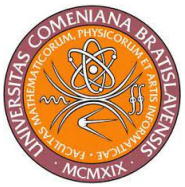
# Data analysis of suicides

Data management

Author: Peter Franček




Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
June 2023

# Contents

# Chapter A

# Introduction

The objective of this project is to analyze worldwide suicide data from https://www.kaggle.com/datasets/omkargowda/suic rates-overview-1985-to-2021?resource=download spanning from 1985 to 2020. The dataset used in this analysis includes information such as the number of suicides, country, year, suicide rate, GDP per capita, and more(Figure A.1). By delving into this extensive dataset, we aim to gain insights into the patterns, trends, and factors associated with suicide rates across different countries and over time.

| | country | year | gender | age | suicides_no | population | suicides/100k pop | GdpForYear | GdpPerCapita | generation |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Albania | 1987 | male | 15-24 years | 21.0 | 312900 | 6.71 | 2,15,66,24,900 | 796.0 | Generation X |
| 1 | Albania | 1987 | male | 35-54 years | 16.0 | 308000 | 5.19 | 2,15,66,24,900 | 796.0 | Silent |
| 2 | Albania | 1987 | female | 15-24 years | 14.0 | 289700 | 4.83 | 2,15,66,24,900 | 796.0 | Generation X |
| 3 | Albania | 1987 | male | 75+ years | 1.0 | 21800 | 4.59 | 2,15,66,24,900 | 796.0 | G.I. Generation |
| 4 | Albania | 1987 | male | 25-34 years | 9.0 | 274300 | 3.28 | 2,15,66,24,900 | 796.0 | Boomers |

Figure A.1: Dataset

Please note that while every effort has been made to ensure the accuracy and reliability of the analysis, limitations inherent to the dataset and analysis approach should be considered.

# Chapter B

# Methods description

home() - main page of Flask web app
country(random_id) - page for a certain country
year_country(country_id,year_id) - page for a certain country with a certain year
fig1-5() - graph methods(returns graph)

# Chapter C

# Data handling

## C.1 Preprocessing of data

First, we processed our data and removed those with NaNs in column 'suicides_no'. Since the data was concatenated with 2 different datasets we had to specially handle the counting of the population for each country for years bigger than in 2016. We also removed some unnecessary columns we did not want to include in our analysis.
At this stage, we also created sub-data frames for faster processing and looked at the correlation table to determine whether some columns connect together(Figure C.1).

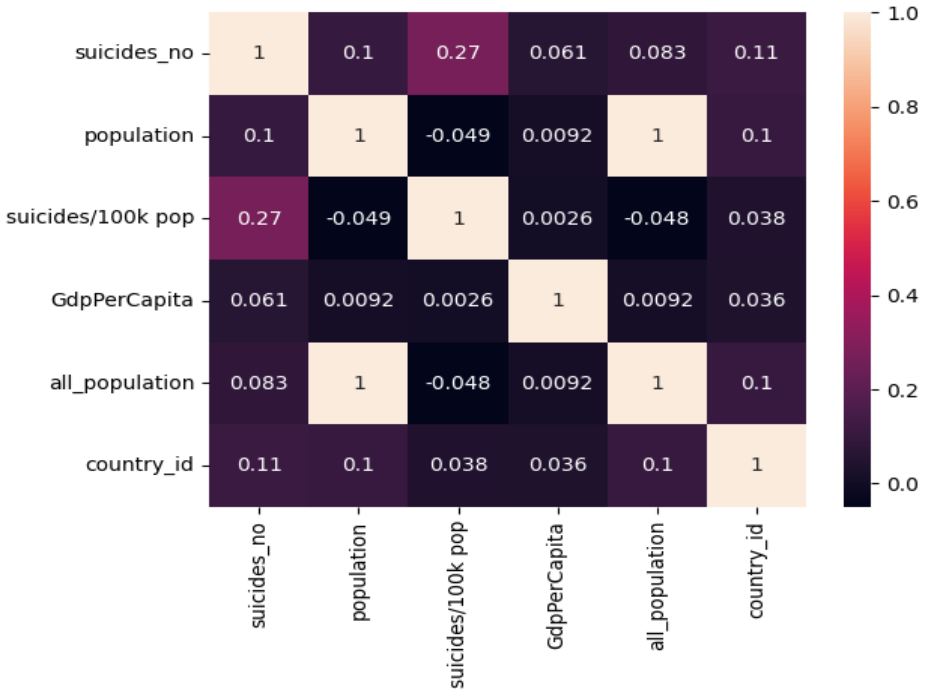



Figure C.1: Correlation table

# Chapter D

# Analysis and results

## D.1 Analysis of world

When we analyzed the number of suicides in the world annually, we came to the conclusion that about 76.5% of all suicides are committed by men. This number is unanimously very high.
When we further looked into the respective gender in both women and men most suicides were committed at the age of 35-54 years.

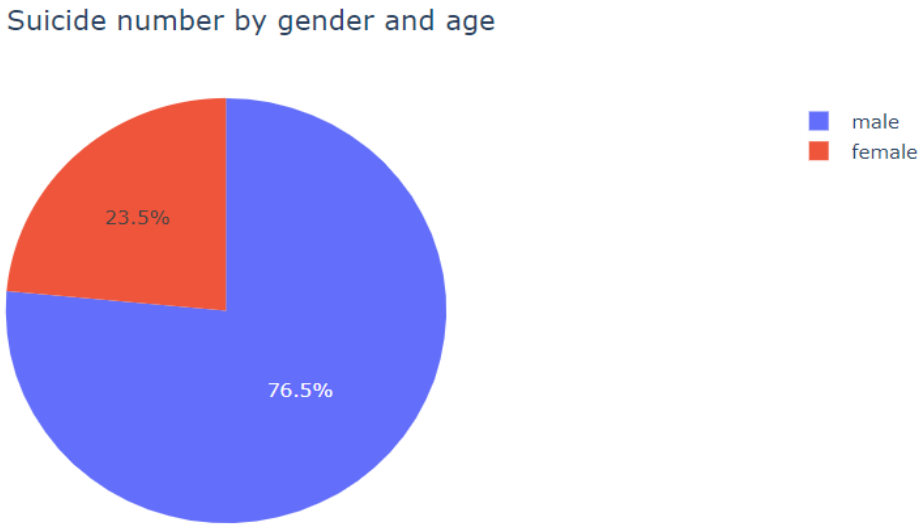



Figure D.1: Male vs Female % of suicides

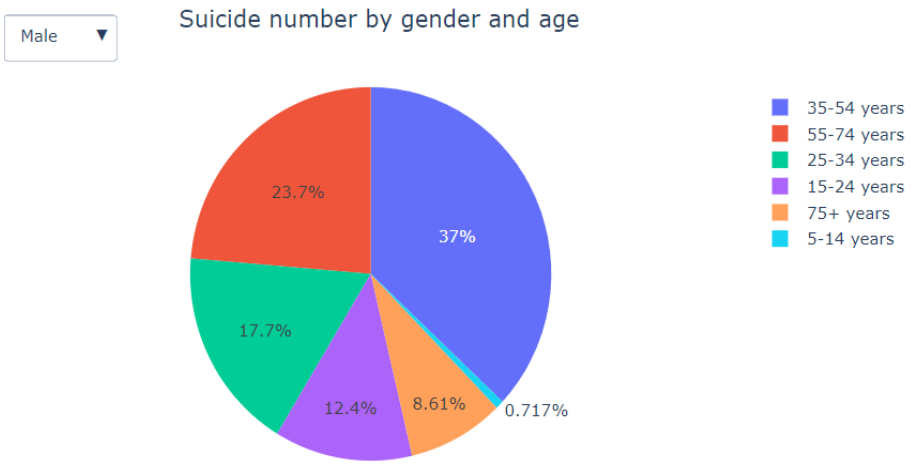



Figure D.2: Male with respect to age

For the overall suicides in respective years, there was not anything special despite the rapid increase from 1998-1990. The gap that you can see in the graph below in the years 2016 and 2020 is due to a small amount of data in these years compared to other years.

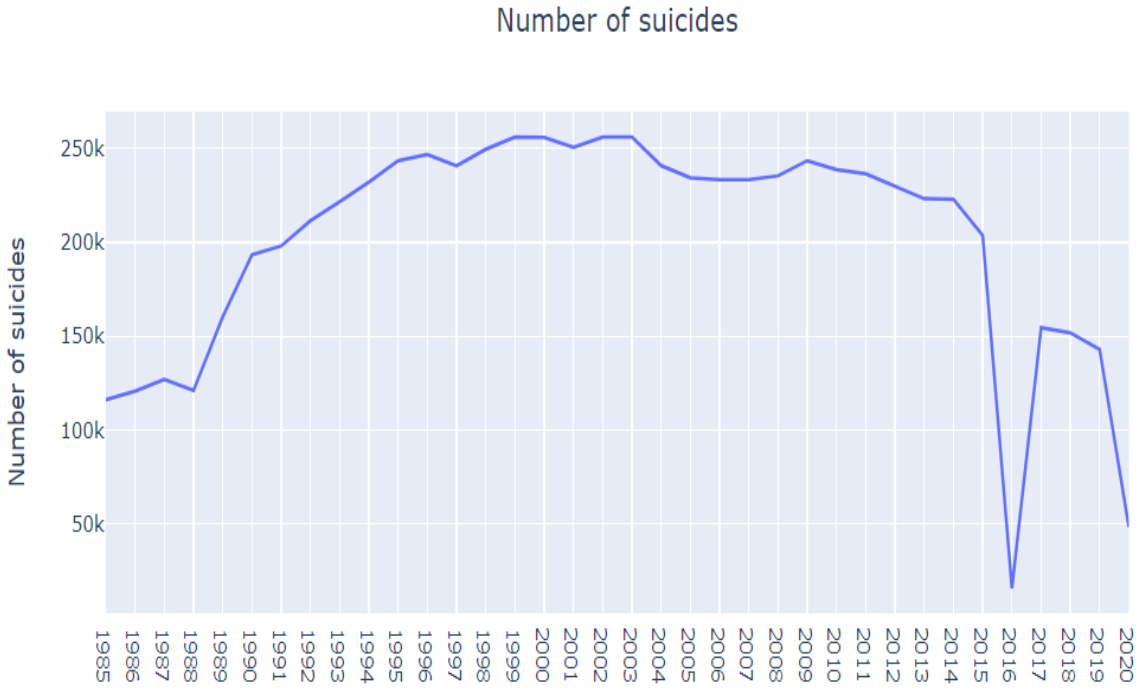



Figure D.3: Suicides per year

## D.2 Analysis of country

When we took a closer look at some specific country we could furthermore analyze whether the GDP per capita correlated with the number of suicides per year. Upon the overall dataset, the correlation was insignificant however when we took closer at some specific countries we could clearly see there might actually be some correlation, and to our surprise, it was directly proportional.

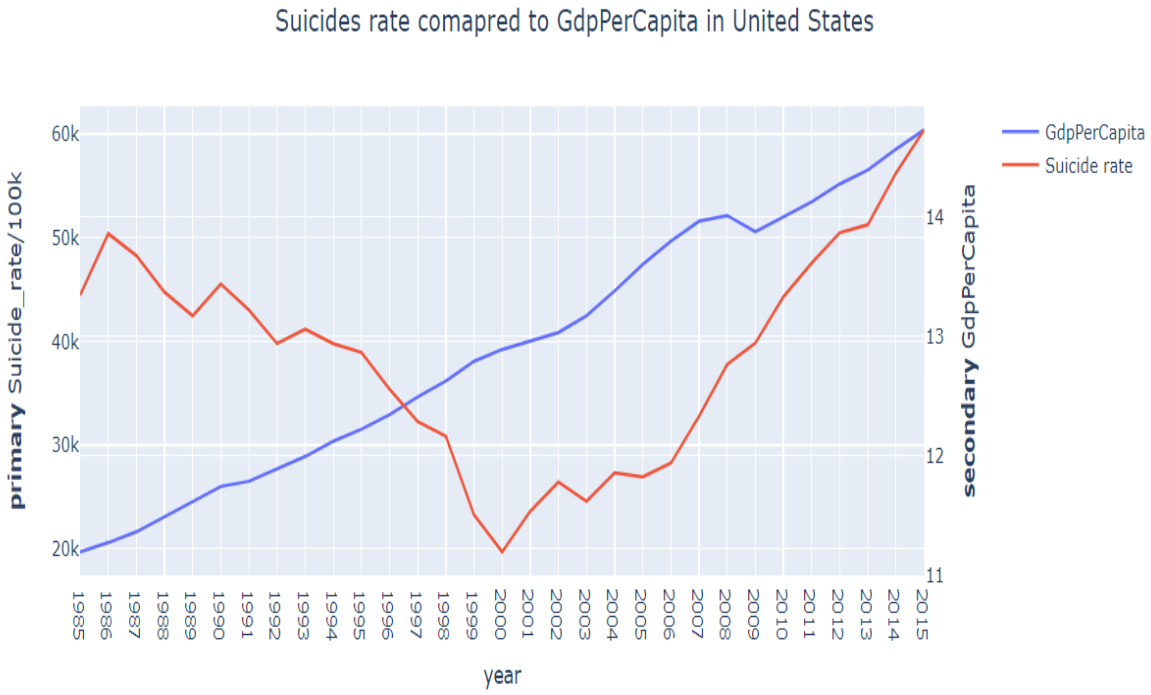



Figure D.4: GDP correlation with the number of suicides in the USA

Then with respect to genders, we tried to show trends using linear regression.

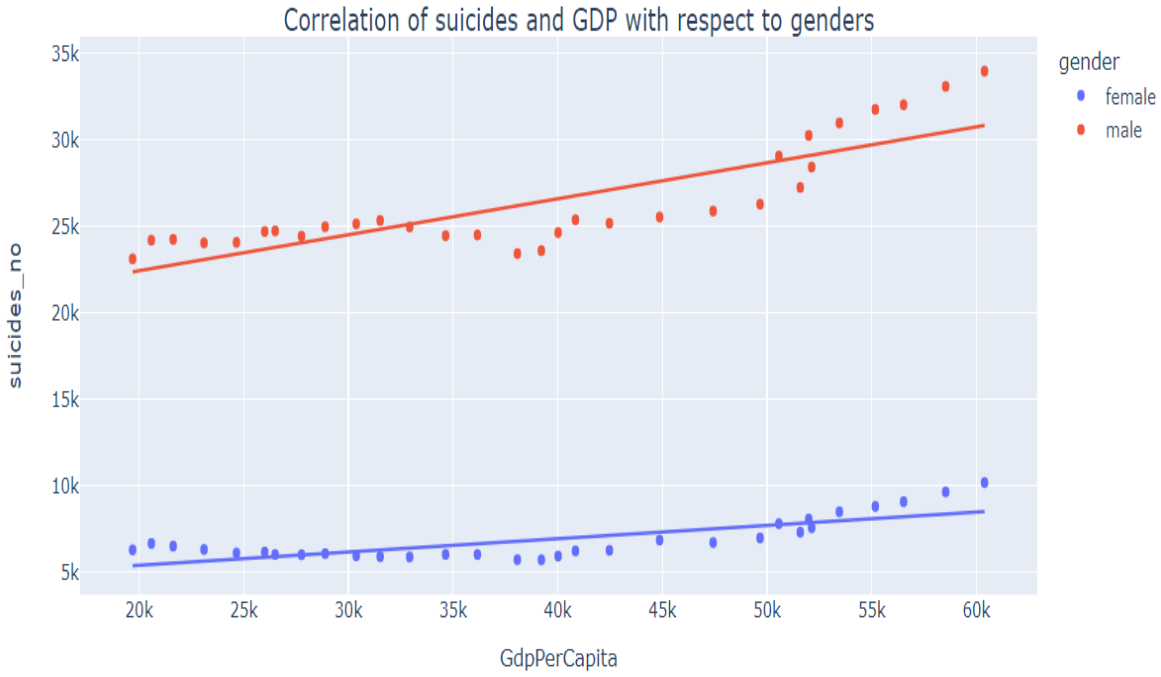



Figure D.5: Correlation with respect to genders

## D.3 Analysis of country in specific year

To dive even further we looked at the specific country in a certain year to see exact data. At this stage we could not make any more of it so we presented the results as given. We found out that our GDP per capita compared to real-time data at that year slightly differed, however in all countries and in all years the approximation was roughly only doubled, thus we did not need to perform any adjustments to our data set.

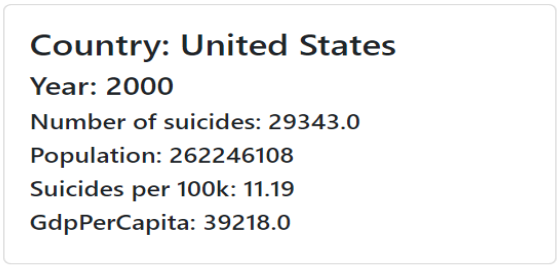



Figure D.6: Stats for the country in year

## D.4 Results

Firstly, the analysis revealed a significant gender disparity in suicide rates, with a higher prevalence among males compared to females. This observation prompts the need for deeper exploration into the underlying causes and societal factors contributing to this imbalance. Understanding why men are more prone to suicide can help in the development of targeted prevention strategies and mental health support systems.
Additionally, the data highlighted a specific age group, namely individuals between the ages of 35 and 54 years, as having the highest suicide rates. This finding emphasizes whether work has also an effect on suicides.

Since the data does not contain other variables, which may further reveal why in some certain countries the rate is so high while in others is not or whether there actually is correspondence with GDP and suicide rate.

Overall, this project has shed light on the worldwide suicide data and emphasized the importance of continued research in this field. By recognizing the limitations of the dataset and the challenges faced in graph creation, we have gained valuable insights into the intricate dynamics of suicide rates. Moving forward, it is crucial to delve further into the causes of suicide, conduct in-depth studies on societal factors, and explore potential interventions to address this global public health concern.

# Chapter E

# Conclusion

In conclusion, the analysis has provided valuable insights into suicide rates across countries and over time. Through the exploration of various variables, interesting findings have emerged.

The process of creating informative and visually appealing graphs proved to be a challenging aspect of this project. However, by utilizing tools such as Plotly, Stackoverflow, and other tools and leveraging data visualization techniques, we were able to effectively communicate the trends and patterns present in the dataset.

An easier task was putting it onto a Flask webpage, however, to design it so it is also visually appealing was certainly a pain to do.

During the project, I have learned that there are also other tools, that would probably make it significantly easier to do this project, but overall it gave me an interesting insight into how some websites are created.