UCLA Extension – Data Science and Visualization

COM SCI X 450.2

Name: Regina Guzanova

Instructor: Ali

Date: 11/21/2020

What are the predictors of life expectancy around the World?

# Question

What are the most important factors/predictors influencing life expectancy around the World?

# Data

Data set belongs to the World Health Organization or WHO and United Nations websites. However, the dataset is from Kaggle. The data set contains 22 columns and 15 years of observations for 182 countries from 2000 to 2015, with a total of 512 Developed countries and 2370 Developing countries. The minimum life expectancy is 36.3 years, and the maximum life expectancy of 89 years. Alcohol consumption ranges from 0.01 up to 17.87 per capita (15+) consumption (in pure alcohol liters). Other variables include health expenditures, multiple diseases, population and income composition of resources, schooling, etc.

Table 1. Data Dictionary.

|  |  |
| --- | --- |
| Country | Country |
| Year | The year 2000-2015 |
| Status | Developed or Developing status |
| Life Expectancy | Life Expectancy in age |
| Adult Mortality Rates | Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population) |
| Infant Deaths | Number of Infant Deaths per 1000 population |
| Alcohol | Alcohol, recorded per capita (15+) consumption (in liters of pure alcohol) |
| Percentage Expenditure | Expenditure on health as a percentage of Gross Domestic Product per capita(%) |
| Hepatitis B | Hepatitis B (HepB) immunization coverage among 1-year-olds (%) |
| Measles | Number of reported cases per 1000 population |
| BMI | Average Body Mass Index of the entire population |
| Under Five Deaths | Number of under-five deaths per 1000 population |
| Polio | Polio (Pol3) immunization coverage among 1-year-olds (%) |
| Total Expenditure | General government expenditure on health as a percentage of total government expenditure (%) |
| Diphtheria | Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%) |
| HIV/AIDS | Deaths per 1 000 live births HIV/AIDS (0-4 years) |
| GDP | Gross Domestic Product per capita (in USD) |
| Population | The population of the country |
| Thinness 1-19 years | Prevalence of thinness among children and adolescents for Age 10 to 19 (% ) |
| Thinness 5-9 years | Prevalence of thinness among children for Age 5 to 9(%) |
| Income Composition of Resources | Human Development Index in terms of income composition of resources (index ranging from 0 to 1) |
| Schooling | Number of years of schooling (years) |

Figure 1: [Human Development Index Composition – United Nations – Human Development Index (HDI)](http://hdr.undp.org/en/content/human-development-index-hdi).

Diagram

Description automatically generated

# Methodology

Preparing data set in Excel for analysis in R.

#Read the data set in R using a data frame named df

df <- read.csv("C:\\Users\\...\\Final project\\Life Expectancy Data.csv", header = TRUE)

#Find NA’s, get a sense of the data set values

x <- summary(df)

x

count(df$Country)

# find the significance values for a matrix of correlations

install.packages("psych")

library(psych)

#matrix of correlations

corr.test(df[4:22])

#[4:22] include only numeric values excluding country, year and development level

Figure 2. Correlation Matrix.

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Graphical user interface, text

Description automatically generated

# Conclusion

The correlation matrix in R Studio shows that Life Expectancy correlates with Adult Mortality Rates 70% (negative or reverse correlation), Income Composition of Resources 73%, and, surprisingly, with Schooling 75%. Weaker correlation with Life Expectancy shows variables such as BMI 57% and HIV AIDS 56% (negative or reverse correlation).

* Another interesting fact was that the Alcohol variable has a high correlation with Schooling 56%.
* Percentage Health Expenditure correlates with GDP 89%.
* BMI variable correlates with Schooling 56%.
* Income composition of Resources is highly correlated with Schooling 80%.

According to the United Nations, Income Composition of Resources represents Human Development Index, including Life Expectancy, Education, and Gross National Income or GNI. The higher the education level, the higher is Gross National Income, the higher is Life Expectancy. Therefore, based on the acquired data set and its further analysis, there are two main predictors for Life Expectancy around the World, namely, Adult Mortality Rates (war, diseases) and schooling (years spent on education).

# Visual format

Figure 3. R Plot - Life Expectancy vs. Number of years of Schooling

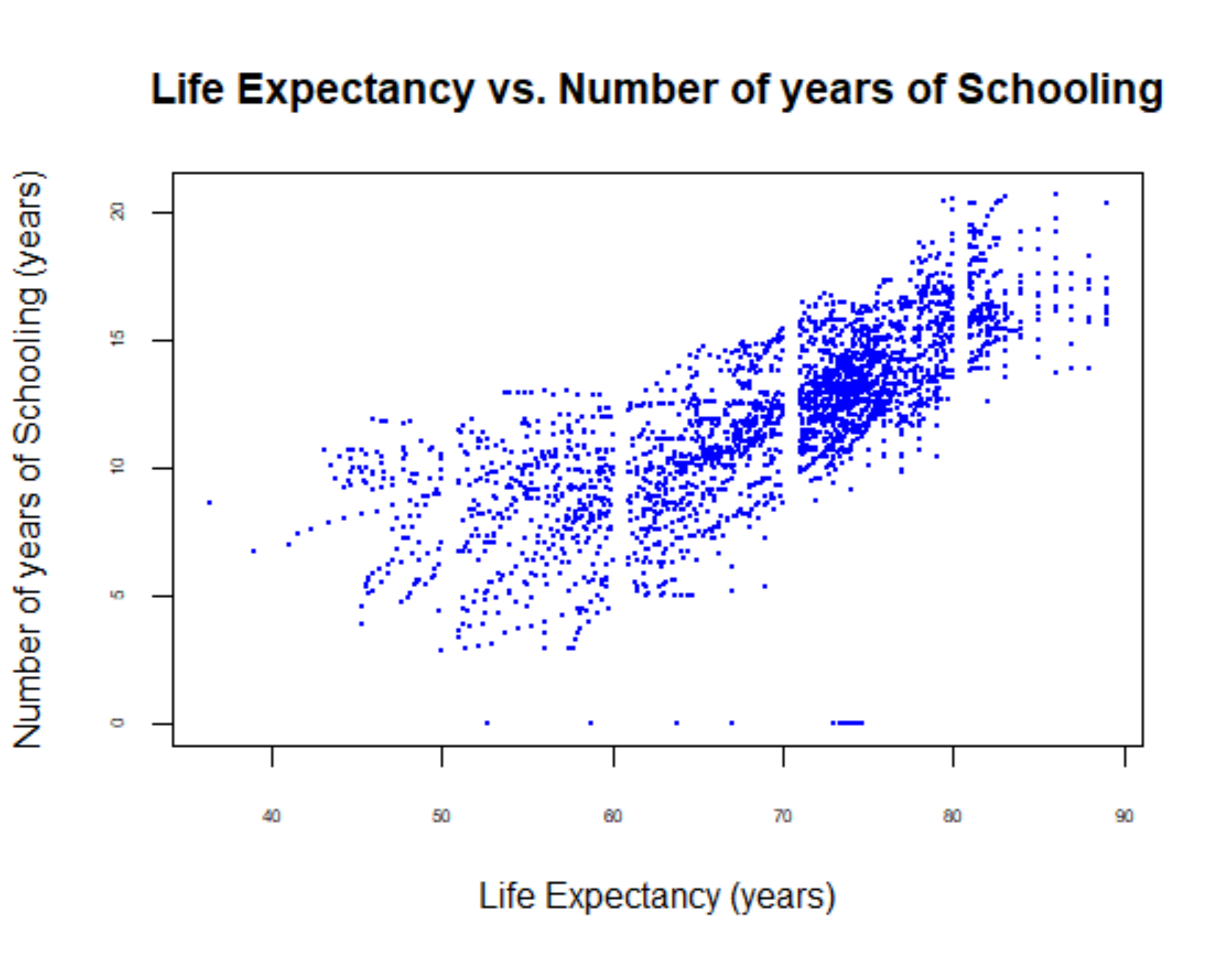


Figure 4. R QQPlot - Life Expectancy vs. Years spent in school

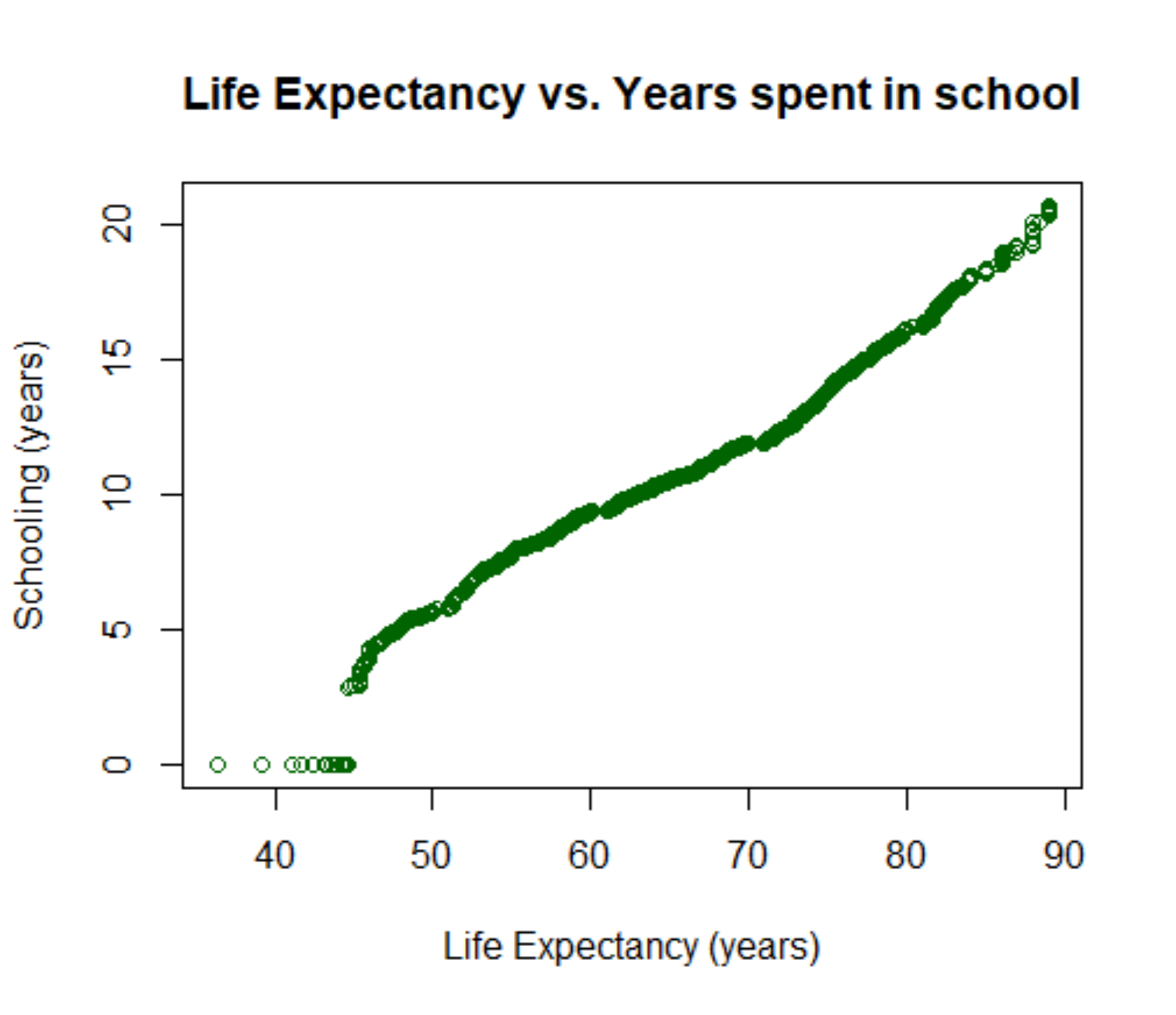


Figure 5. R Plot - Life Expectancy vs. Adult Mortality Rates

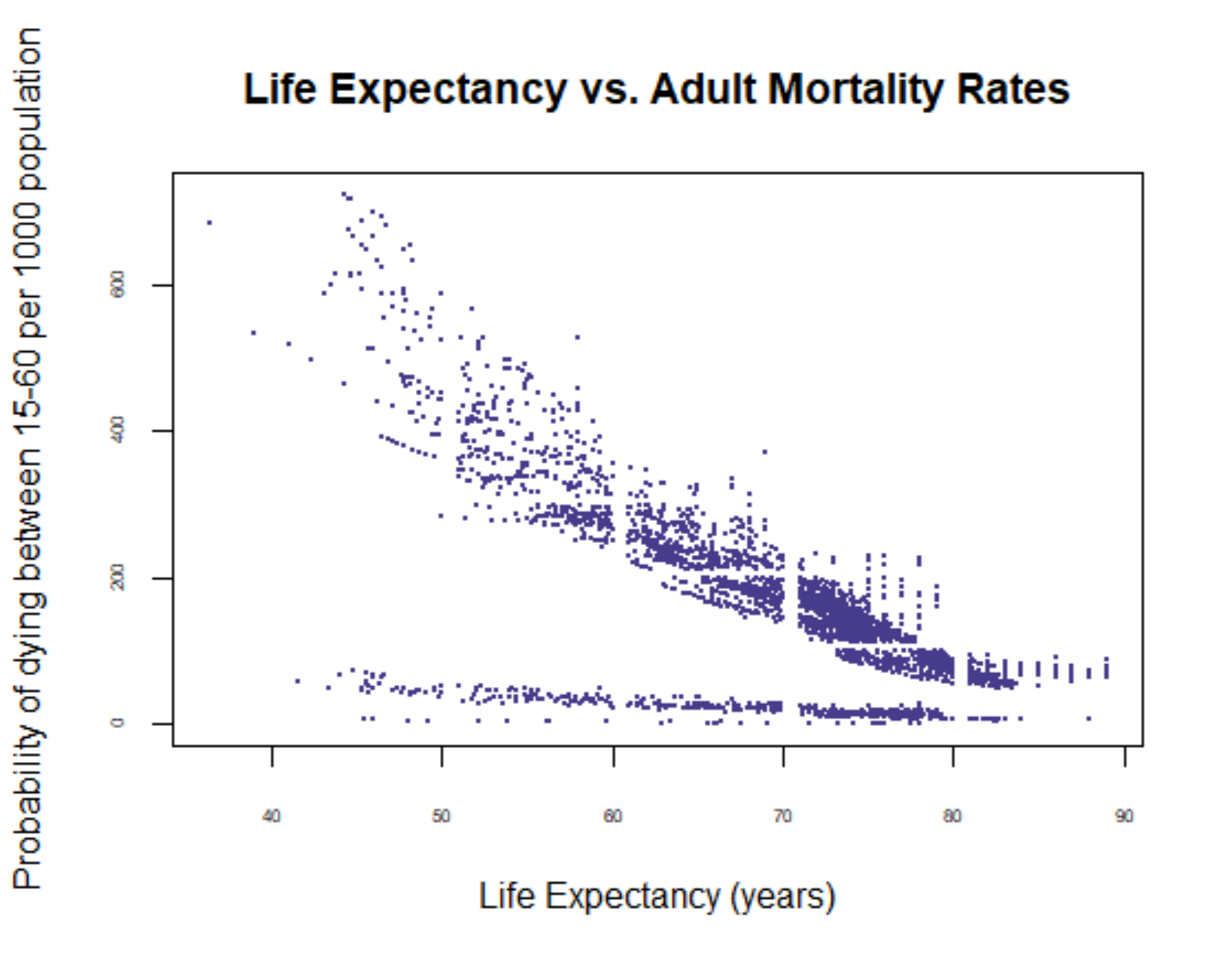


Figure 6. R Plot - Life Expectancy vs. Income Composition of Resources

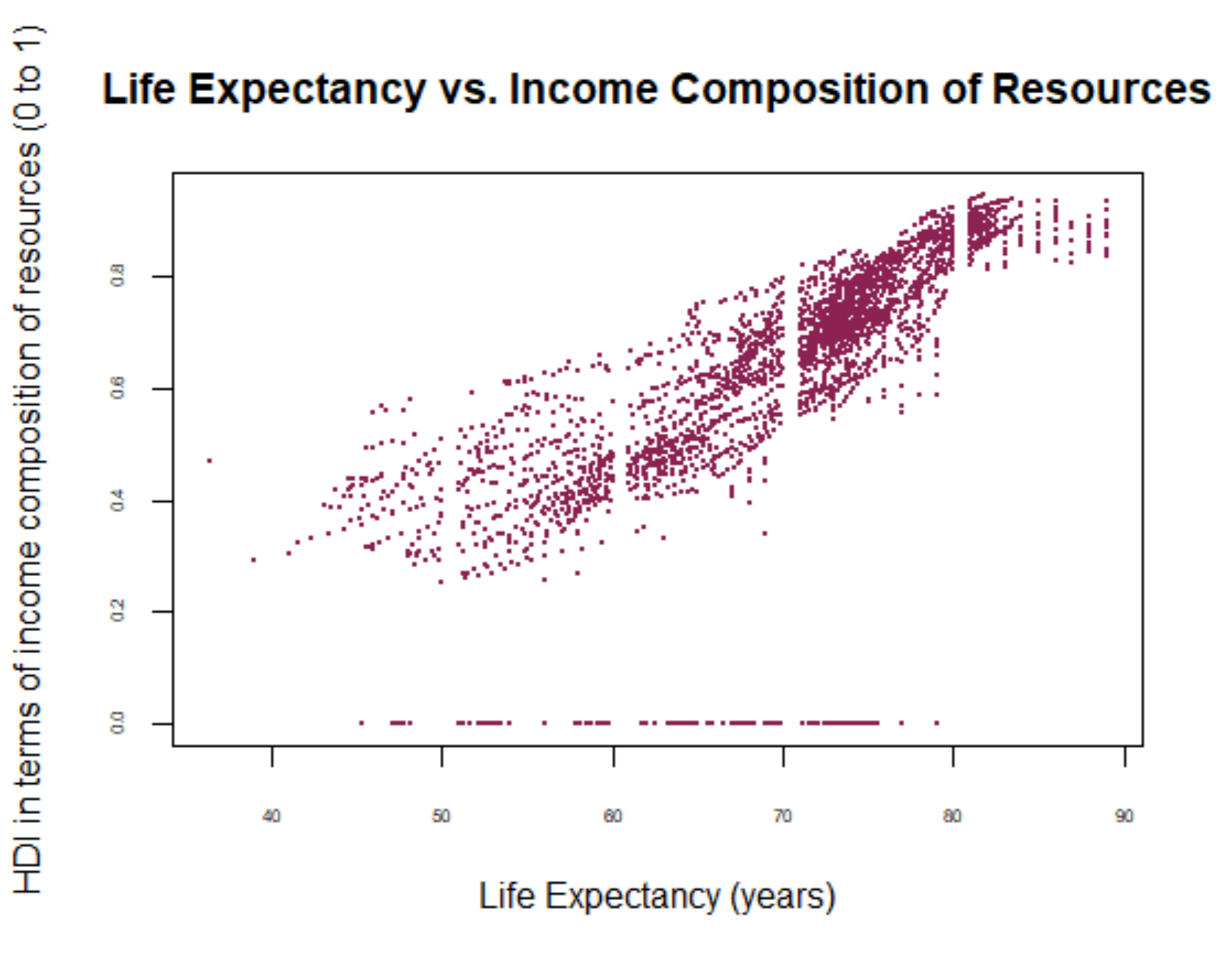
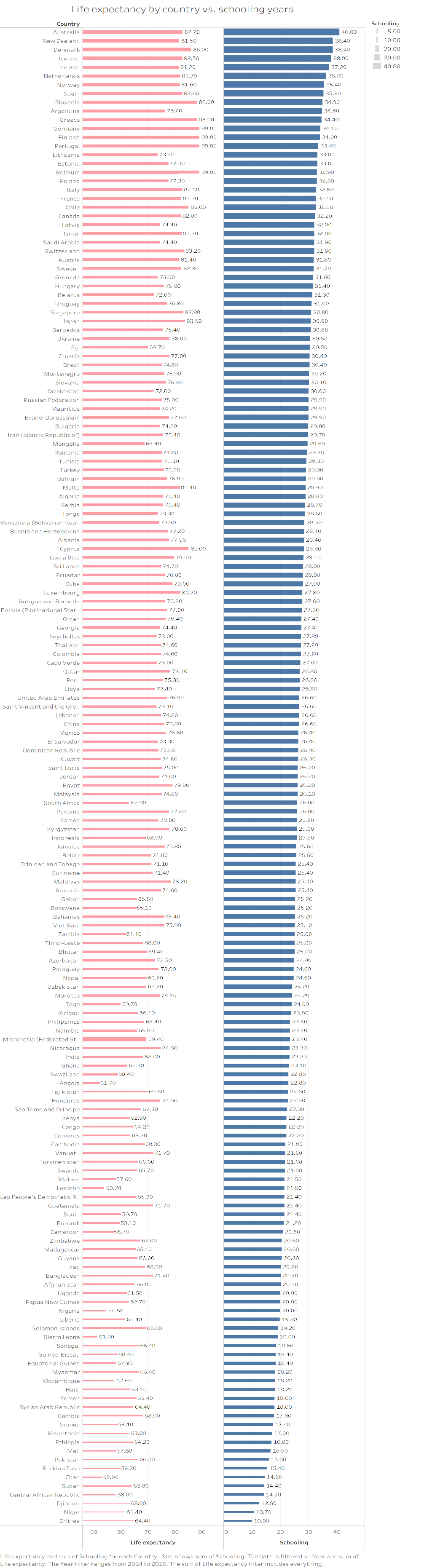
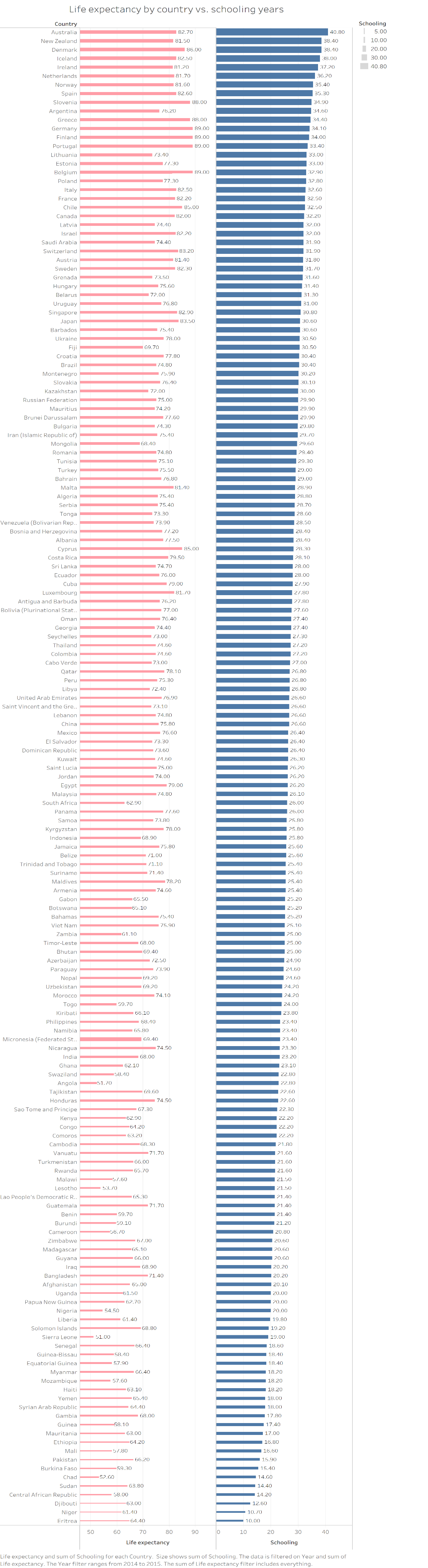


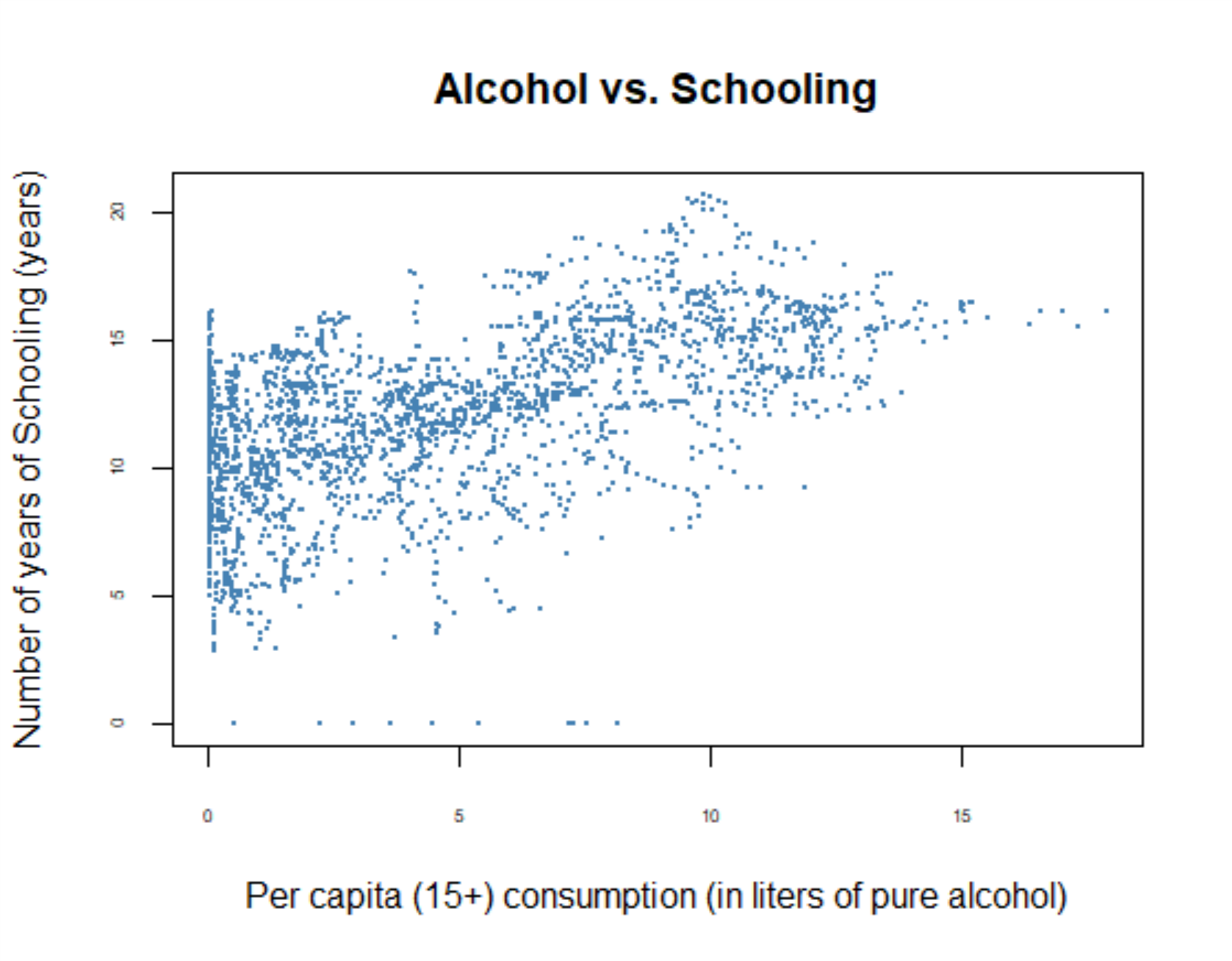
Figure 7. Tableau – Country’s Life expectancy and Schooling





# Interesting Facts

Figure 8. R Plot – Alcohol vs. Schooling



The higher the number of years spent in school increases the alcohol consumption levels. However, it is not a cause-and-effect relationship—this correlation is most likely connected to the Gross National Income of a country.