

Proposal

Life savers

2019-11-06

The group members

Xue Jin(xj2249); Yimeng Shang(ys3298); Sibeil Liu(sl4660); Yi Liu(yl4358); Yue Lai(yl4359)

The tentative project title: Serious facts about life

The motivation for this project

Suicide is a major public health concern globally. According to WHO, about 800,000 people died by suicide in 2016, of which 79% occurred in low- and middle-income countries. Overall, suicide accounts for 1.4% of all deaths worldwide, and has become the second leading cause of death among 15-29 age group. Our group aims to summarize the general trend in global suicide rate over time, and explore possible influential factors related to suicide, including age, gender, country and socioeconomic status. Every suicide is tragic but also preventable. Therefore, we hope to arouse public awareness of suicide prevention and provide illustrative background information for further research and development of preventive strategies.

The intended final products

In the end, we will finish a report, in which we will discuss the trend of suicide over the recent 30 years worldwide and the association between potential variables. Besides, we will create a webpage and a narrated screencast for our project.

The anticipated data sources

We plan to utilize a compiled dataset pulled from four other datasets by locations and time. In the dataset, we have HDI (Human development index) from the United Nations Development Program (2018), GDP per year from World Bank (2018), and suicide information from Suicide in the Twenty-First Century [dataset] and World Health Organization Suicide prevention (2018). We will analyze the dataset to find signals and draw conclusions.

The planned analyses / visualizations / coding challenges

The expected analysis include three parts. First, a general description of data would be applied, including mean, range, variance, and other statistical measurements of the rates. Distribution of suicide rates among countries, genders, years would be shown, thus the “dangerous” group with higher suicide rates can be identified. According to the year-rate results, the trend of suicide rate could also be detected globally and domestically. Second, doing proportion tests is necessary when we want to know if there are significant differences between groups. Third, linear regression methods can be used to explore the potential relationship between suicide rates and other variables.

We use histogram or boxplot to show the rate distribution. Bar plots are to compare rates in groups. Line charts would be utilized to show the general trends and linear relationship. Tables are to show the statistical values in each proportion test.

How to find the most proper ways to visualize data would be the biggest challenge.

The planned timeline

TIME	TASKS
1st, Nov - 7th, Nov	Reading scientific papers; written proposal document; group members works assignment
8th, Nov - 13th, Nov	Scraping and cleaning the data; initial data exploration
14th, Nov - 15th, Nov	Project review meeting
16th, Nov - 26th, Nov	Exploratory analysis; Website establishment
27th, Nov - 4th, Dec	Written report giving detailed project description; further analysis
5th, Dec - 9th, Dec	Preparation for presentation