Suicide Data Analysis Project Proposal

Stat Stars March 29th, 2019

Section 1. Introduction

The introduction should introduce your general research question and your data (where it came from, how it was collected, what are the cases, what are the variables, etc.).

This dataset provides statistics for suicides worldwide between 1985 and 2016. It has data that was collected via compilation of data from the UN Development Program, the World Bank, Kaggle, and the World Health Organization. The purpose of compiling the data was "to find signals correlate to increased suicide rates among different cohorts globally, across the socio-economic spectrum" (Kaggle). Each occurrence provided represents the country, year, sex, age (in groups), number of suicides, population, suicide rate for 100 thousand people, country-year, gdp for the year (in USD), the gdp per capita (in USD), and the generation of a single suicide. This will allow us to look at a single occurrence, but also group based on incomes, gender, year, country, or some combination of the variables.

Our overarching research question guiding our investigation is:

do location, year, and economic situation contribute to suicide rates?

This research question will allow us to see where and when economic changes, like the crash in 2008 affected the suicide rate, and for how long. This is an important investigation to conduct because understanding the relationship between economics and mental health is important, given that so much of society is attached to and operated with the purpose of earning more money. Furthermore, it will allow us to see certain groups-like men or women, high or low socioeconomic status were affected more or less so than other groups. This will allow us to understand the impact economic changes has on the metal health and suicide rates. It will also be interesting to see the progression of changes in suicide rates and how a person of a certain intersection of identities is affected differently. We will be able to use linear modelling to make limited predictions of changes in suicide rates if a crash were to ocurr. We are purposely leaving the research question broad, so that we can compare a variety of the statistics and not become too focused too early on in the investigation.

We acknowledge the pitfalls of the data. For instance, there is not detailed collection information such as method as it is a compilation of other data. Furthermore, it is known that suicides sometimes go falsely reported or unreported at all- for instance, suicides have been occasionally reported as farming or hunting accidents- so we will be conscious of this as we conduct the investigation. It has also come to our attention that not every nation is represented, which will limit our ability to compare truly worldwdide, but we will be able to prepare nation to nation and regions to region.

The data can be found here:

https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016

Section 2. Data analysis plan

Section 3. Data