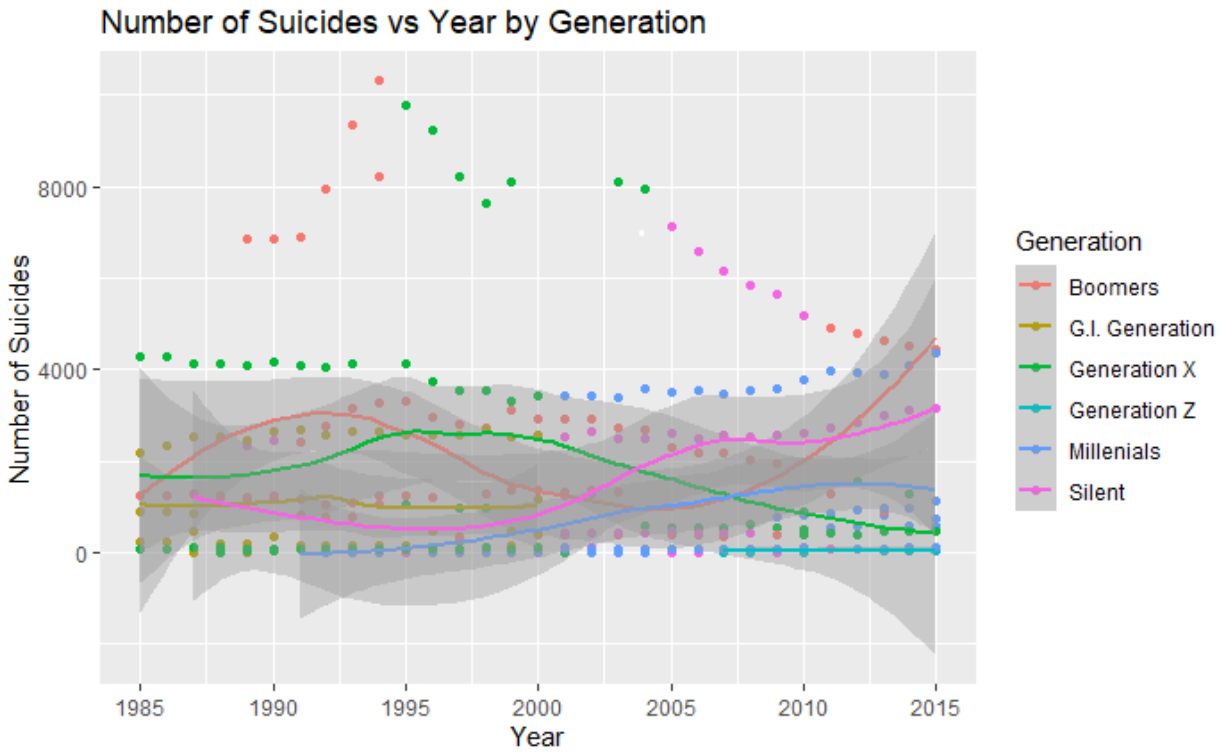
Number of Suicides based on GDP

- The below chart shows the correlation between the number of suicides and the GDP.



Number of Suicides based on different generations over the Years

- The below graph shows the trend in the number of suicides for different generations over the years.



PRELIMINARY ANALYSIS

1. Referencing the “Number of suicides vs. Country” graph in the exploratory data analysis, we can see that the United States has the maximum number of Suicides of the years (1985 - 2016). Some european countries follow close behind.
2. Referencing the “Number of suicides overall from 1985-2016” graph in the exploratory data analysis, it is seen that the number of suicides have steadily increased from 1980’s to 2000’s and then became somewhat stagnant.
3. Referencing the “Number of suicides over the years based on Age Groups” graph in the exploratory data analysis, as the years go by the number of suicides among different age groups have declined. It is also apparent that, among the age groups, the 75+ group has the highest number of suicides.

4. Referencing the “Number of suicides based on Sex” graph in the exploratory data analysis, the proportion of number of suicides is significantly higher in males compared to females.
5. Referencing the “Number of suicides based on GDP” graph in the exploratory data analysis, the relationship between the GDP and the number of suicides is dependent, while there is no correlation between suicide count and GDP per capita. This difference could be due to the GDP increasing as the population increases. Thus this could explain the correlation.
6. Referencing the “Number of suicides based on different generations over the years” graph in the exploratory data analysis, no trend is seen between the different generations. However, outliers from each generation can be observed.

CHALLENGES

1. The first challenge when dealing with the dataset was coming across the columns with null values. To fix this, these columns were removed from the analysis.
2. Some columns were redundant, and therefore unnecessary for the analysis. This could have led to problems when performing tests on the data. These columns were also removed from the dataset.

SUPPLEMENTARY INFORMATION

1. To predict the number of suicide counts based on the factors in the dataset, a linear regression model will be built on the data.
2. An executive summary will be written to give an overview of the project and its intention.
3. A conclusion will be written to sum up the project and provide information on the null hypothesis and whether or not to reject it.