# 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.

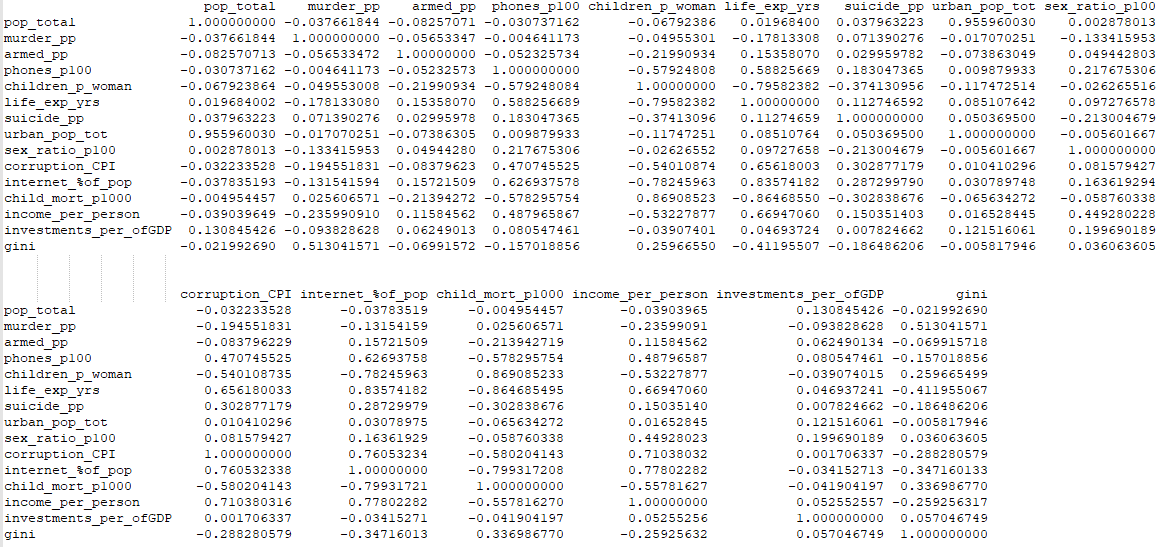
## Motivation

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



## Scatterplots matrix

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

