# ISQS 6350 Multivariate Analysis: Introduction

Mikaela Pisani, Jereamy Riggs, Roger Valdez, Marcin Grzechowiak

## Dataset

The dataset used is from Gapminder World, which contains data from several sources. It contains information from different indicators about the population for each country.

The indicators considered for this study are:

| Indicator | Source |
|---|---|
| Population total | Various Sources |
| Murder total | WHO Global Burden of Disease |
| Armed forces personnel total | World Development Indicators |
| Cell Phones per 100 people | World Bank |
| Fertility rate total | Various Sources |
| Urban Population | World Bank |
| Suicide total | WHO Global Burden of Disease |
| Life expectancy | Various Sources |
| Corruption Perception Index | Transparency International |
| Internet users (% of pop) | World Bank |
| Child Mortality | Various Sources |
| Income per person | Various Sources |
| Sex Ratio | UN Population Division |
| Investment (as % of GDP) | World Bank |
| Inequality Index (GINI) | World Bank |

All indicators' values come from the year 2016. The only exception is Sex Ratio. We assume that sex ratio did not change significantly between 2015 and 2016, and using data from 2015 is a good approximation of year 2016.

In the following table it is provided the meaning for each of the variables used in the study.

| Name | Shortcut | Meaning | Units |
|---|---|---|---|
| Population total | pop_total | Total population | total |
| Murder per person | murder_pp | Calculated from the total divided by the population | per person |
| Armed forces personnel total | armed_pp | Calculated from the total divided by the population | total |
| Cell Phones per 100 people | phones_p100 | Mobile cellular telephone subscription | 100 people |
| Fertility rate total | children_p_woman | Children per woman total fertility | total |
| Urban Population | urban_pop_tot | People that live in urban areas | total |
| Suicide total | suicide_pp | Total number of estimated deaths from self-inflicted injury. | total |
| Life expectancy | life_exp_yrs | Average number of years a child would live. | year |
| Corruption Perception Index | corruption_CPI | International score of perception of corruption. Higher values indicates less corruption. | 0-100 |
| Internet users (% of pop) | internet_%of_pop | Percentage of individuals using internet | percentage |
| Child Mortality | child_mort_p1000 | Death of children under 5 per 1000 born. | 100 born |
| Income per person | income_per_person | GDP/Capita, inflation adjusted $ | Per person |

| Sex Ratio* | sex_ratio_p100 | Male/Female per 100 among all age groups | Per 100 |
|---|---|---|---|
| Investment (as % of GDP) | investments_per_ofGDP | Gross capital formation. Includes fixed assets plus net changes. | percentage |
| Inequality Index (GINI) | gini | Gini shows income inequality in a society. Higher is more inequality | 0-100 |

**\*Sex Ratio -** is the only variable which data comes from 2015.

# Motivation

Our motivation is focused on understanding and comparing countries from the economical, social and government perspectives. This study can be useful for many people. For instance, for a person that wants to move to another country, it might be important to consider these aspects to make the decision. In addition, it could be useful for the World Health Organization when determining which countries it should support in each field. Finally, firms that want to open an office in another country might be interested in this data as careful considerations are made when making such a large business decision.

# Analysis

Provided a basic analysis to show the correlation between these variables.

# Correlation matrix

```
                     pop_total     murder_pp     armed_pp  phones_p100 children_p_woman life_exp_yrs   suicide_pp urban_pop_tot sex_ratio_p100
pop_total          1.000000000 -0.037661844 -0.08257071 -0.030737162      -0.06792386   0.01968400  0.037963223   0.955960030     0.002878013
murder_pp         -0.037661844  1.000000000 -0.05653347 -0.004641173      -0.04955301  -0.17813308  0.071390276  -0.017070251    -0.133415953
armed_pp          -0.082570713 -0.056533472  1.00000000 -0.052325734      -0.21990934   0.15358070  0.029959782  -0.073863049     0.049442803
phones_p100       -0.030737162 -0.004641173 -0.05232573  1.000000000      -0.57924808   0.58825669  0.183047365   0.009879933     0.217675306
children_p_woman  -0.067923864 -0.049553008 -0.21990934 -0.579248084       1.00000000  -0.79582382 -0.374130956  -0.117472514    -0.026265516
life_exp_yrs       0.019684002 -0.178133080  0.15358070  0.588256689      -0.79582382   1.00000000  0.112746592   0.085107642     0.097276578
suicide_pp         0.037963223  0.071390276  0.02995978  0.183047365      -0.37413096   0.11274659  1.000000000   0.050369500    -0.213004679
urban_pop_tot      0.955960030 -0.017070251 -0.07386305  0.009879933      -0.11747251   0.08510764  0.050369500   1.000000000    -0.005601667
sex_ratio_p100     0.002878013 -0.133415953  0.04944280  0.217675306      -0.02626552   0.09727658 -0.213004679  -0.005601667     1.000000000
corruption_CPI    -0.032233528 -0.194551831 -0.08379623  0.470745525      -0.54010874   0.65618003  0.302877179   0.010410296     0.081579427
internet_%of_pop  -0.037835193 -0.131541594  0.15721509  0.626937578      -0.78245963   0.83574182  0.287299790   0.030789748     0.163619294
child_mort_p1000  -0.004954457  0.025606571 -0.21394272 -0.578295754       0.86908523  -0.86468550 -0.302838676  -0.065634272    -0.058760338
income_per_person -0.039039649 -0.235990910  0.11584562  0.487965867      -0.53227877   0.66947060  0.150351403   0.016528445     0.449280228
investments_per_ofGDP 0.130845426 -0.093828628  0.06249013  0.080547461      -0.03907401   0.04693724  0.007824662   0.121516061     0.199690189
gini              -0.021992690  0.513041571 -0.06991572 -0.157018856       0.25966550  -0.41195507 -0.186486206  -0.005817946     0.036063605


                  corruption_CPI internet_%of_pop child_mort_p1000 income_per_person investments_per_ofGDP         gini
pop_total           -0.032233528     -0.03783519     -0.004954457       -0.03903965          0.130845426 -0.021992690
murder_pp           -0.194551831     -0.13154159      0.025606571       -0.23599091         -0.093828628  0.513041571
armed_pp            -0.083796229      0.15721509     -0.213942719        0.11584562          0.062490134 -0.069915718
phones_p100          0.470745525      0.62693758     -0.578295754        0.48796587          0.080547461 -0.157018856
children_p_woman    -0.540108735     -0.78245963      0.869085233       -0.53227877         -0.039074015  0.259665499
life_exp_yrs         0.656180033      0.83574182     -0.864685495        0.66947060          0.046937241 -0.411955067
suicide_pp           0.302877179      0.28729979     -0.302838676        0.15035140          0.007824662 -0.186486206
urban_pop_tot        0.010410296      0.03078975     -0.065634272        0.01652845          0.121516061 -0.005817946
sex_ratio_p100       0.081579427      0.16361929     -0.058760338        0.44928023          0.199690189  0.036063605
corruption_CPI       1.000000000      0.76053234     -0.580204143        0.71038032          0.001706337 -0.288280579
internet_%of_pop     0.760532338      1.00000000     -0.799317208        0.77802282         -0.034152713 -0.347160133
child_mort_p1000    -0.580204143     -0.79931721      1.000000000       -0.55781627         -0.041904197  0.336986770
income_per_person    0.710380316      0.77802282     -0.557816270        1.00000000          0.052552557 -0.259256317
investments_per_ofGDP 0.001706337    -0.03415271     -0.041904197        0.05255256          1.000000000  0.057046749
gini                -0.288280579     -0.34716013      0.336986770       -0.25925632          0.057046749  1.000000000
```

# Scatterplots matrix

(please find the correlation matrix plot in the attached file to display it in better quality)