|  |  |
| --- | --- |
| Suicide Rate Overview from 2001 to 2016  Summary report | **Author:**  Amine Jalali  432733  **Course:**  Data Science for Life Science, Hanze University of Applied Science  **Module:**  Programming 1    **Module coordinators:**  Fenna Feenstra |
| **Published Date:**  2/5/2022 |  |

# Introduction

The worldwide suicide rate has been quite high for many years, around 700,000 people take their own lives annually. There is no question that this tragedy (being higher than the mortality caused by war and homicides) not only affects families but also affects the entire society. Undoubtedly, suicide does not belong only to a certain segment of society or a particular country, but it can be said that living conditions affect its rate.

This report is a summary of data analysis of the suicide rate based on some well-known factors. the key factors will be determined by statistical test. Finally, for giving more clear idea of what the information means,  data is visualized. Data is taken from Kaggle website[[1]](#footnote-1) applied to analyse the suicide rate between 2001 and 2016. The first (“master.csv”) involves examining the suicide rate based on data such as country, year, gender, age, number of suicide per population, HDI index, GDP and GDP per capita, and the second ("concap.csv") file includes geographical information of countries. It should be noted that some unnecessary columns such as population and generation (that depends on the birth year) were dropped.

Research questions:

* Does gender affects suicide rate?
* Is there a differences in suicides caused by a specific age group?
* Is there a differences in suicides caused by specific continent?
* Is there any correlation between gdp and suicides?

# Procedures

Taken processes are summarized as below:

1. Creating a new data frame based on year, sex, and age for each country;
2. Merging data frame with main data frame and filled nan values:
   1. There was no data for suicides for specific years or ages, assuming that there were not noticeable suicides and filled it with zero;
   2. Filling GDP in two ways: using data from the other rows in the same year or using the average of GDP of that country;
3. Changing the type of 'country', 'year', 'age', and 'gender' to category;
4. Preparing another data frame from "concap.csv";
5. Merging theses two data frames.
6. Mapping gdp to gdp\_level according to the worldbank[[2]](#footnote-2) website.

# Data inspection

|  |  |  |
| --- | --- | --- |
| Figure 1 - Suicide rates are higher in men than women. | Figure 2 - As people get older, the suicide rate increases. | Figure 3 - With increasing GDP, the suicide rate increases. |
| Figure 4 - Suicide do not have a definite trend 2002- 2016. | Figure 5 - Suicide rates vary widely on some continents. | Figure 6 - As men and women age, suicide rates increase. |
|  | Figure 7- Percentage of correlation:   * The suicide rate increases by 30% with age * Suicide rate are 35% higher in men than women * Suicide rate decreased 11% over years. * it does not show any correlaion between gdp\_level and suicide, only 2% * it does not show any correlaion between gdp and suicide, only -1% | |

# Statistical test

ANOVA Assumption:

1. Normality - approximate normally distribution: Although the test and histogram show that data is not normally distributed, it is assumed that data is normally distributed according to Central Limit Theorem. In addition, the number of observations is large enough;
2. Independently Observations: countries are independent;
3. The dependent variable should be continuous: suicides values are continuous
4. Homogeneity of variances; according to Levene’s test, age, gender, gdp\_level and continent have unequal variances and are not hemogenic

Since, data is not hemogenic ANOVA test result cannot be considered reliable and I applied Welch

T-Test as well.

**Answer to research questions:**

## Does gender affects suicide rate? Yes, suicides rate in men is greater than women

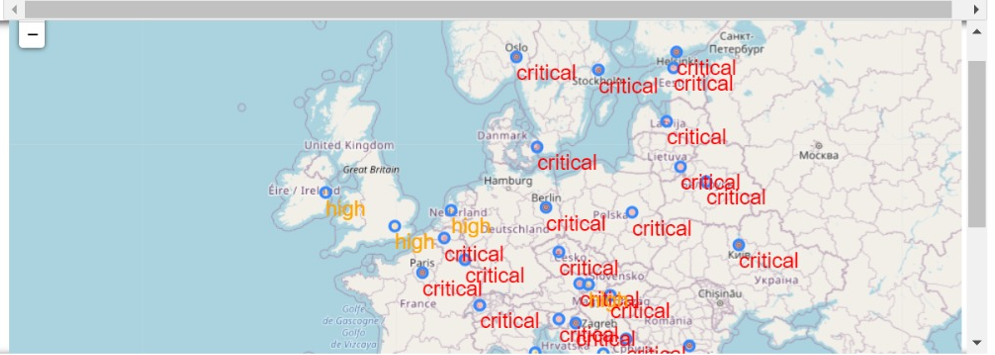
## Is there a differences in suicides caused by a specific age group? Yes, rate of suicides almost increased by aging.

## Is there a differences in suicides caused by specific continent? Yes,for some continents, suicides rate is significantly different.

## Is there any correlation between gdp and suicides? No,it is very low and negative.

# Visualisation

* 1. Rate of suicides on map



* 1. Some graphs has been shown on dashboard.



* 1. In various tabs, suicides pie charts are available on notebook (Figure 13)

1. <https://www.kaggle.com/mjella45/suicides-by-country/notebook> [↑](#footnote-ref-1)
2. <https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2020-2021> [↑](#footnote-ref-2)