Proposal

Life savers

2019-11-28

The group members

Xue Jin(xj2249); Yimeng Shang(ys3298); Sibei 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. In 2016, about 800,000 people died by suicide. Overall, suicide has become the second leading cause of death among 15-29 age group. According to CDC, the annual U.S. suicide rate increased 24% between 1999 and 2014. In this project, our goal aims to summarize the suicide rate in 50 states in the USA, and explore possible influential variables related to suicide, including age, gender, GDP, education, gun ownership, weather and drug. 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 association between potential variables and suicide rate in 50 states in the USA. 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 states. In the dataset, we have Number of Guns Ranked by State from the ATF, Educational Attainment by Metropolitan Area, suicide information from . We will analyze the dataset and draw conclusions.

The planned analyses / visualizations / coding challenges

The expected analysis include three parts. First, a general distribution of suicide among countries, genders and age over time would be shown. Secondly, focusing on the data in America, regression methods would be used to explore the aasociation between suicide rates and other variables. Finally, we would construct a predictive model for suicide rate in America.

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. Shiny is employed to construct interactive graphics for suicide among groups.

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