# Suicides in the World

Data Analysis

# Dataset Overview

- The dataset used is Suicides between 1985 and 2015 around the world.
- Dataset collected by www.kaggle.com

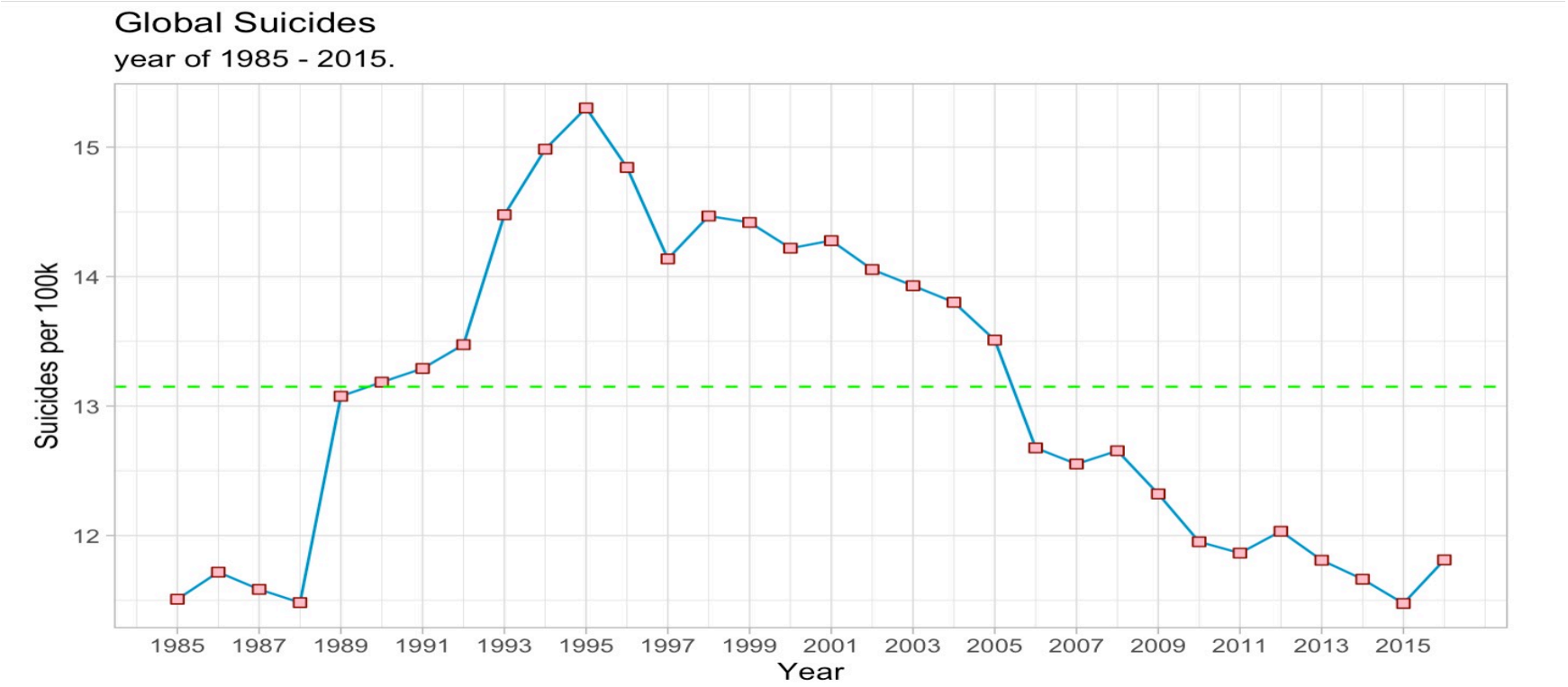
- Input variables : country, year, sex, age, suicides_no, population,

suicides/100k.pop, country-year, HDI_for_year, gdp_for_year ($), gdp_per_capita ($) and generation

# Summary statistics

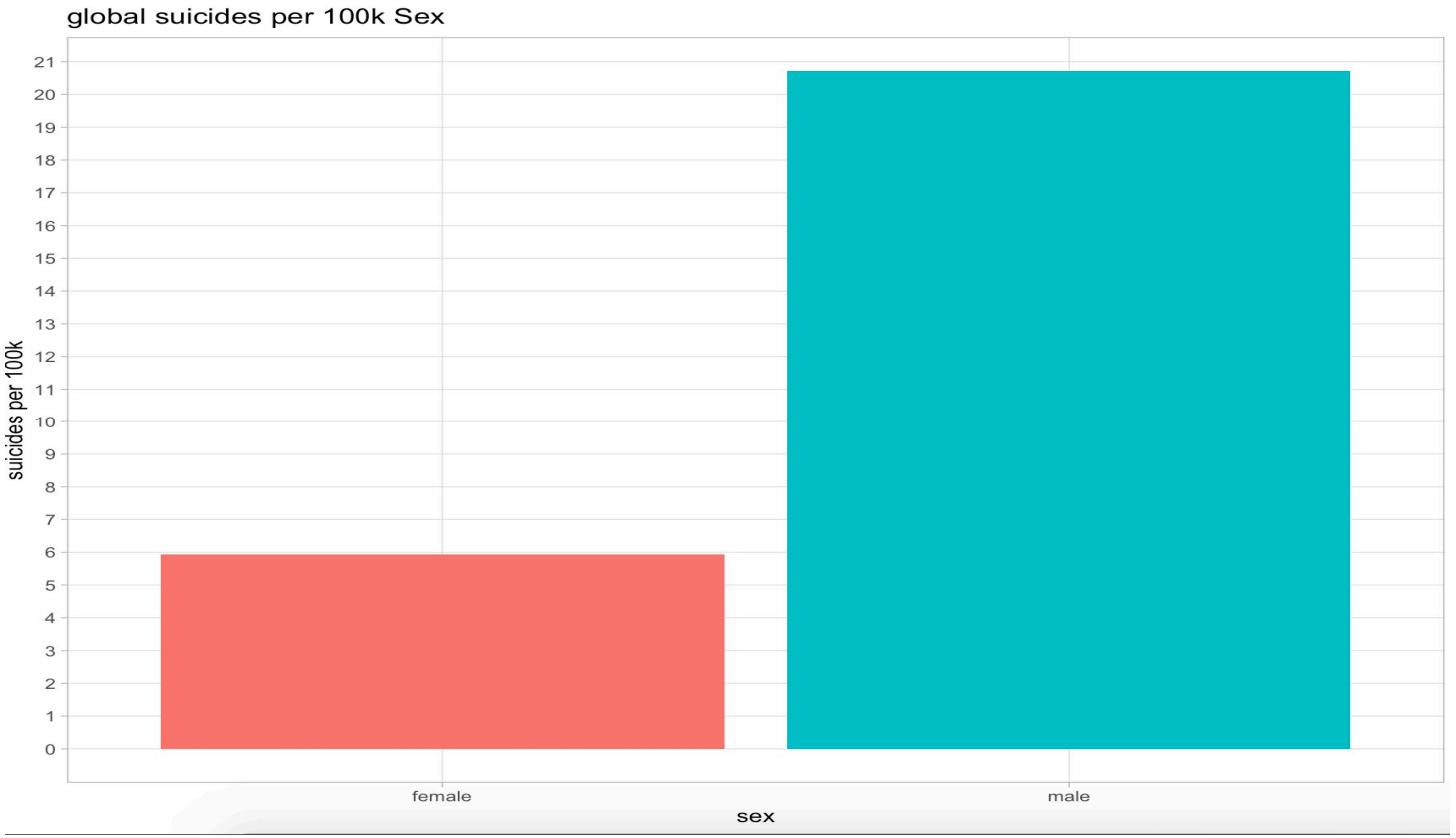
```
  country              year          sex                age             suicides_no       population
Length:27820       Min.   :1985   Length:27820       Length:27820       Min.   :    0.0   Min.   :     278
Class :character   1st Qu.:1995   Class :character   Class :character   1st Qu.:    3.0   1st Qu.:   97498
Mode  :character   Median :2002   Mode  :character   Mode  :character   Median :   25.0   Median :  430150
                   Mean   :2001                                         Mean   :  242.6   Mean   : 1844794
                   3rd Qu.:2008                                         3rd Qu.:  131.0   3rd Qu.: 1486143
                   Max.   :2016                                         Max.   :22338.0   Max.   :43805214

suicides.100k.pop country.year        HDI.for.year   gdp_for_year....   gdp_per_capita....  generation
Min.   :  0.00    Length:27820       Min.   :0.483   Length:27820       Min.   :   251     Length:27820
1st Qu.:  0.92    Class :character   1st Qu.:0.713   Class :character   1st Qu.:  3447     Class :character
Median :  5.99    Mode  :character   Median :0.779   Mode  :character   Median :  9372     Mode  :character
Mean   : 12.82                       Mean   :0.777                      Mean   : 16866
3rd Qu.: 16.62                       3rd Qu.:0.855                      3rd Qu.: 24874
Max.   :224.97                       Max.   :0.944                      Max.   :126352
                                     NA's   :19456
```

It is the global average suicide rate from 1985 - 2015:

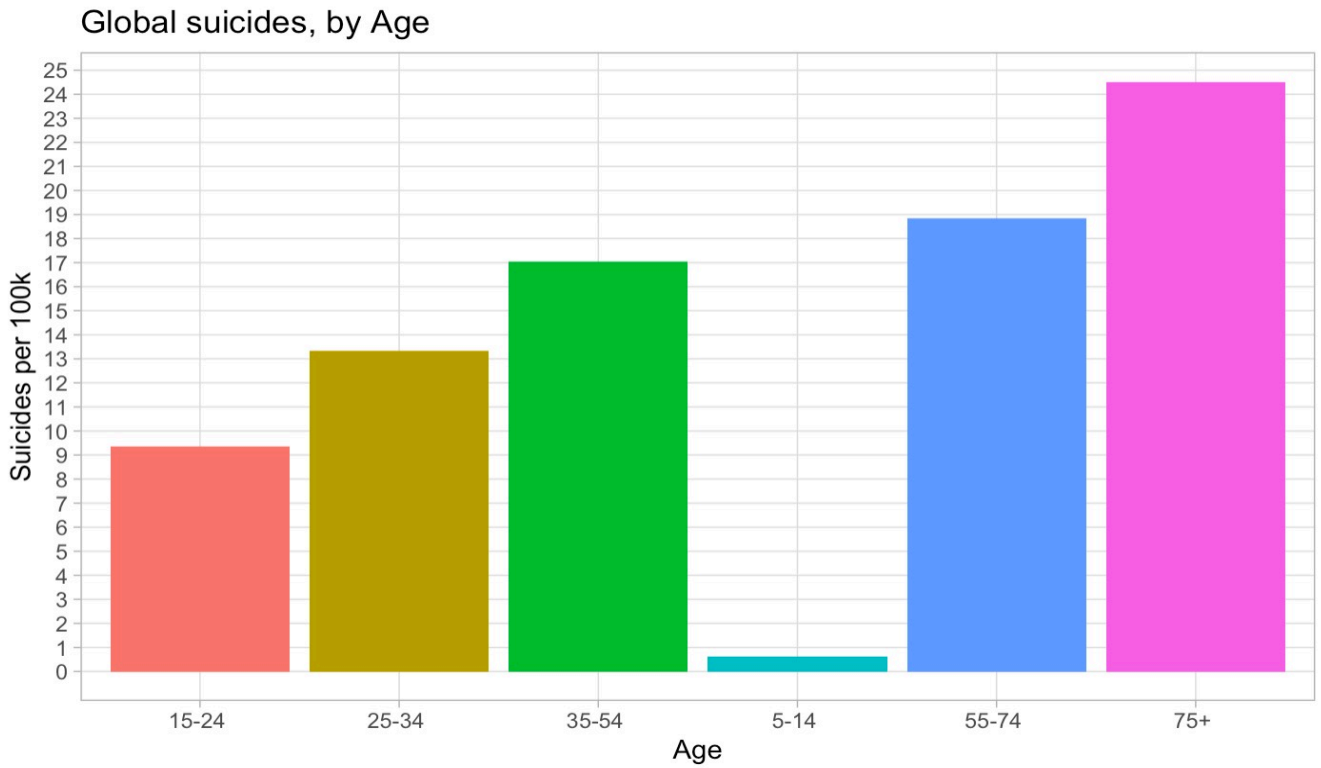



suicide decreasing after year of 1995

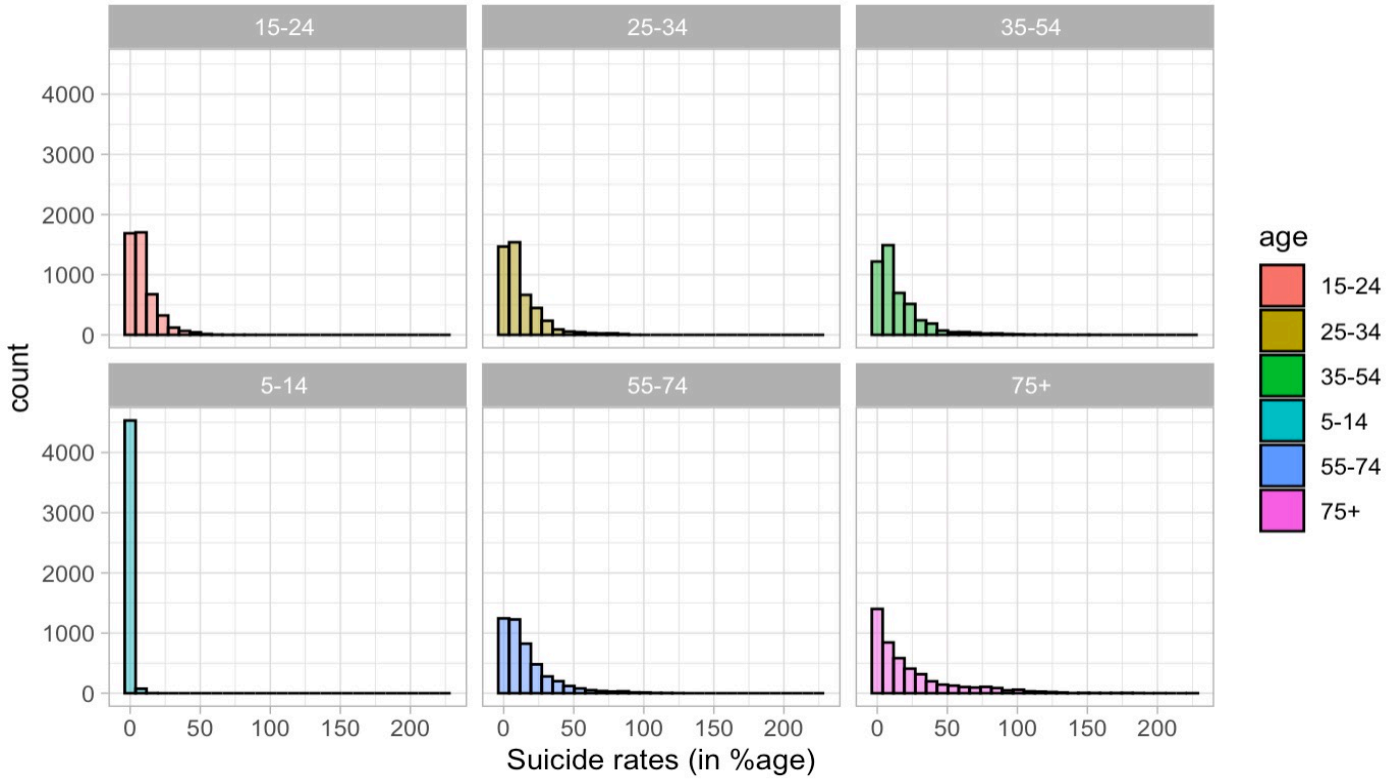
Suicide rates are mostly increased by men population rather than woman



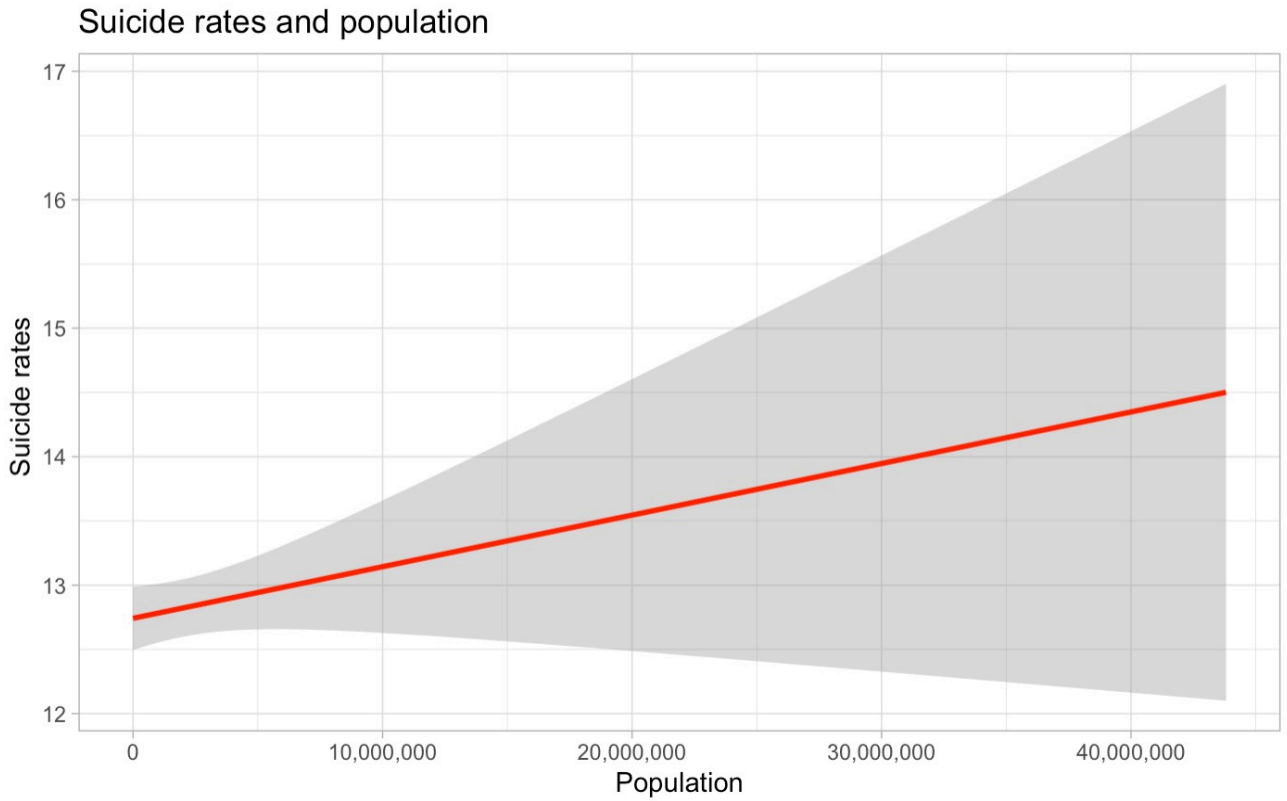
Globally, the likelihood of suicide increases with age



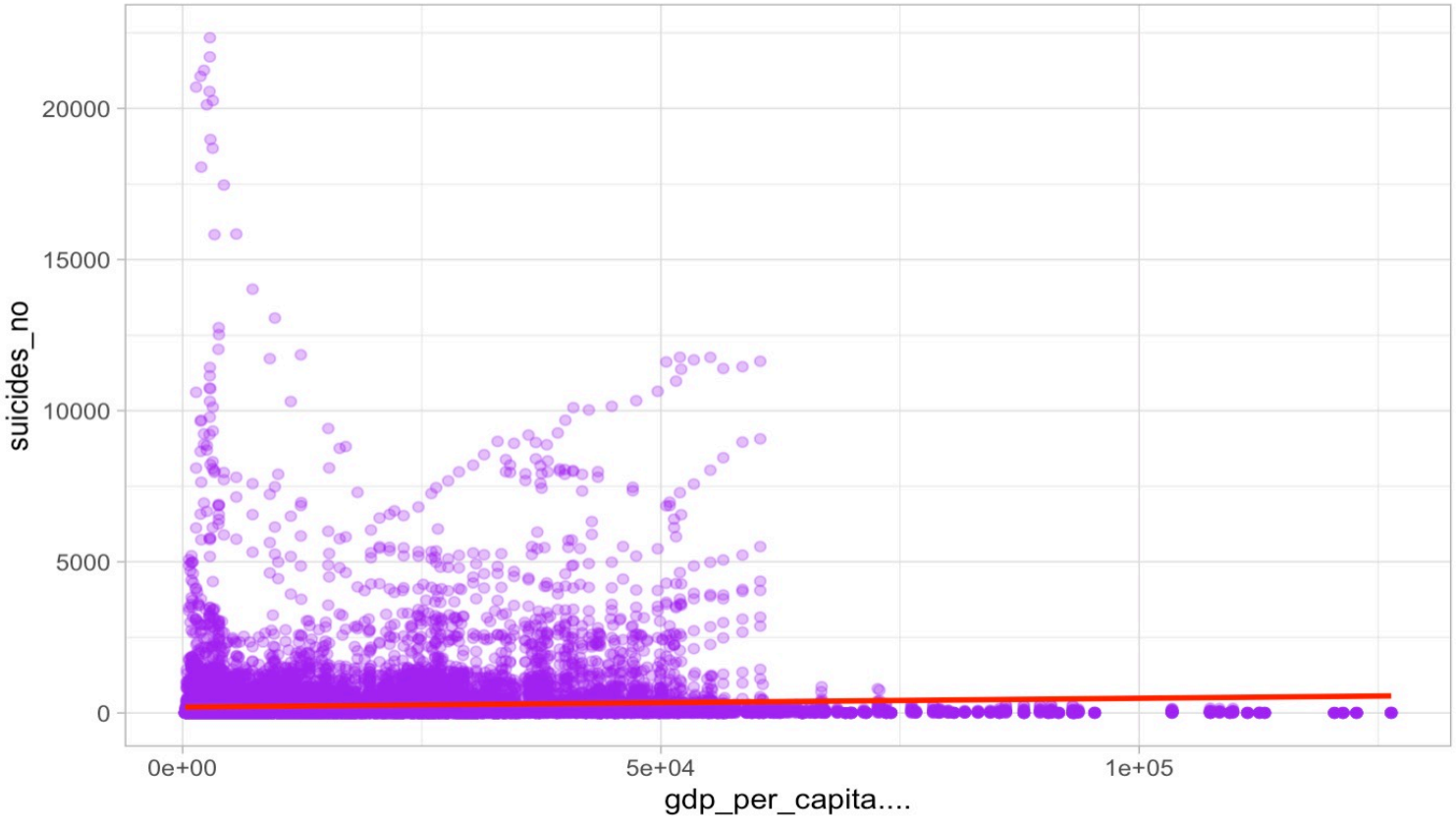
Distribution of suicide rates across age groups



# Suicide rates do not depend on population



Higher standards of living do not necessarily indicate lower suicide rates.



# Add row numbers to data frame

| plot_id <int> | country <chr> | year <int> | sex <chr> | age <fctr> | suicides_no <int> | population <int> | suicides.100k.pop <dbl> | country.year <chr> |
|---|---|---|---|---|---|---|---|---|
| 1 | Albania | 1987 | male | 15-24 | 21 | 312900 | 6.71 | Albania1987 |
| 2 | Albania | 1987 | male | 35-54 | 16 | 308000 | 5.19 | Albania1987 |
| 3 | Albania | 1987 | female | 15-24 | 14 | 289700 | 4.83 | Albania1987 |
| 4 | Albania | 1987 | male | 75+ | 1 | 21800 | 4.59 | Albania1987 |
| 5 | Albania | 1987 | male | 25-34 | 9 | 274300 | 3.28 | Albania1987 |
| 6 | Albania | 1987 | female | 75+ | 1 | 35600 | 2.81 | Albania1987 |
| 7 | Albania | 1987 | female | 35-54 | 6 | 278800 | 2.15 | Albania1987 |
| 8 | Albania | 1987 | female | 25-34 | 4 | 257200 | 1.56 | Albania1987 |
| 9 | Albania | 1987 | male | 55-74 | 1 | 137500 | 0.73 | Albania1987 |
| 10 | Albania | 1987 | female | 5-14 | 0 | 311000 | 0.00 | Albania1987 |

-10 of 27,820 rows | 1-9 of 13 columns

Previous 1 2 3 4 5 6 ... 100 N

# summary(suicide_new)

```
   plot_id        country              year          sex                age        suicides_no     
Min.   :    1   Length:27820       Min.   :1985   Length:27820       15-24 :4642   Min.   :    0.0  
1st Qu.: 6956   Class :character   1st Qu.:1995   Class :character   25-34 :4642   1st Qu.:    3.0  
Median :13910   Mode  :character   Median :2002   Mode  :character   35-54 :4642   Median :   25.0  
Mean   :13910                      Mean   :2001                      5-14  :4610   Mean   :  242.6  
3rd Qu.:20865                      3rd Qu.:2008                      55-74 :4642   3rd Qu.:  131.0  
Max.   :27820                      Max.   :2016                      75+   :4642   Max.   :22338.0  

  population       suicides.100k.pop country.year        HDI.for.year   gdp_for_year....  
Min.   :     278   Min.   :  0.00    Length:27820       Min.   :0.483   Length:27820      
1st Qu.:   97498   1st Qu.:  0.92    Class :character   1st Qu.:0.713   Class :character  
Median :  430150   Median :  5.99    Mode  :character   Median :0.779   Mode  :character  
Mean   : 1844794   Mean   : 12.82                       Mean   :0.777                     
3rd Qu.: 1486143   3rd Qu.: 16.62                       3rd Qu.:0.855                     
Max.   :43805214   Max.   :224.97                       Max.   :0.944                     
                                                        NA's   :19456                     
gdp_per_capita....  generation       
Min.   :   251     Length:27820      
1st Qu.:  3447     Class :character  
Median :  9372     Mode  :character  
Mean   : 16866                       
3rd Qu.: 24874                       
Max.   :126352                       
```

# MODEL BUILDING

```
fit <- lm(population~.,data = training)
summary(fit)
```

```
country.yearBelgium2012                    -8.552e+04  1.732e+06  -0.049 0.960632
country.yearBelgium2013                    -1.598e+05  1.107e+06  -0.144 0.885274
country.yearBelgium2014                            NA         NA      NA       NA
country.yearBelize1990                      2.353e+06  1.848e+07   0.127 0.898666
country.yearBelize1995                      2.495e+06  1.463e+07   0.171 0.864590
country.yearBelize2000                      1.438e+06  1.080e+07   0.133 0.894064
country.yearBelize2005                      1.033e+06  6.957e+06   0.148 0.881991
country.yearBelize2010                      3.658e+05  3.176e+06   0.115 0.908302
country.yearBelize2011                      3.720e+05  2.407e+06   0.155 0.877197
country.yearBelize2012                      3.206e+05  1.684e+06   0.190 0.849021
country.yearBelize2013                      7.492e+04  1.042e+06   0.072 0.942660
country.yearBelize2014                             NA         NA      NA       NA
country.yearBosnia and Herzegovina2011     -9.848e+05  2.448e+06  -0.402 0.687508
country.yearBosnia and Herzegovina2014             NA         NA      NA       NA
country.yearBrazil1985                     -1.288e+06  2.232e+07  -0.058 0.953982
country.yearBrazil1990                      1.874e+04  1.848e+07   0.001 0.999191
country.yearBrazil1995                      3.736e+06  1.464e+07   0.255 0.798593
country.yearBrazil2000                      6.282e+05  1.080e+07   0.058 0.953598
country.yearBrazil2005                      6.554e+05  6.960e+06   0.094 0.924988
country.yearBrazil2010                      2.719e+06  3.199e+06   0.850 0.395368
country.yearBrazil2011                      3.034e+06  2.433e+06   1.247 0.212523
 [ reached getOption("max.print") -- omitted 1299 rows ]
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1489000 on 5168 degrees of freedom
  (13595 observations deleted due to missingness)
Multiple R-squared:  0.8773,    Adjusted R-squared:  0.8604
F-statistic: 52.04 on 710 and 5168 DF,  p-value: < 2.2e-16
```

fit2 <- lm(population ~ year + suicides_no + sex,data = training)
summary(fit2)

```
Call:
lm(formula = population ~ year + suicides_no + sex, data = training)

Residuals:
      Min         1Q     Median         3Q        Max
-44794106  -1257970   -739244   -195529  32872445

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.368e+06  5.172e+06  -1.231    0.218
year         3.960e+03  2.584e+03   1.533    0.125
suicides_no  2.825e+03  2.523e+01 111.953   <2e-16 ***
sexmale     -7.884e+05  4.420e+04 -17.835   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3052000 on 19470 degrees of freedom
Multiple R-squared:  0.3917,    Adjusted R-squared:  0.3916
F-statistic:  4179 on 3 and 19470 DF,  p-value: < 2.2e-16
```

# Conclusion

1-Suicide are higher among men than women
2-Higher standards of living do not necessarily indicate lower suicide rates.
3-Suicide numbers were decreasing up till 2010, after which they started rising again.
4-Middle-aged people are more likely to commit suicides globally.
5-The year, sex, and population are strong indicators when used to predict the number of suicides for a country.
6-There is a weak positive relationship between a countries GDP (per capita) and suicide rate

# Functions

| Functions | Usage |
|---|---|
| read.cvs() | Read csv file |
| head() | Show the first part of the data frame |
| dim() | Get the dimension of the data frame |
| sum() | Count the missing value in the data frame |
| subset() | Select variables in the column when cleaning the data |
| str() | Display the structure and output the information of each variables after cleaned the data |
| as. numeric () | This generic function is to convert the data frame column from descriptive values to numeric values |
| sum() | Provide summary of the dataset |
| stat.desc() | Get the descriptive statistics of the dataset |
| boxplot() | For data visualization of the giving data values |
| bagplot() | |
| ggplot | |
| filter() | Find out the usual values |
| count() | Show the count of the categorical variable and range |
| set. Seed() | Generate 10k diamond samples |
| ggpairs() | Build scatterplot matrix |
| colmeans() | Calculate the overall mean for each variables |
| predict() | Make predictions from modules of model-fitting functions |
| exp() | For exponential model fitting |