Suicide analysis: educational research project

Nikita Chuikin

1   Introduction

Definition. In this work suicide is accepted in classic Durkheim's definition: "...all cases of death resulting directly or indirectly from a positive or negative act of the victim himself, which he knows will produce this result." The key factor is intention to commit suicide but as it will be discussed later it is not always possible to establish whether the death was voluntary.

   Why is it important to study suicides? Besides moral costs and being marker of social decease, a high rate of suicide has it's economic costs. It usually includes direct medical costs and indirect future earnings. However, estimates of these costs differ:

1. $260 691 per person who committed suicide (Stoudemire, Frank, Hedemark, Kamlet, Blazer, 1986)

2. $33 000 per person who attempted suicide (Palmer, 1995)

3. $141 923 per person who's death cause was self-inflicted injury (Donald Shepard, 2015). 97% of costs are indirect (lost of productivity).

4. $27.43 millions per person died from self-inflicted injury. (Ian R.H. Rockett, Bina Ali, Eric D. Caine, Donald S. Shepard, 2023). Authors estimated quality of life losses which gave the predominant sum.

5. Bijou Yang and David (2007) Lester argue that previous scholars did not take into account potential savings from non-treatment depression cases and savings from pensions and nursing. If we do, we will get $5.07 billion net savings from suicides

   All of scholars above are related to the United States statistic. Estimates have not been corrected to inflation and approximately represent costs by a year of publishing the paper.

Statistical issues. Official national statistics of suicide rates underestimate suicides due to a lack of possibilities to determine whether the death was intentional or not (I.R.H. Rockett, N.D. Kapusta , J.H. Coben JH, 2014; P.N. Cooper, C.M. Milroy, 1995). Futhermore, one of the most reliable source of world mortality World Health Organisation (WHO) has no statistics exactly about suicides but only about 'self-inflicted injures' which will be used in this analysis.

Emile Durkheim - a classic of sociology - had studied suicides in his book of the same name published in 1897. He analyzed specific statistics and suggested 3 main types of suicides: egoistic, altruistic and anomic one. WHO's data can be used to verify the last type by analysing particular countries which will be done in this analysis on the Russian case example.

The goal of the analysis is to explore fundamental suicide statistics and valuate Durkheim's explanations of suicides.

2   Data preparation and exploration

On this stage the data is being uploaded from WHO mortality database, extra columns with duplicated information (for instance, Age group code and Age group contain the same info for out analysis), Nan values will be deleted.

Main target values for this analysis are death rate per 100 000 population and percentage of suicides out of total deaths. The data is provided since 1950.

Due to the specific of the data, variance coefficient for death rate is 56% > 33% (and 140% through all data set). That's why further below median method is used for all groupings.

3   Suicide, age and gender

Let's group the data set by sex and age and see what people is most of all in risk group. The most frequently suicide affects people over age 80. In spite of buzz about teen's high suicide rates, they commit suicide less than any other people do. Distribution by sex reveals that women's suicide is spread in the middle-age (45-70) while death rate of men increases over the years.

Dynamic of suicide rates by sex *presents similar tendency for both genders*: decreasing suicide rates. Also, the figure 3 confirms the data on the fig. 2 - *men more frequently commit suicide* (there is no year since 1950 which in women suicide rate is more that men's one).

It's interesting to measure variation between men and women death rate each year to give numeric estimation of how two time series replicate each other. Another words, the goal is to measure whether two curves act the same over time. The variation coefficient of distance between two curves can be used for our purpose. It is not applicable for measurement opposite tendencies or assumption that one curve can change more than the other one (for instance, x^2 and e^x) but our goal is free from these biases. So, variation coefficient is 0.14 (14%) - therefore, *there are no statistically significant differences between dynamic patterns of men and women suicides*.

While distinctions of suicide rates are obvious, the reason of it is unclear. It's essential to keep in mind the process of gathering the data. P.N. Cooper and C.M. Milroy (1995) noted that women's suicide can be under-reported because of difference in methods of committing it: "*Over 50 per cent of females chose poisoning with solid or liquid compared to less than one third of males. This method of suicide had the second lowest rate of suicide verdicts at 40 per cent of cases.*" Using the self-injury statistics smooths out these distortions but does not remove them.
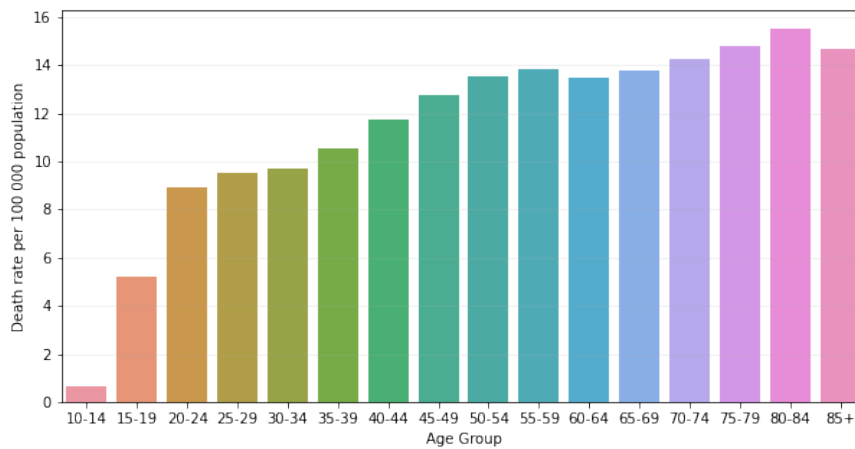

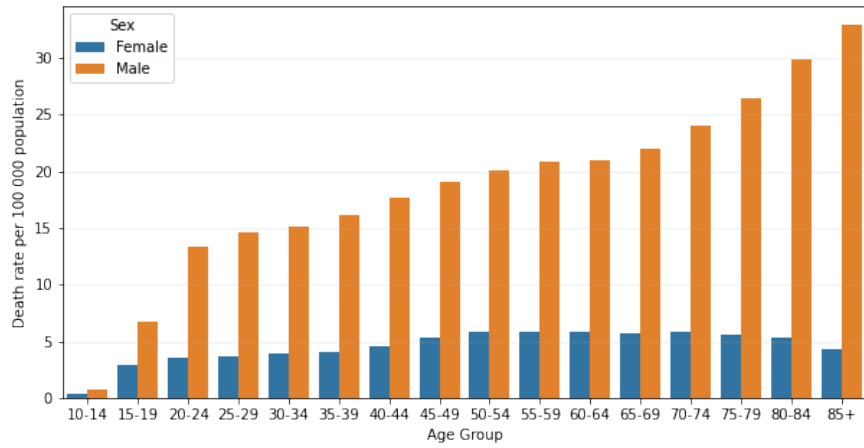Figure 1. Mean world suicide death rate distribution by age groups since 1950

Figure 2. Mean world suicide death rate distribution by sex and age groups since 1950
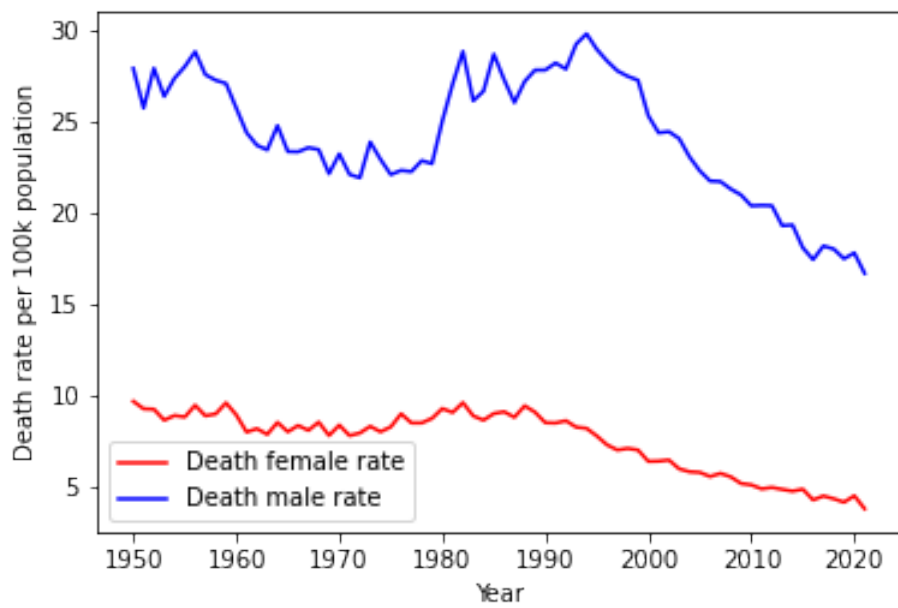

Figure 3. Suicide death rate per 100k population by sex

4 Russian case

Looking at the first figure which presents suicide rate dynamic one can see a significant increase both suicide rate and percentage of suicide in all deaths

in 1990s. It can be explained with:

- **a high number of crimes** at that period

- **economic crisis and poverty**

- **anomie state** (in terms of Durheim's definition). (Durkheim, E. Suicide: A Study of Sociology. 1897). Durkheim's explanation was a crisis of values when demand of people from different stratum are not limited by social norms of consumption.

However, the figure 5 with the same indicators but for the whole world shows similar dynamic in the period. Hardly criminogenic situation was the main factor all over the world and economic poverty can't be extended to all countries. Anomie state also do not explain rising suicide rates over the world because it is deeply connected with culture which is not the same for all societies. That's why *either Durkheim's definition of anomie state should be enhanced or anomie state is not enough for rising suicide rates.*
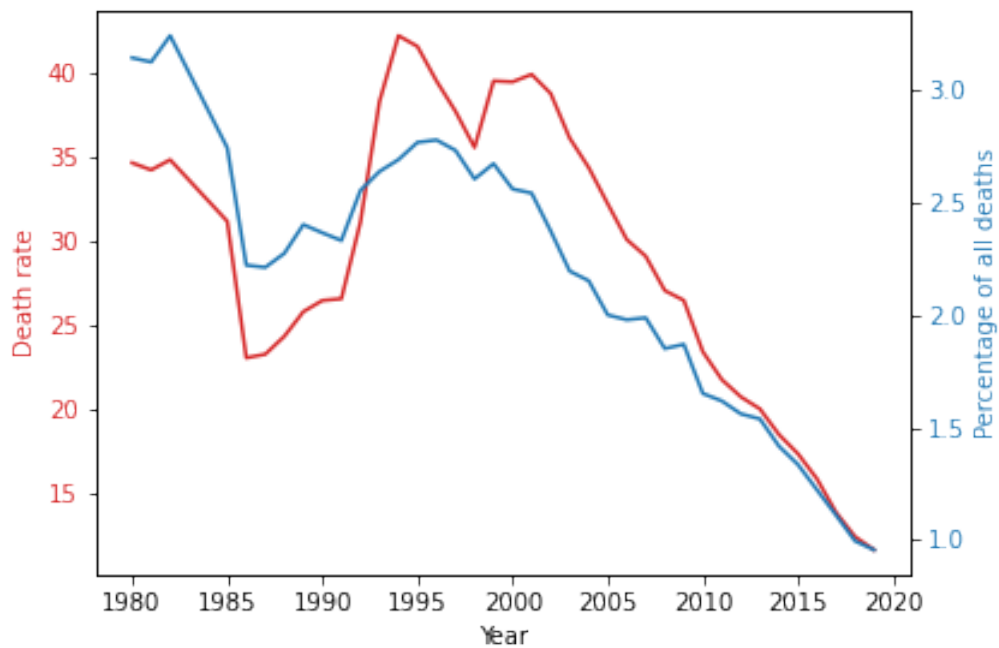

Figure 4. Suicide rate dynamically in Russia since 1980 to 2020
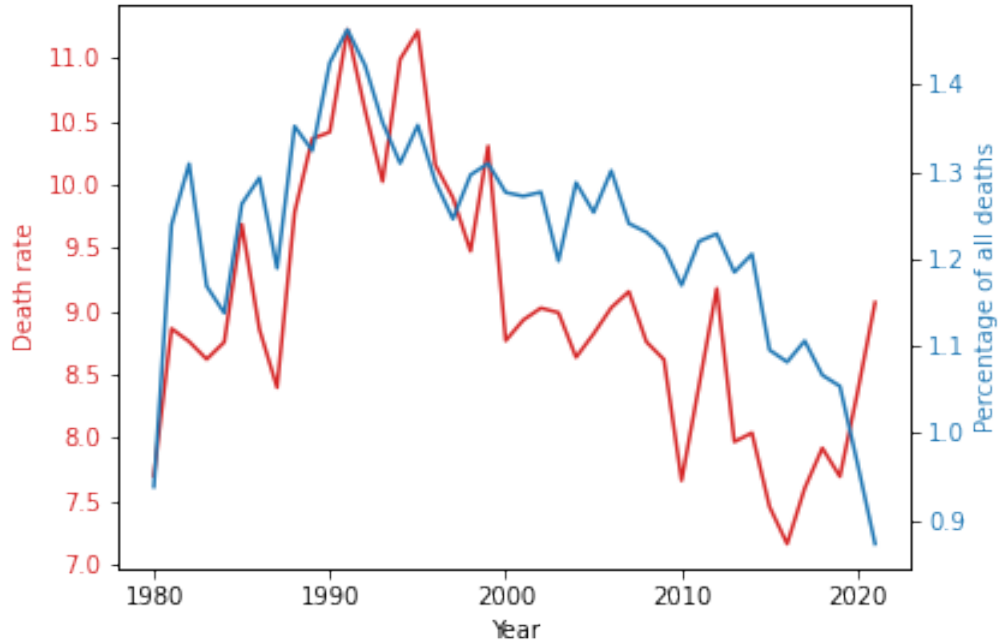
Figure 5. Mean suicide rate across the world since 1980 to 2020
Suicide rate dynamically in Russia since 1980 to 2020

5   Suicide and economic factors

The significant part of literature is dedicated to establish association between suicide and poverty (J. Bantjes, V. Iemm, 2016). Emile Durkeim (1897) argued that low-income is protective factor from suicide mainly because of 2 reasons: affluence can lead to anomie or egoistic suicide, poor communities are more united, so the feeling of social integration prevent from intentional self-inflicted deaths. Fabio Ferretti and Anna Coluccia (2009) gave a supporting statistics for the first Durkeim's explanation using discriminant analysis - European countries with low healthcare expenditure, education level and high annual GDP growth have high suicide rates while states with less thriving economies have also less suicides. Nevertheless, analysis in this study considered unemployment rate is non significant but Anja Baumbach, Gabriel Gulis (2014) came to opposite conclusion studying the same countries: "...*when social spending is low in a country, unemployment rate has a stronger effect on suicide mortality*". Another study (S. Burrows, N. Auger, M. Roy, C. Alix, 2010) revealed a correlation between suicides and inequality which can be an explanation for the results of Fabio Ferretti and Anna Coluccia (2009). Case-study in Portugal (Paula Santana, Cláudia Costa, Graça Cardoso, Adriana

Loureiro, João Ferrão, 2015) also committed an association between material deprivation, economic crisis and suicides.

Thus, GDP growth can influence on suicide through:

- healthcare and social insurance expenditure

- economic inequality

- unemployment

But can GDP be directly related to suicides? *The purpose of this part is to visualize some statistics, explore correlation between annual GDP growth and suicides and make a linear regression to demonstrate the possibilities of it's interpretations.* The same results can be established by pairwise comparison of categorical data.

Firstly, let's match classification of the country by World Bank (high, upper middle, lower middle and low income) for 2021 and death rate dynamic. It is not quite correct to use classification just for one year and extrapolate it to all times but assume that relative position of countries have not changed much. Fig. 6 presents that lower middle countries suicide mortality have been between high and upper middle most of time expect the period 1975 - 1986. High income countries usually have also the highest suicide rates. One can already spot non-linear dependence from economic factors and this hypothesis will be confirmed later.

Fig. 7 is an interactive map which on suicide rate by countries were displayed. Despite a lack of statistics for African countries, examples for both positive and negative relation between income and suicides can be found.

Finally, we will get a quantitative estimate of the association. World Bank's data set with annual GDP growth by years and countries was merges to the original data set to explore a correlation. Annual growth is considered to represent changes in financial position of residents which are more essential than just a static position. Fig. 8 display a distribution of investigated variables - another proof that there is no linear association between them, hence regression analysis can be used. Correlation coefficient gives numeric evidence for this claim, it is 0.0032 and 0.0010 for all data set and data set without outliers respectively - *there is no significant association between GDP growth and suicide mortality.*

At last, let's discuss making linear regression where variables are original data set's columns: age group, sex and region. Each value of age group was converted to the middle of a particular interval and scaled. Since sex and region are categorical variables, they have been coded with one-hot encoding in order to get relevant results. It's important to note that all weights of categorical variables only can be compared with each other. Both linear models

7

with and without regularization were tested. Even Lasso model gives edge solutions and make some of weights equal to zero. Median absolute error is used as a metric since data set contains outliers. Lasso model performs better within statistical margin of error, therefore regression without regularization will be used for analysis.

Sorting weights of trained model one can see results that have been obtained later - men commit suicide more frequently than women. Also, regions with the highest suicide rates are Europe and Oceania, America and Asia are with the lowest ones. *This is indicative because one of the richest region (Europe) is close to poor rural region of Oceania - another proof of absence of correlation.*
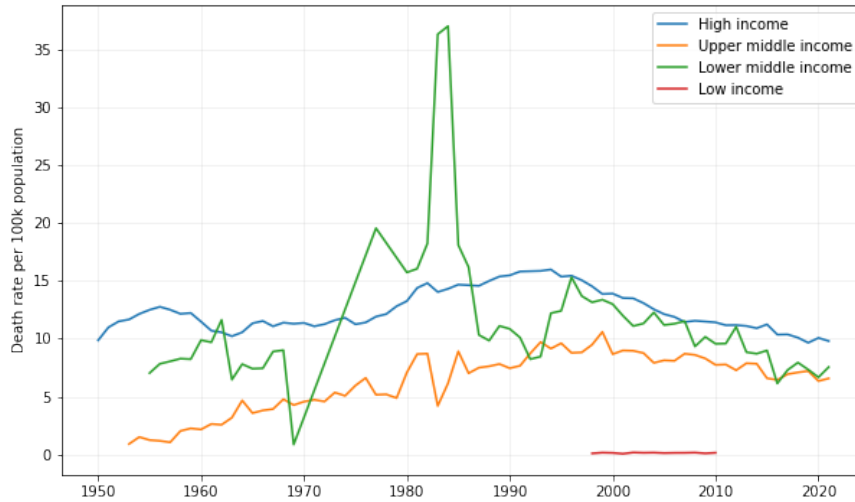

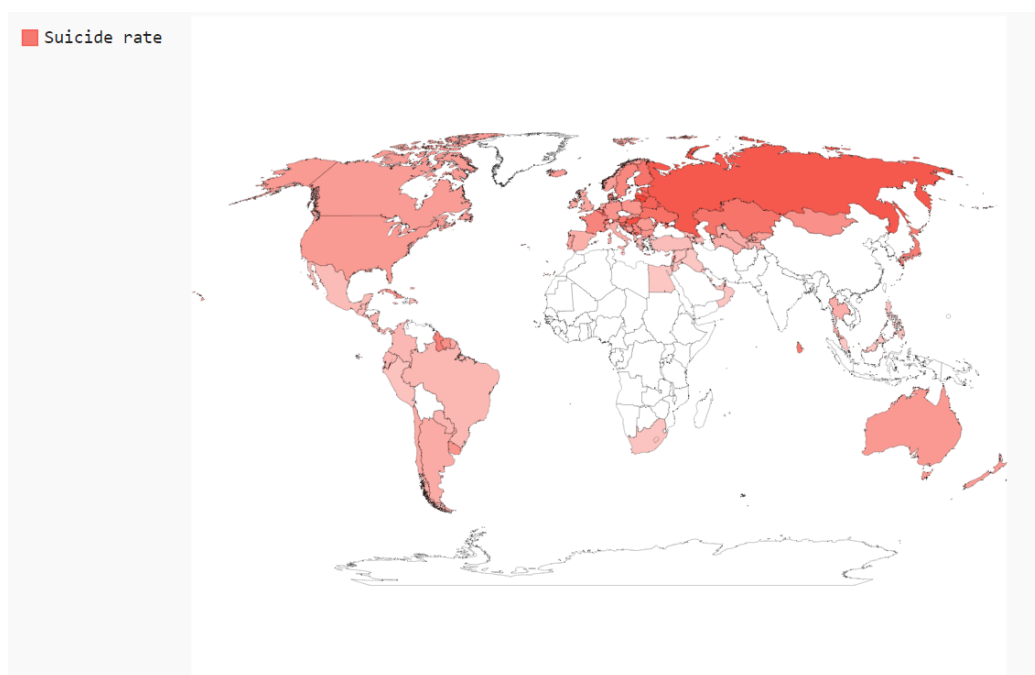Figure 6. Dynamic of death rate by income groups
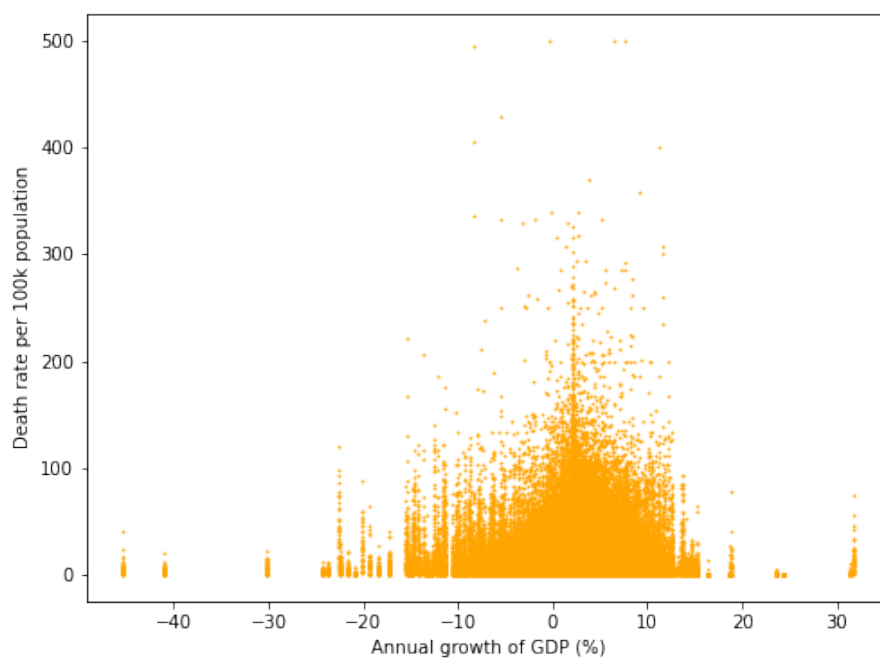
Figure 7. Suicide rate distribution on global map

Figure 8. Dependence of suicide rates and annual growth

6 Results

- The most frequent age for suicide is over 80 and 45-70 for women specifically. Suicide mortality of men is steadily increasing with age.

- Men commit suicide 197 % more often accordingly to the global data. A part of this effect can be explained by difference in methods of suicide men and women choose for themselves.

- Men and women suicide rate show the same tendency deviating within the margin of error.

- Despite Durkheim's hypothesis, anomie state in society is not enough to raise increase of suicides.

- There is no correlation between annual GDP growth and suicides. Even if economic welfare influence through healthcare system, inequality or unemployment, direct association is excluded.