



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Sistemas de Gestão de Dados

Projecto de Análise de Dados

<Suicide rates>

Miguel Arieiro <2014197166> marieiro@student.dei.uc.pt
Pawel Urbanowicz <2018314759> uc2018314759@student.uc.pt

Index

1. Introduction	2
1.1. Context	2
1.2. Used software	2
2. Data Analysis	3
2.1. Correlation matrix	3
2.2. World suicide trend	4
2.3. Suicides by continent	5
2.3.1. Average suicides per capita	5
2.3.2. Suicide trends by continent	6
2.4. Gender comparison	7
2.5. Age comparison	8
2.5.1. 5 - 14 years old	8
2.5.2. 15 - 24 years old	9
2.5.3. 25 - 34 years old	9
2.5.4. 35 - 54 years old	10
2.5.5. 55 - 74 years old	10
2.5.6. 75 years old	11
2.5.7. Conclusions	11
2.6. Suicide rates by country	12
2.7. Influence of GDP per capita on suicide rates	13
2.7.1. Linear regression	13
2.8. A closer look on Europe	14
2.8.1. Influence of GDP per capita on suicide rates	15
2.8.2. Influence of GDP on suicide rates	15
2.8.3. Influence of political stability on suicide rates	16
2.8.4. Influence of inflation on suicide rates	17
3. Conclusions	17
4. References	18

1. Introduction

In a world of chaos and apparent randomness there's information hidden in plain sight and that's what we're set to find and present on this report.

Along this document we'll analyse a few datasets obtained on different websites such as Kaggle, in an attempt to find interesting, relevant and perhaps information about suicide rates. In order to perform this task we've used the Python 3 programming language and some available libraries with which we've prepared and transformed our datasets, generating various useful outputs such as charts and tables.

1.1. Context

Suicide is the act of intentionally causing one's own death.

Suicide is among the leading causes of death in many countries. This is a concerning fact and hugely problematic on those countries, which is why it needs to be studied and understood.

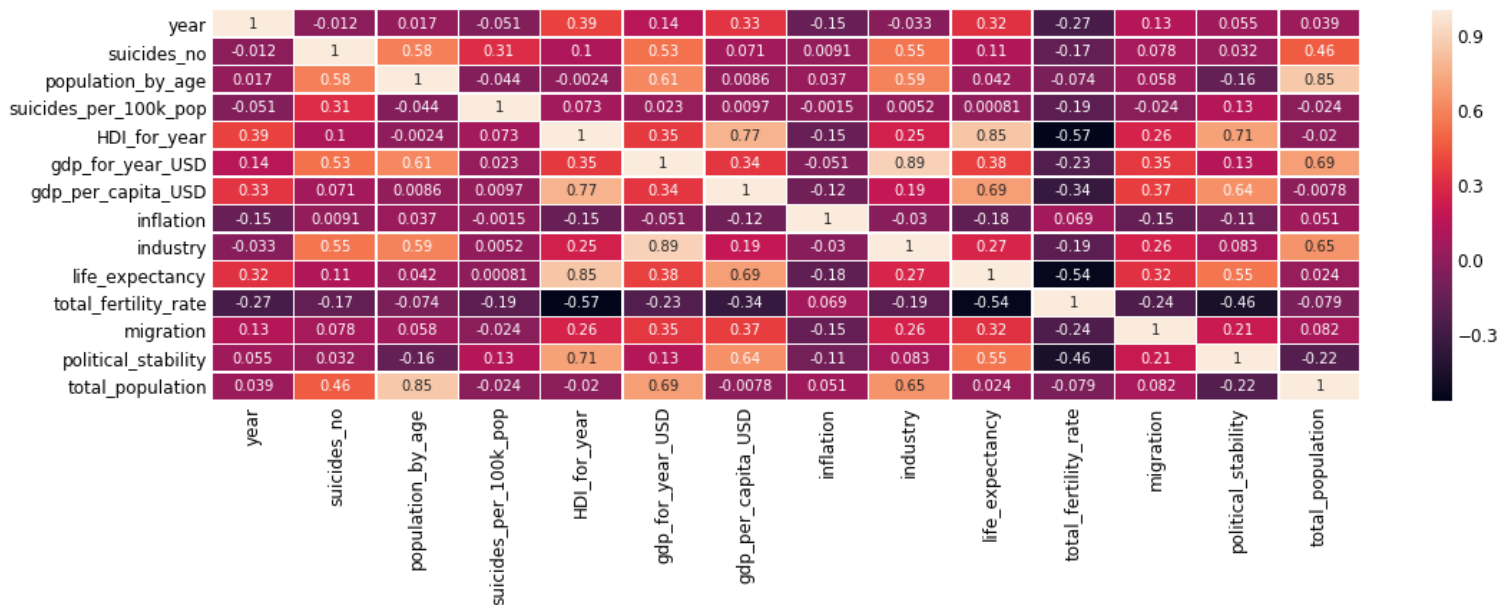
It can have many causes and influencing factors such as depression, bipolar disorder, anxiety, personality disorders, and substance abuse, as well as other factors. On this report we'll look for such other risk factors which can influence suicide rates across big populations.

1.2. Used software

- Jupyter Notebook
- Python 3.7
 - NumPy
 - Matplotlib
 - Pandas
 - Seaborn

2. Data Analysis

2.1. Correlation matrix

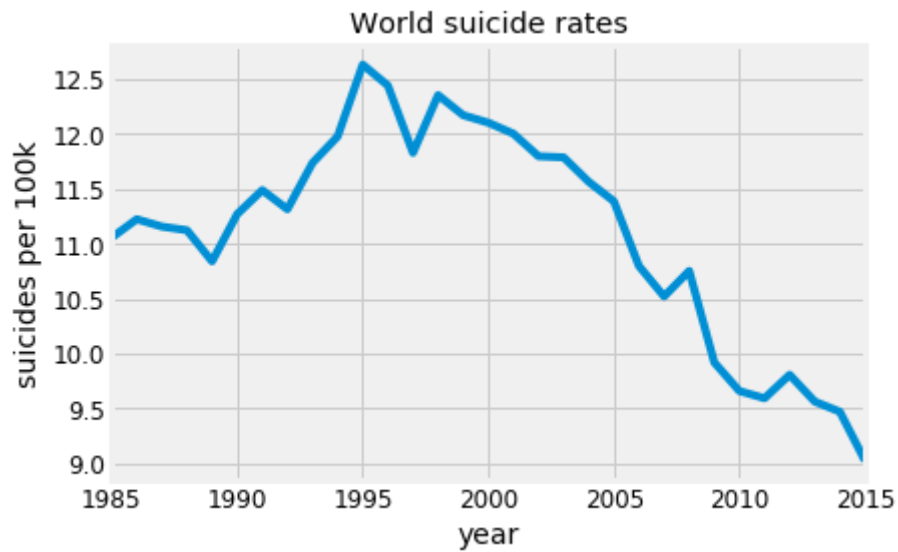


When looking at the correlation matrix we can easily spot some expected strong correlations across variables such as positive correlations between suicides per capita (suicides_per_100k_pop) and number of suicides (suicides_no), or between GDP for the year (gdp_for_year_USD) and GDP per capita (gdp_per_capita_USD).

We're also presented with several other between our main variable of study - suicides per capita - and apparently unrelated variables. One of such cases, is the presence of a positive correlation between the amount of suicides per capita and the Human Development Index (HDI) for the year (HDI_for_year). This seems to imply that most developed countries have higher suicide rates than undeveloped ones, which seems counterintuitive.

There also seems to be a negative correlation between fertility rate (fertility_rate) and suicides per capita, which is also heavily negatively correlated with the HDI.

2.2. World suicide trend

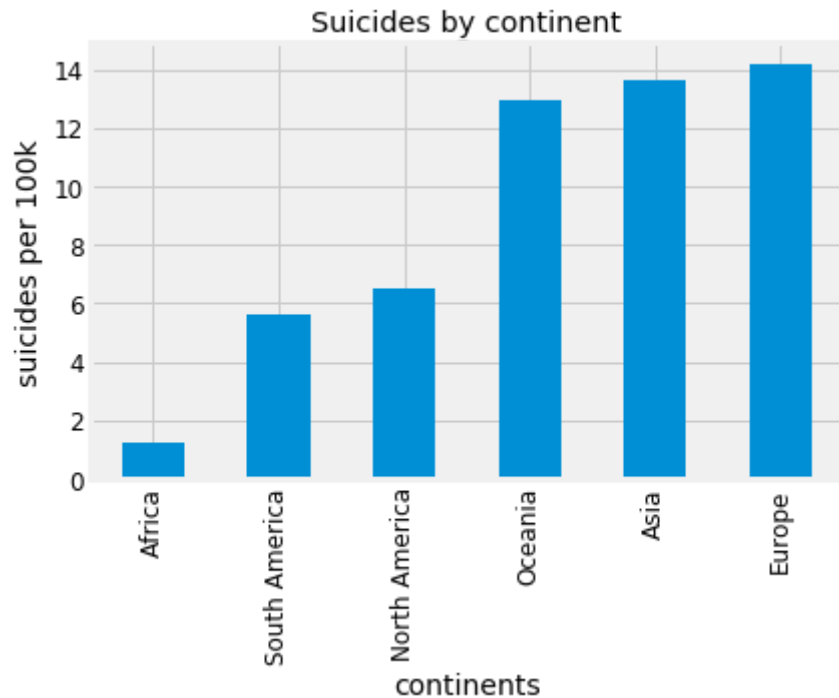


- **1985:** 11.056
- **2015:** 9.021
- **Peak:** 12.635
- **Low:** 9.021

We can spot a general increase in world suicide rate trend up until it peaks in 1995 at ~12.635 per 100k population. This increase is followed by steady decline which it maintained until the end of our time interval - 2015 - at which point there were ~ 9.021 suicides per one hundred thousand people, the lowest value over this 30 year period.

2.3. Suicides by continent

2.3.1. Average suicides per capita

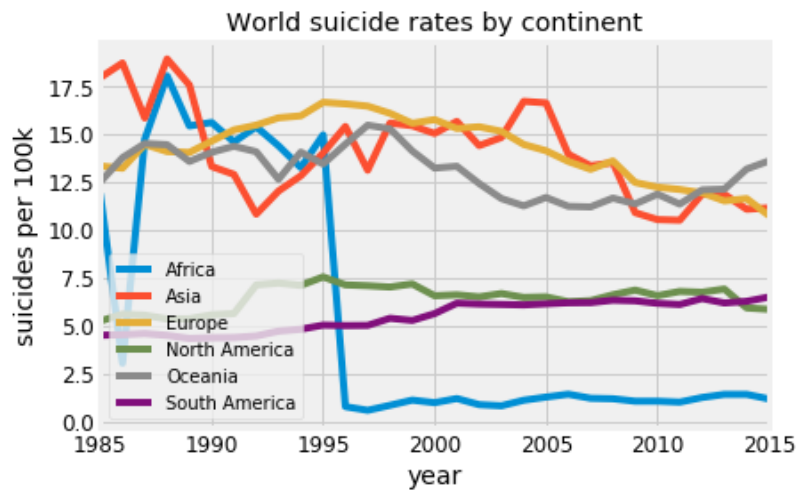


- **Africa:** 1.237
- **Asia:** 13.645
- **Europe:** 14.185
- **North America:** 6.503
- **Oceania:** 12.966
- **South America:** 5.581

At a glance we're able to see Europe at the top with 14.185 suicides per 100 thousand people and Africa at the bottom with 1.237.

Looking at this chart we can see a huge divide across continents in terms of suicides per capita. On one group there's Europe, Asia and Oceania, and on the other one there is North America, South America and Africa. The lowest suicide rate on the first group (Oceania) is over twice the highest of the second one (North America).

2.3.2. Suicide trends by continent



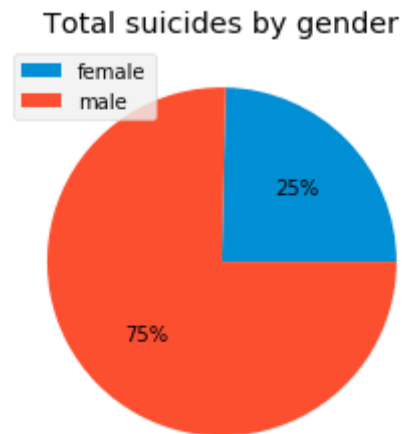
- **Africa:**
 - **1985:** 11.906
 - **2015:** 1.166
 - **Peak:** 18.054
 - **Low:** 0.572
- **Asia:**
 - **1985:** 17.961
 - **2015:** 11.156
 - **Peak:** 18.940
 - **Low:** 10.509
- **Europe:**
 - **1985:** 13.342
 - **2015:** 10.754
 - **Peak:** 16.673
 - **Low:** 10.754
- **North America:**
 - **1985:** 5.234
 - **2015:** 5.851
 - **Peak:** 7.547
 - **Low:** 5.234
- **Oceania:**
 - **1985:** 12.503
 - **2015:** 13.610
 - **Peak:** 15.477
 - **Low:** 11.2063
- **South America:**
 - **1985:** 4.494
 - **2015:** 6.496
 - **Peak:** 6.496
 - **Low:** 4.326

Looking at this chart we can clearly see a downward trend over the last decade across Asia, on the other side there's been an overall increase of the amount of suicides per capita in Oceania.

We can also spot a huge drop of Africa's suicide rate around 1986, followed by a sharp increase which peaked at 18.054 in 1988. It then stayed high up until 1995 at which point there was a sharp decline followed by a period of relative stability until 2015.

Both North and South America's suicide rates have been relatively steady since the beginning of this century.

2.4. Gender comparison



- **Male:** 1 053 040
- **Female:** 3 219 283

From this chart alone it's clear that there's a huge discrepancy between male and female's suicide rates. Despite there being less males than females females are only responsible for 25% of the cases.



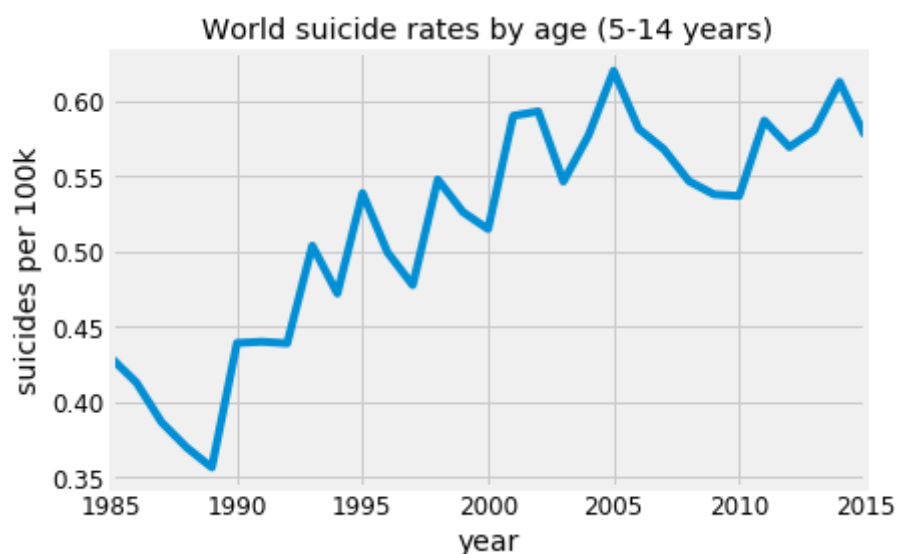
- **Male:**
 - **1985:** 15.616
 - **2015:** 14.048
 - **Peak:** 19.511
 - **Low:** 14.048
- **Female:**
 - **1985:** 6.621
 - **2015:** 4.185
 - **Peak:** 6.726
 - **Low:** 4.185

By taking a glance at this graph we can see a clear divide between male and female even when adjusted for population differences. To start males seem to commit suicide way more often than females, having a suicide rate at least twice as high as their female counterparts, ending this time period with almost three times as many suicides as the opposite genre.

While females' suicide rate seem to be decreasing since 1985, males' suicide rate has suffered an increase up until 1995, at which point it reached its maximum. Only after 1995 did it start slowing decreasing until the lowest value recorded during this thirty year period for the genre: 14.048.

2.5. Age comparison

2.5.1. 5 - 14 years old

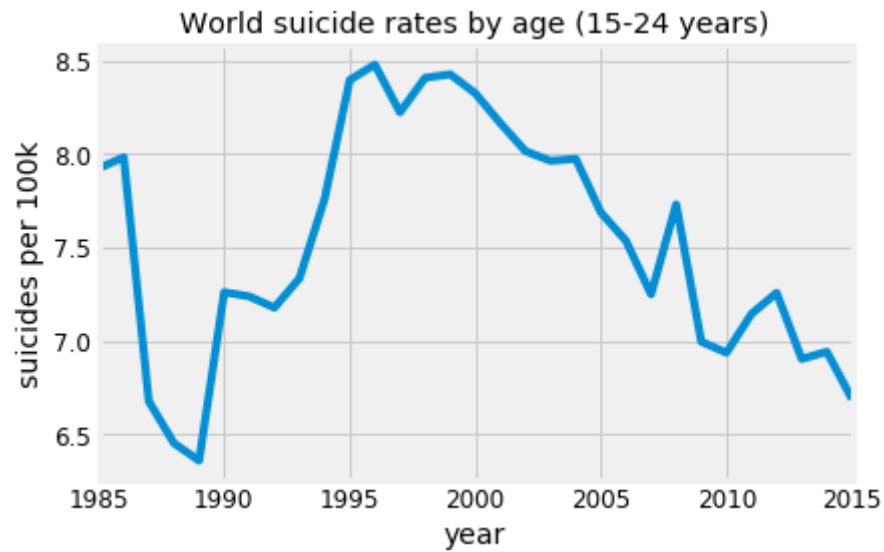


- **1985:** 0.429
- **2015:** 0.577
- **Peak:** 0.6203
- **Low:** 0.357

Looking at this chart we can see a sharp decrease since the start of this interval up until 1989, from which point we register an overall increase of the suicide rates. Unlike the remaining age groups, this group's suicide rate kept increasing well up until the current century with a peak of 0.577 per 100 thousand people in 2005 and another slightly lower one in 2014.

It's also the only age group with a higher suicide rate in 2015 than in 1985.

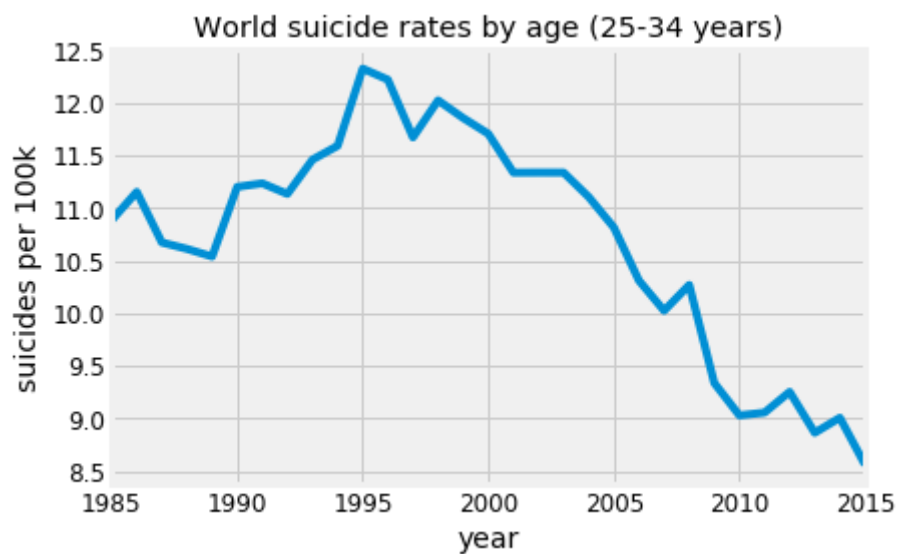
2.5.2. 15 - 24 years old



- **1985:** 7.923
- **2015:** 6.688
- **Peak:** 8.478
- **Low:** 6.359

On both the previous and this age group there is a substantial decrease up until 1989, followed by a sharp increase from 6.688 to 8.478 suicides per 100 thousand in 1996. This trend started then inverting and has been on decline since 1998.

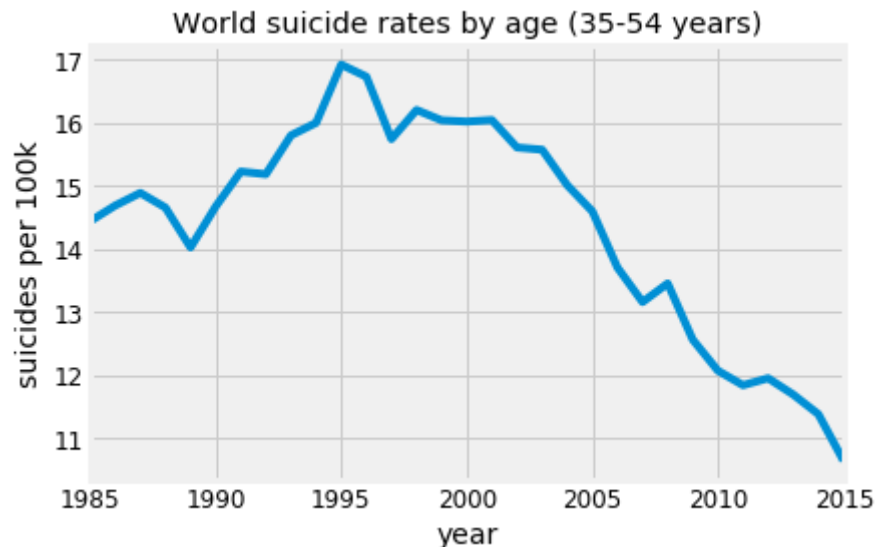
2.5.3. 25 - 34 years old



- **1985:** 10.883
- **2015:** 8.563
- **Peak:** 12.321
- **Low:** 8.563

On this chart we can spot a sharp increase at the end of the 80s, peaking in 1995, at which point the trend inverted and this age group's suicide rate started declining until the end of the studied time period, reaching a minimum of 8.563 suicides per 100k.

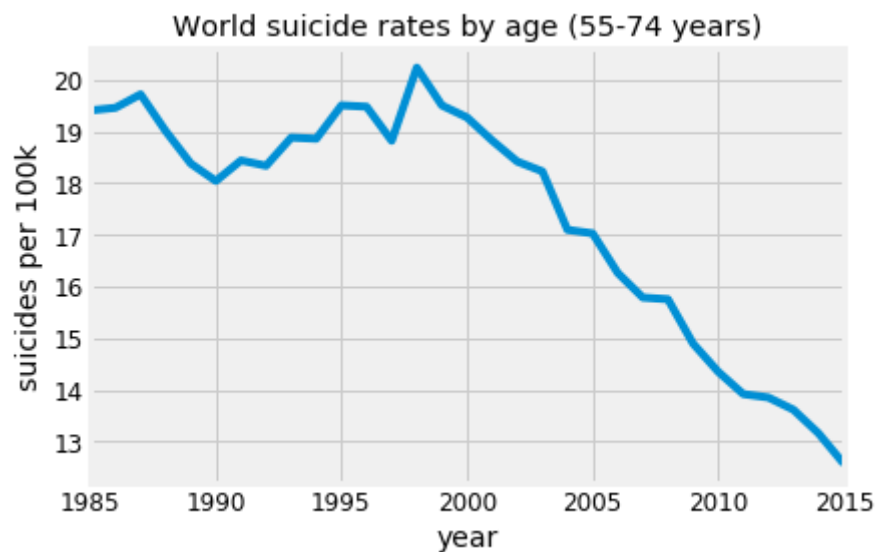
2.5.4. 35 - 54 years old



- **1985:** 14.444
- **2015:** 10.649
- **Peak:** 16.926
- **Low:** 10.649

Unlike the previous age group there's only a slight drop around 1988 followed by the same increase up until 1995 at which point it reaches a maximum of 16.926. From that point on there's a huge drop culminating in the lowest value recorded in the last 30 years for this age group: 10.649.

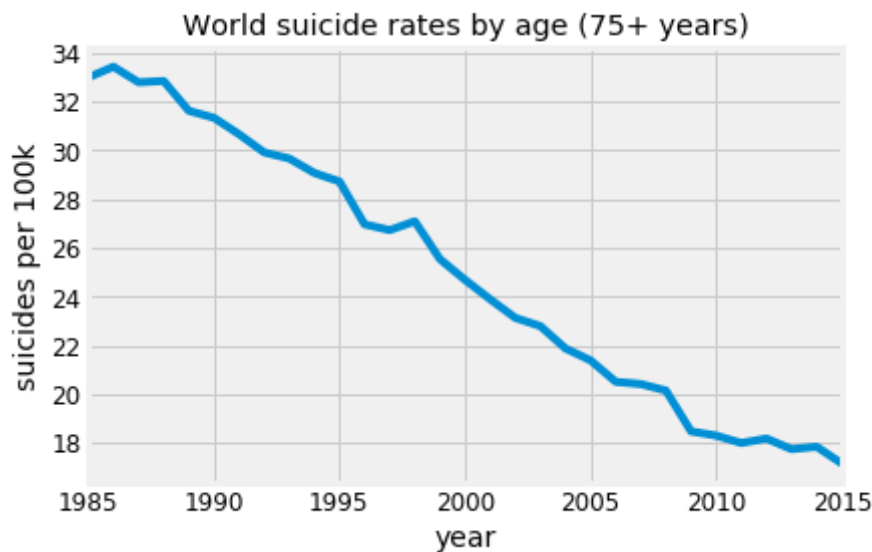
2.5.5. 55 - 74 years old



- **1985:** 19.409
- **2015:** 12.572
- **Peak:** 20.238
- **Low:** 12.572

On this age interval the drop in suicide rates at the end of the 80s is a lot less pronounced and it is followed by only a slight increase in suicide rates. Like with the previous charts we're able to see a huge drop at the end of the 90s. This downwards trend was maintained until the end of this time period, at which point this group's suicide rate was at its lowest value during the studied time interval: 12.572 suicides per 100k.

2.5.6. 75 years old



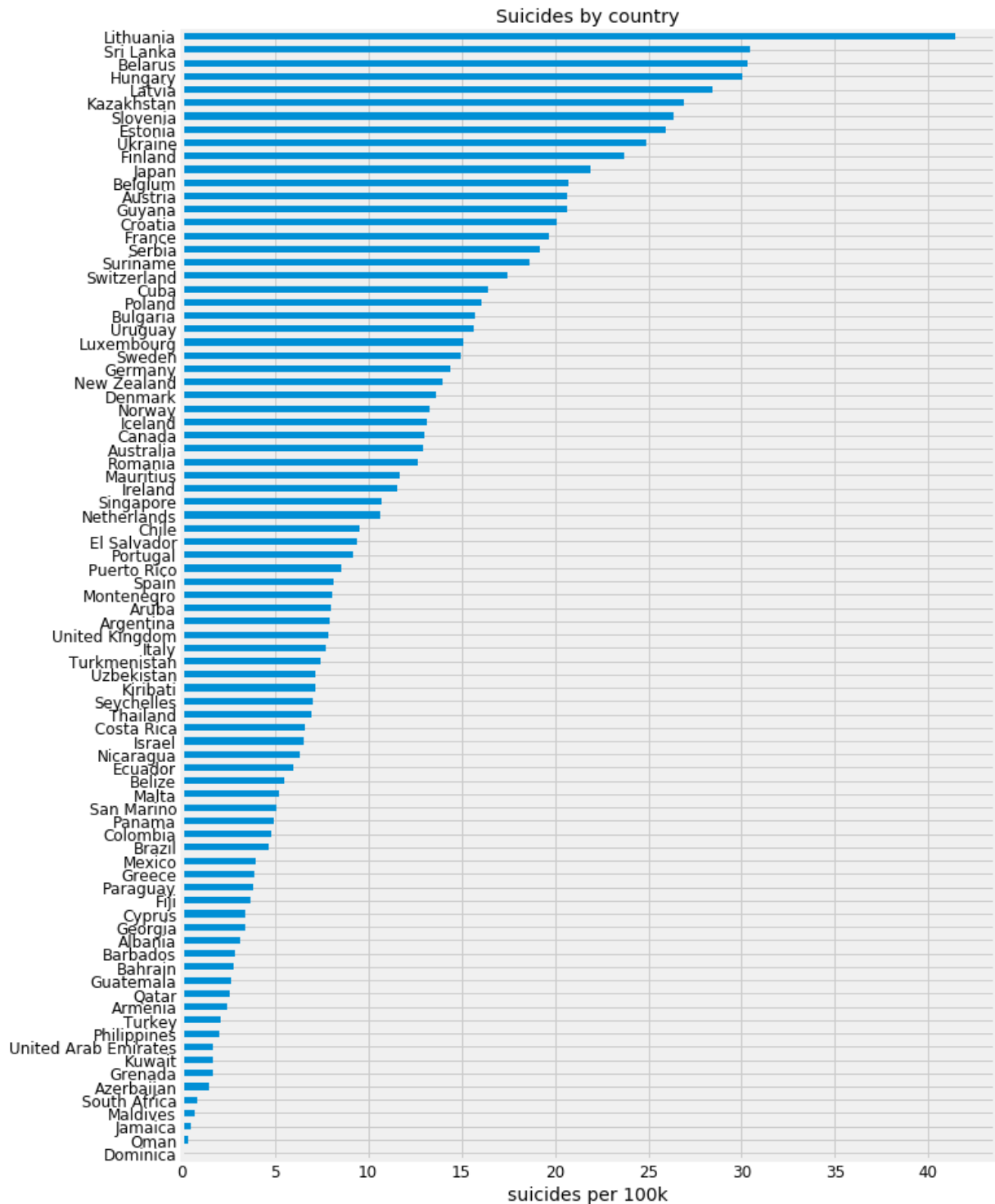
- **1985:** 32.978
- **2015:** 17.156
- **Peak:** 33.415
- **Low:** 17.156

The suicide rate of people over 75 years old has been quickly declining over this period of time. It started at 32.978 (per 100k) and was at a minimum in 2015 (17.156). Despite this it's still the highest among all the age groups.

2.5.7. Conclusions

Suicide rates seem to increase with age, being 5-15 the interval with the least suicides per capita and 75+ the one with the most. Despite this, suicide rates have been quickly declining since the mid-90s across all age groups besides younger one, in which they have suffered a slight increase.

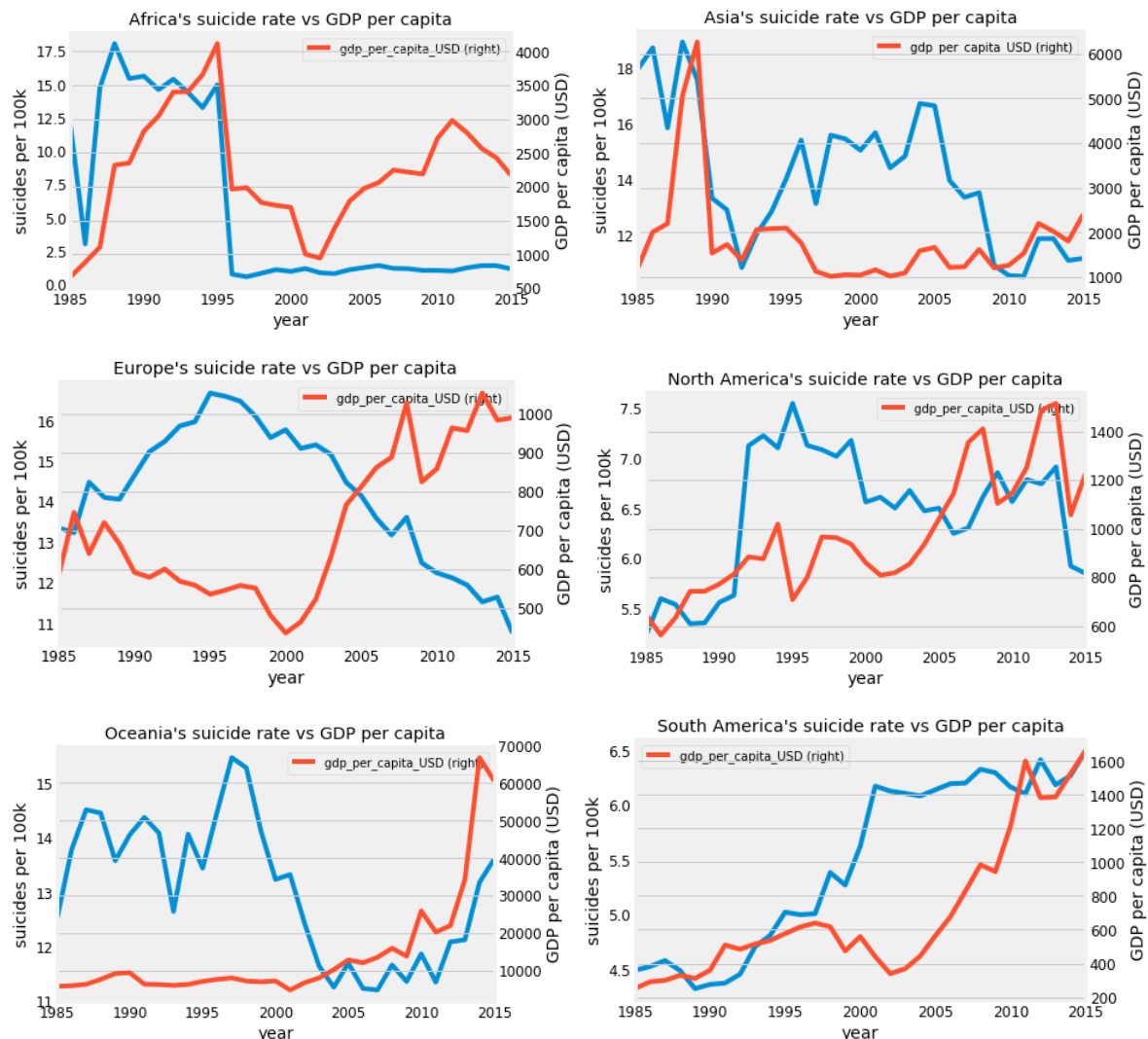
2.6. Suicide rates by country



When analysing this chart we can spot a clear outlier: Lithuania. It has over 10 more suicides per 100k population than the following country (Sri Lanka).

Looking at the top 20 we can clearly spot a large majority of European countries, which was expected after analysing the charts on 2.3.1. And 2.3.2..

2.7. Influence of GDP per capita on suicide rates

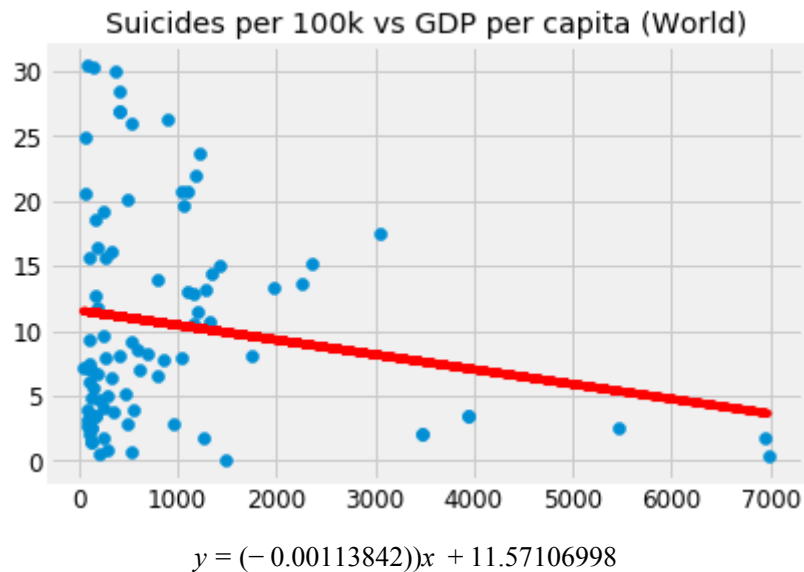


Looking at these charts we see that while in some continents, such as Europe, there seems to be a clear correlation between suicides per capita and gdp per capita, this relationship doesn't seem so clear on others such as Oceania.

When taking a closer look at Europe's suicide rate per capita vs GDP per capita, we quickly realize that there seems to be a negative correlation between both variables. In other words, the higher the average GDP per capita, the lower the suicide rate. On the other side we have Africa's data, and Oceania's recent rise of the suicide rate per 100k which seems to coincide with an augment of its GDP per capita.

2.7.1. Linear regression

In order to get a better idea of the relationship between world's suicide rates and GDP per capita we'll take a look at its linear regression plots. After removing any outliers we're presented with the following char:



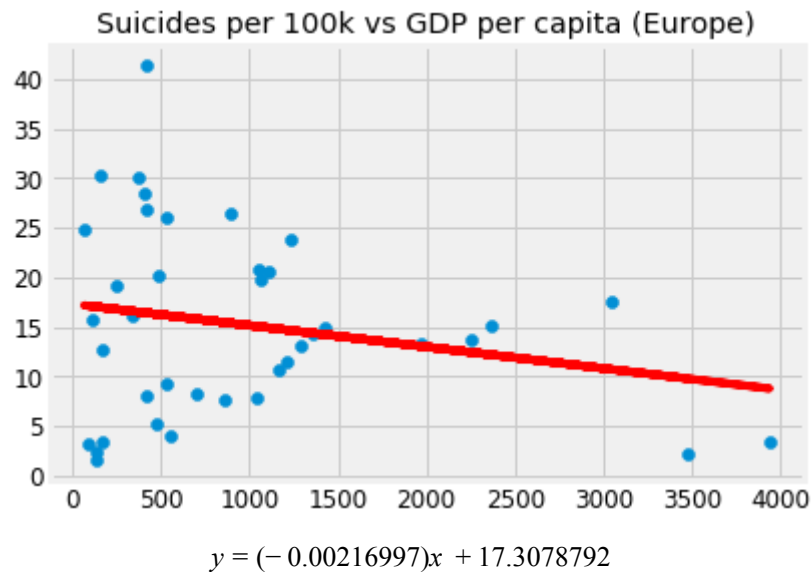
With a p-value of 0.083 we are able to conclude with over 90% confidence level that GDP per capita is negatively associated with suicide rates, although it has a low R-squared value (0.034) , which means it's influence on the suicide rate is very small.

2.8. A closer look on Europe

On this section we'll focus on the European continent as a whole. We'll look into other variables and how they influence European suicide rates. In order to do that we'll use mainly regression and K-means clustering.

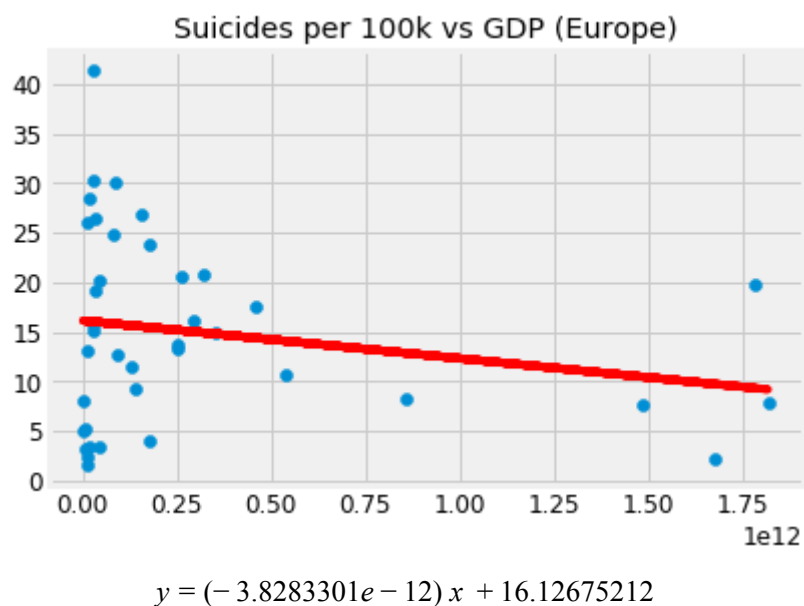
The following charts compare the average suicide rates of each European country over the studied time frame (30 years) and compare them with different variables we thought these values may be influenced by. Prior to the generation of each regression model we've removed any outliers on both variables.

2.8.1. Influence of GDP per capita on suicide rates



This chart represents the comparison between suicide rates over Europe and their corresponding countries' average GDP per capita. With a p-value of 0.188 it's impossible to conclude anything about the influence of GDP per capita on European suicide rates with a good confidence level. With this in mind, the model seems to imply that there might be a small negative influence (R-squared 0.046) with 1 less suicide per 460.835864 increase of the GDP per capita.

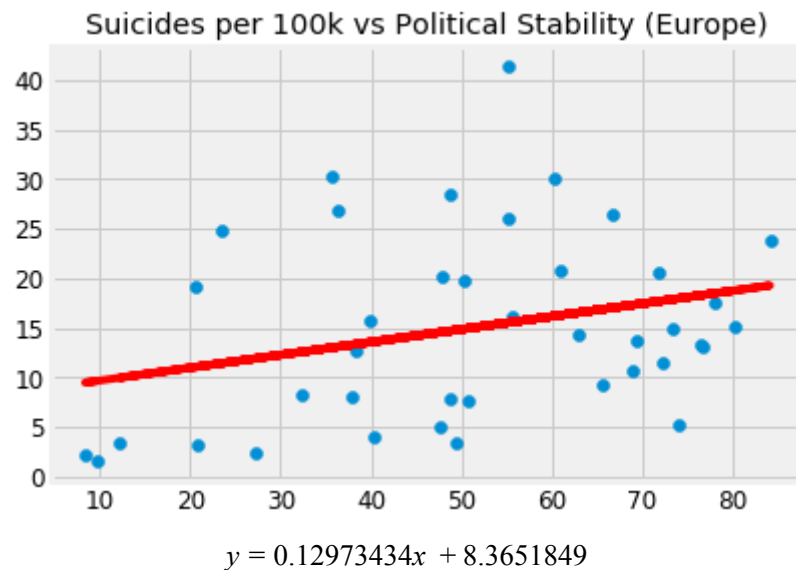
2.8.2. Influence of GDP on suicide rates



This chart represents the comparison between suicide rates over Europe and their corresponding countries' average GDP. With a p-value of 0.219, much like with GDP per capita we're unable to conclude anything with a high level of confidence based on this

model. Even with such a low confidence level we could only conclude that GDP has a small influence on the number of suicides per capita due to a low R-squared value of 0.041. Much like with the previous model's GDP per capita, according to this model, an increase of GDP seems to be negatively associated with the European countries' suicide rates with 1 less suicide per 100 000 population per 261210.5 million USD increase of the GDP.

2.8.3. Influence of political stability on suicide rates

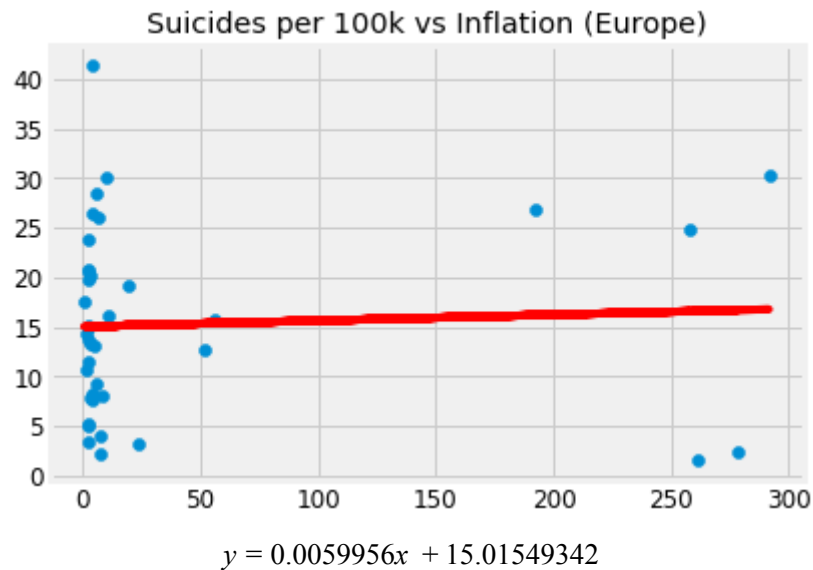


Political stability is a number between 0 and 100 attributed based on a country's political stability. The higher the country's political stability, the higher the score.

With a relatively low p-value of 0.076 it seems that we can declare with over 90% confidence that there's an association between both suicide rate per capita and political stability in Europe. However by looking at the R-squared value of 0.080 we quickly realize that despite being the variable with the most influence on the suicide rates up until now, it still has nothing but a small influence on those.

According to this linear regression in order to lower the number of suicides per 100k population by one, a country would require a decrease of -7.70805941 points on the political stability index.

2.8.4. Influence of inflation on suicide rates



With a p-value of 0.743 we can't reject the hypothesis that inflation has no influence on suicides per capita.

3. Conclusions

After analysing all this data we found several unexpected results.

To start with suicides per capita seem to be declining on most of the world, across all ages but children aged between 5 and 14 years old. Despite this it seems way more prevalent among older age brackets, with people over 75 years old committing suicide as often as 1.36 times per capita as the previous age group (55-74). There's also a huge discrepancy between genres. Males committed three quarters of the world's suicides, despite representing less than half of the global population.

We've also found data supporting that suicide rate might be slightly negatively correlated with a country's GDP per capita and political stability, although with a low confidence level (over 90%).

With all this in mind we can only conclude that suicide is a complex with many causes and influences that can't be easily isolated. For this reason we believe that it needs to be heavily studied by different fields in order to be better understood and hopefully have it's frequency reduced.

4. References

- <https://matplotlib.org/>
- <https://www.numpy.org/>
- <https://pandas.pydata.org/> [
- <https://seaborn.pydata.org/>
- <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>
- <https://www.kaggle.com/ammon1/demographic>
- <https://www.kaggle.com/juanumusic/countries-iso-codes>
- https://datahub.io/JohnSnowLabs/country-and-continent-codes-list#resource-country-and-continent-codes-list_zip