Correlation of Location, Year, Economic Status, and Demographic Information with Suicide Rate

STAT 420, Summer 2020, Kun Hu, Madeline Old, Yuwen Xiang, Patrick Willhalm

07/17/2020

- Description of data file
- Background on the data set
- Statement of Interest

Description of data file

This file compares the number of suicides per 100,000 across categories stratified by age group, sex, country, and year. There are 27820 observations across 12 variables. Variables include country, year, sex, age group, number of suicides, suicide rate, Human Development Index (HDI) and Gross Domestic Product (GDP).

Background on the data set

This dataset is a compilation of data from the UN Development Program, World Bank, and the World Health Organization. https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016 (https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016)

Statement of Interest

Suicide is unique as a public health issue in that it is 100% preventable. Stopping deaths from cancer requires developing new cancer drugs, but suicide is a social and economic rather than technical problem. Here we intend to find the correlation among different social status (includes year, sex and age) and the economic situation with the suicide rate with an eye toward suicide prevention. Depending on social status, we may find some causal links with suicide risk.

```
suicide <- read.csv('master.csv') %>%
  as_tibble(.)
summary(suicide)
```

```
##    country              year          sex               age
## Length:27820       Min.   :1985   Length:27820      Length:27820
## Class :character   1st Qu.:1995   Class :character  Class :character
## Mode  :character   Median :2002   Mode  :character  Mode  :character
##                    Mean   :2001
##                    3rd Qu.:2008
##                    Max.   :2016
##
##   suicides_no      population       suicides.100k.pop country.year
## Min.   :    0   Min.   :     278   Min.   :  0.00   Length:27820
## 1st Qu.:    3   1st Qu.:   97498   1st Qu.:  0.92   Class :character
## Median :   25   Median :  430150   Median :  5.99   Mode  :character
## Mean   :  243   Mean   : 1844794   Mean   : 12.82
## 3rd Qu.:  131   3rd Qu.: 1486143   3rd Qu.: 16.62
## Max.   :22338   Max.   :43805214   Max.   :224.97
##
##   HDI.for.year   gdp_for_year....  gdp_per_capita....  generation
## Min.   :0       Length:27820      Min.   :    251    Length:27820
## 1st Qu.:1       Class :character  1st Qu.:   3447    Class :character
## Median :1       Mode  :character  Median :   9372    Mode  :character
## Mean   :1                         Mean   :  16866
## 3rd Qu.:1                         3rd Qu.:  24874
## Max.   :1                         Max.   : 126352
## NA's   :19456
```

```
suicide
```

```
## # A tibble: 27,820 x 12
##    country   year sex    age    suicides_no population suicides.100k.p…
##    <chr>    <int> <chr> <chr>        <int>      <int>            <dbl>
##  1 Albania   1987 male  15-2…           21     312900             6.71
##  2 Albania   1987 male  35-5…           16     308000             5.19
##  3 Albania   1987 fema… 15-2…           14     289700             4.83
##  4 Albania   1987 male  75+ …            1      21800             4.59
##  5 Albania   1987 male  25-3…            9     274300             3.28
##  6 Albania   1987 fema… 75+ …            1      35600             2.81
##  7 Albania   1987 fema… 35-5…            6     278800             2.15
##  8 Albania   1987 fema… 25-3…            4     257200             1.56
##  9 Albania   1987 male  55-7…            1     137500             0.73
## 10 Albania   1987 fema… 5-14…            0     311000             0
## # … with 27,810 more rows, and 5 more variables: country.year <chr>,
## #   HDI.for.year <dbl>, gdp_for_year.... <chr>, gdp_per_capita.... <int>,
## #   generation <chr>
```