title: “HarvardX Data Science Capstone”
author: “Jorge Castro Martínez”
date: “2024-10-07”
output:
html_document:
toc: yes
toc_depth: 2
number_sections: yes
code_folding: show
highlight: tango
theme: united
pdf_document:
toc: yes
toc_depth: 2
number_sections: yes
latex_engine: xelatex
highlight: tango
urlcolor: blue

Introduction

#Although life expectancy is understood as the number of years that a person would expect to live on average from the moment of birth, it is important to note that, throughout their life, the conditions of the environment and the interaction of people have a direct impact. some factors such as demographic structure, income composition, level of immunization, among other determinants in health, environment and others.

#In this sense, the following document reveals the current situation, progress and needs in the countries that make up the OECD based on the data repository for 193 countries of the World Health Observatory for the period 2000-2015.

#For this, exploratory data analysis is developed, as the first encounter with the data, with the objective of detailing and delimiting what is behind the data, as well as exploitative visualization by identifying what type of variables are present, in addition to identifying the number of observations, their periodicity and distribution.Therefore,the first part constitutes the univariate and bivariate exploitative analysis to know the independent variables and apply learned codes for all countries.

##And finally in the second section, the application of machine learning is carried out, through different machine learning techniques such as the Principal Component Analysis,the multiple linear regression model and classification that allow applying different algorithms and predicting results for decision making based on data and with scientific processes.But in this second moment, I only apply it to the group of countries of interest, the OECD countries.

#It should be noted that this project is the final presentation for the edX course, “HarvardX PH125.9x Data Science: Capstone. Its analysis can identify strengths and the need for attention to increase the life expectancy of the population in different countries.

#The data set aims to answer the following key questions:

#Do the various initially chosen prediction factors really affect life expectancy? What are the predictive variables that really affect life expectancy? #Should a country with a lower life expectancy value (<50) increase its healthcare spending to improve its average life expectancy? #How do infant and adult mortality rates affect life expectancy? #Does life expectancy have a positive or negative correlation with eating habits, lifestyle, alcohol and different diseases, etc.? #Does life expectancy have a positive or negative relationship with alcohol consumption? #Do developed or developing countries tend to have lower life expectancy?

Methodology

#This study focuses on transformation and cleaning of the data from the life expectancy database of the World Health Organization (WHO) from the period of 2010-2015, an exploratory data analysis - EDA was carried out, applying the different codes learned in the course for univariate and bivariate analyses.

#Subsequently, the machine learning technique is applied, related to the analysis of principal components for OECD countries, with the idea of analyzing and understanding the data set and visualizing it in a better way, capturing the most important and representative information of the database of the data, by reducing its dimension and grouping into correlated variables.

#Additionally, multiple and logistic regression is carried out for the variables and countries evaluated, in order to determine which are the main determining factors for achieving life expectancy.

First part: Exploratory data analysis -EDA

The complete database is considered for all 193 countries

```
#Loading packages
#install.packages("readxl")
#installed.packages("faraway")
#install.packages("corrr")
#install.packages("ggcorrplot")
#install.packages("FactoMineR")
#install.packages("factoextra")
#install.packages("faraway")
#install.packages("corrplot")
#install.packages("psych")
#install.packages("lmtest")

#Loading libraries
library(tidyverse)
library(ggplot2)
library(readxl)
library(haven)
library(faraway)
library(readr)
library(dplyr)
library(corrr)
library(ggcorrplot)
library("FactoMineR")
library("factoextra")
library(faraway)
library(corrplot)
library(psych)
library(janitor)
```

```r
library(car)
library (lmtest)
library(tseries)

#Read the csv file
setwd("C:/Users/jorge.castro/Desktop/PROYECTOS/Life EXp")

datas <- read.table("Life Expectancy Data.csv", header= TRUE,sep="," )
```

Analysis of dataframe

```r
str(datas)

## 'data.frame':    2914 obs. of  22 variables:
##  $ Country                        : chr  "Afghanistan" "Afghanistan"
"Afghanistan" "Afghanistan" ...
##  $ Year                           : int  2015 2014 2013 2012 2011 2010
2009 2008 2007 2006 ...
##  $ Status                         : chr  "Developing" "Developing"
"Developing" "Developing" ...
##  $ Life.expectancy                : num  65 59.9 59.9 59.5 59.2 58.8
58.6 58.1 57.5 57.3 ...
##  $ Adult.Mortality                : int  263 271 268 272 275 279 281
287 295 295 ...
##  $ infant.deaths                  : int  62 64 66 69 71 74 77 80 82 84
...
##  $ Alcohol                        : num  0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.03 0.02 0.03 ...
##  $ percentage.expenditure         : num  71.3 73.5 73.2 78.2 7.1 ...
##  $ Hepatitis.B                    : int  65 62 64 67 68 66 63 64 63 64
...
##  $ Measles                        : int  1154 492 430 2787 3013 1989
2861 1599 1141 1990 ...
##  $ BMI                            : num  19.1 18.6 18.1 17.6 17.2 16.7
16.2 15.7 15.2 14.7 ...
##  $ under.five.deaths              : int  83 86 89 93 97 102 106 110
113 116 ...
##  $ Polio                          : int  6 58 62 67 68 66 63 64 63 58
...
##  $ Total.expenditure              : num  8.16 8.18 8.13 8.52 7.87 9.2
9.42 8.33 6.73 7.43 ...
##  $ Diphtheria                     : int  65 62 64 67 68 66 63 64 63 58
...
##  $ HIV.AIDS                       : num  0.1 0.1 0.1 0.1 0.1 0.1 0.1
0.1 0.1 0.1 ...
##  $ GDP                            : num  584.3 612.7 631.7 670 63.5
...
##  $ Population                     : num  33736494 327582 31731688
3696958 2978599 ...
##  $ thinness..1.19.years           : num  17.2 17.5 17.7 17.9 18.2 18.4
18.6 18.8 19 19.2 ...
```

```
##  $ thinness.5.9.years            : num  17.3 17.5 17.7 18 18.2 18.4
18.7 18.9 19.1 19.3 ...
##  $ Income.composition.of.resources: num  0.479 0.476 0.47 0.463 0.454
0.448 0.434 0.433 0.415 0.405 ...
##  $ Schooling                      : num  10.1 10 9.9 9.8 9.5 9.2 8.9
8.7 8.4 8.1 ...
```

```
head(datas,n=5)
```

```
##       Country Year     Status Life.expectancy Adult.Mortality
infant.deaths
## 1 Afghanistan 2015 Developing            65.0             263
62
## 2 Afghanistan 2014 Developing            59.9             271
64
## 3 Afghanistan 2013 Developing            59.9             268
66
## 4 Afghanistan 2012 Developing            59.5             272
69
## 5 Afghanistan 2011 Developing            59.2             275
71
##   Alcohol percentage.expenditure Hepatitis.B Measles  BMI
under.five.deaths
## 1    0.01              71.279624          65    1154 19.1
83
## 2    0.01              73.523582          62     492 18.6
86
## 3    0.01              73.219243          64     430 18.1
89
## 4    0.01              78.184215          67    2787 17.6
93
## 5    0.01               7.097109          68    3013 17.2
97
##   Polio Total.expenditure Diphtheria HIV.AIDS       GDP Population
## 1     6              8.16         65      0.1 584.25921   33736494
## 2    58              8.18         62      0.1 612.69651     327582
## 3    62              8.13         64      0.1 631.74498   31731688
## 4    67              8.52         67      0.1 669.95900    3696958
## 5    68              7.87         68      0.1  63.53723    2978599
##   thinness..1.19.years thinness.5.9.years
Income.composition.of.resources
## 1                 17.2               17.3
0.479
## 2                 17.5               17.5
0.476
## 3                 17.7               17.7
0.470
## 4                 17.9               18.0
0.463
## 5                 18.2               18.2
```

```
0.454
##   Schooling
## 1      10.1
## 2      10.0
## 3       9.9
## 4       9.8
## 5       9.5
```

```
tail(datas,5)
```

```
##       Country Year     Status Life.expectancy Adult.Mortality
infant.deaths
## 2910 Zimbabwe 2004 Developing            44.3             723
27
## 2911 Zimbabwe 2003 Developing            44.5             715
26
## 2912 Zimbabwe 2002 Developing            44.8              73
25
## 2913 Zimbabwe 2001 Developing            45.3             686
25
## 2914 Zimbabwe 2000 Developing            46.0             665
24
##      Alcohol percentage.expenditure Hepatitis.B Measles  BMI
under.five.deaths
## 2910    4.36                      0          68      31 27.1
42
## 2911    4.06                      0           7     998 26.7
41
## 2912    4.43                      0          73     304 26.3
40
## 2913    1.72                      0          76     529 25.9
39
## 2914    1.68                      0          79    1483 25.5
39
##      Polio Total.expenditure Diphtheria HIV.AIDS       GDP Population
## 2910    67              7.13         65     33.6 454.36665   12777511
## 2911     7              6.52         68     36.7 453.35116   12633897
## 2912    73              6.53         71     39.8  57.34834     125525
## 2913    76              6.16         75     42.1 548.58731   12366165
## 2914    78              7.10         78     43.5 547.35888   12222251
##      thinness..1.19.years thinness.5.9.years
Income.composition.of.resources
## 2910                  9.4                9.4
0.407
## 2911                  9.8                9.9
0.418
## 2912                  1.2                1.3
0.427
## 2913                  1.6                1.7
0.427
```

```
## 2914                 11.0               11.2
0.434
##      Schooling
## 2910       9.2
## 2911       9.5
## 2912      10.0
## 2913       9.8
## 2914       9.8
```

```
ncol(datas)
```

```
## [1] 22
```

```
nrow(datas)
```

```
## [1] 2914
```

```
names(datas)
```

```
##  [1] "Country"                          "Year"
##  [3] "Status"                           "Life.expectancy"
##  [5] "Adult.Mortality"                  "infant.deaths"
##  [7] "Alcohol"                          "percentage.expenditure"
##  [9] "Hepatitis.B"                      "Measles"
## [11] "BMI"                              "under.five.deaths"
## [13] "Polio"                            "Total.expenditure"
## [15] "Diphtheria"                       "HIV.AIDS"
## [17] "GDP"                              "Population"
## [19] "thinness..1.19.years"             "thinness.5.9.years"
## [21] "Income.composition.of.resources" "Schooling"
```

```
#datas
```

Variable transformation

```
dato<-datas


#Categorical variables
datac <-dato %>%
mutate(across(contains(c("Country","Status")),~as.factor(.x)))

#Numerics variables
datan <-dato %>%
mutate(across(contains(c("Year","Life.expectancy","Adult.Mortality","infa
nt.deaths","Alcohol","percentage.expenditure","Diphtheria","HIV.AIDS","Me
asles","GDP","Population","thinness..1.19.years","thinness.5.9.years","In
come.composition.of.resources","Schooling","Total.expenditure","Polio","B
MI","Hepatitis.B")),~as.numeric(.x)))

data <-mutate(datac,datan)
#data
```

Data cleaning and transformation

```
#It is placed as a comment so as not to extend the deployment too much of codes
#View(data)

anyNA(data)

## [1] TRUE

sum(is.na(data))

## [1] 2468

#Transform to numeric

Life.expectancy <-as.numeric(unlist(data[,-1]))
Adult.Mortality <-as.numeric(unlist(data[,-2]))
infant.deaths <-as.numeric(unlist(data[,-3]))
Alcohol <-as.numeric(unlist(data[,-4]))
percentage.expenditure <-as.numeric(unlist(data[,-5]))
Hepatitis.B <-as.numeric(unlist(data[,-6]))
Measles <-as.numeric(unlist(data[,-7]))
BMI <-as.numeric(unlist(data[,-8]))
under.five.deaths <-as.numeric(unlist(data[,-9]))
Polio <-as.numeric(unlist(data[,-10]))
Total.expenditure <-as.numeric(unlist(data[,-11]))
Diphtheria <-as.numeric(unlist(data[,-12]))
HIV.AIDS <-as.numeric(unlist(data[,-13]))
GDP <-as.numeric(unlist(data[,-14]))
Population <-as.numeric(unlist(data[,-15]))
thinness..1.19.years <-as.numeric(unlist(data[,-16]))
thinness.5.9.years <-as.numeric(unlist(data[,-17]))
Income.composition.of.resources <-as.numeric(unlist(data[,-18]))
Schooling <-as.numeric(unlist(data[,-19]))

# Class type identification
class(data$Life.expectancy)

## [1] "numeric"

class(data$Hepatitis.B)

## [1] "numeric"

#It is placed as a comment so as not to extend the deployment too much of codes
#View(data)
```

Mean value imputation

```
data$Life.expectancy[is.na(data$Life.expectancy)] <-
mean(data$Life.expectancy, na.rm=T) ##cambiar por la media de
```

```
expectactiva de vida
data$Adult.Mortality[is.na(data$Adult.Mortality)] <-
mean(data$Adult.Mortality, na.rm=T) ##cambiar por la media de
Adult.Mortality
data$infant.deaths[is.na(data$infant.deaths)] <-mean(data$infant.deaths,
na.rm=T) ##cambiar por la media de infant.deaths
data$Alcohol[is.na(data$Alcohol)] <-mean(data$Alcohol, na.rm=T) ##cambiar
por la media de Alcohol
data$percentage.expenditure[is.na(data$percentage.expenditure)] <-
mean(data$percentage.expenditure, na.rm=T) ##cambiar por la media de
percentage.expenditure
data$Hepatitis.B[is.na(data$Hepatitis.B)] <-mean(data$Hepatitis.B,
na.rm=T) ##cambiar por la media de Hepatitis.B
data$Measles[is.na(data$Measles)] <-mean(data$Measles, na.rm=T) ##cambiar
por la media de Measles
data$BMI[is.na(data$BMI)] <-mean(data$BMI , na.rm=T) ##cambiar por la
media de BMI
data$under.five.deaths[is.na(data$under.five.deaths)] <-
mean(data$under.five.deaths, na.rm=T) ##cambiar por la media de
under.five.deaths
data$Polio[is.na(data$Polio)] <-mean(data$Polio, na.rm=T) ##cambiar por
la media de Polio
data$Total.expenditure[is.na(data$Total.expenditure)] <-
mean(data$Total.expenditure, na.rm=T) ##cambiar por la media de
Total.expenditure
data$Diphtheria[is.na(data$Diphtheria)] <-mean(data$Diphtheria, na.rm=T)
##cambiar por la media de Diphtheria
data$HIV.AIDS[is.na(data$HIV.AIDS)] <-mean(data$HIV.AIDS, na.rm=T)
##cambiar por la media de HIV.AIDS
data$GDP[is.na(data$GDP)] <-mean(data$GDP, na.rm=T) ##cambiar por la
media de GDP
data$Population[is.na(data$Population)] <-mean(data$Population, na.rm=T)
##cambiar por la media de Population
data$thinness..1.19.years[is.na(data$thinness..1.19.years)] <-
mean(data$thinness..1.19.years, na.rm=T) ##cambiar por la media de
thinness..1.19.years
data$thinness.5.9.years[is.na(data$thinness.5.9.years)] <-
mean(data$thinness.5.9.years, na.rm=T) ##cambiar por thinness.5.9.years
data$Income.composition.of.resources[is.na(data$Income.composition.of.res
ources)] <-mean(data$Income.composition.of.resources, na.rm=T) ##cambiar
por Income.composition.of.resources
data$Schooling[is.na(data$Schooling)] <-mean(data$Schooling, na.rm=T)
##cambiar por Schooling
```

Review of null and na.

###Also, it is possible to obtain values missing with data <- na.omit(data) or drop_na()

```
sum(is.na(data$Life.expectancy))
```

```
## [1] 0
sum(is.null(data$Life.expectancy))#no hay
## [1] 0
sum(is.na(data))
## [1] 0
```

DESCRIPTION OF VARIABLES

##assessing the concentration of data, identifying anomalies, outliers, relationships and others ## Univariate analysis

```
datas %>% glimpse
## Rows: 2,914
## Columns: 22
## $ Country                         <chr> "Afghanistan", "Afghanistan",
"Afghani…
## $ Year                            <int> 2015, 2014, 2013, 2012, 2011,
2010, 20…
## $ Status                          <chr> "Developing", "Developing",
"Developin…
## $ Life.expectancy                 <dbl> 65.0, 59.9, 59.9, 59.5, 59.2,
58.8, 58…
## $ Adult.Mortality                 <int> 263, 271, 268, 272, 275, 279,
281, 287…
## $ infant.deaths                   <int> 62, 64, 66, 69, 71, 74, 77,
80, 82, 84…
## $ Alcohol                         <dbl> 0.01, 0.01, 0.01, 0.01, 0.01,
0.01, 0.…
## $ percentage.expenditure          <dbl> 71.279624, 73.523582,
73.219243, 78.18…
## $ Hepatitis.B                     <int> 65, 62, 64, 67, 68, 66, 63,
64, 63, 64…
## $ Measles                         <int> 1154, 492, 430, 2787, 3013,
1989, 2861…
## $ BMI                             <dbl> 19.1, 18.6, 18.1, 17.6, 17.2,
16.7, 16…
## $ under.five.deaths               <int> 83, 86, 89, 93, 97, 102, 106,
110, 113…
## $ Polio                           <int> 6, 58, 62, 67, 68, 66, 63, 64,
63, 58,…
## $ Total.expenditure               <dbl> 8.16, 8.18, 8.13, 8.52, 7.87,
9.20, 9.…
## $ Diphtheria                      <int> 65, 62, 64, 67, 68, 66, 63,
64, 63, 58…
## $ HIV.AIDS                        <dbl> 0.1, 0.1, 0.1, 0.1, 0.1, 0.1,
```

```
0.1, 0.1…
## $ GDP                             <dbl> 584.25921, 612.69651,
631.74498, 669.9…
## $ Population                      <dbl> 33736494, 327582, 31731688,
3696958, 2…
## $ thinness..1.19.years            <dbl> 17.2, 17.5, 17.7, 17.9, 18.2,
18.4, 18…
## $ thinness.5.9.years              <dbl> 17.3, 17.5, 17.7, 18.0, 18.2,
18.4, 18…
## $ Income.composition.of.resources <dbl> 0.479, 0.476, 0.470, 0.463,
0.454, 0.4…
## $ Schooling                       <dbl> 10.1, 10.0, 9.9, 9.8, 9.5,
9.2, 8.9, 8…

#Status-categorical variable

Stated <-unique(data$Country)
Stated

##   [1] "Afghanistan"
##   [2] "Albania"
##   [3] "Algeria"
##   [4] "Angola"
##   [5] "Antigua and Barbuda"
##   [6] "Argentina"
##   [7] "Armenia"
##   [8] "Australia"
##   [9] "Austria"
##  [10] "Azerbaijan"
##  [11] "Bahamas"
##  [12] "Bahrain"
##  [13] "Bangladesh"
##  [14] "Barbados"
##  [15] "Belarus"
##  [16] "Belgium"
##  [17] "Belize"
##  [18] "Benin"
##  [19] "Bhutan"
##  [20] "Bolivia (Plurinational State of)"
##  [21] "Bosnia and Herzegovina"
##  [22] "Botswana"
##  [23] "Brazil"
##  [24] "Brunei Darussalam"
##  [25] "Bulgaria"
##  [26] "Burkina Faso"
##  [27] "Burundi"
##  [28] "Côte
dIvoire,2015,Developing,53.3,397,57,,0,83,65,28,79,81,,83,1.9,,,5.5,5.5,,
\nCôte dIvoire"
##  [29] "Côte
```

```
dIvoire,2013,Developing,52.3,412,59,3.15,0,8,48,26.8,81,79,5.81,8,2.4,,,5
.8,5.7,,\nCôte dIvoire"
##  [30] "Côte
dIvoire,2011,Developing,51.7,419,60,3.13,0,62,628,25.6,83,58,6.42,62,3.3,
,,6.1,6,,\nCôte dIvoire"
##  [31] "Côte
dIvoire,2009,Developing,51,426,60,2.92,0,81,183,24.4,84,77,6.41,81,3.7,,,
6.5,6.4,,\nCôte dIvoire"
##  [32] "Côte
dIvoire,2007,Developing,49.9,443,61,2.58,0,76,5,23.2,87,75,6.35,76,5.3,,,
6.8,6.7,,\nCôte dIvoire"
##  [33] "Côte
dIvoire,2005,Developing,48.7,466,63,3.11,0,76,115,22.1,90,87,5.39,76,6.1,
,,7.2,7.1,,\nCôte dIvoire"
##  [34] "Côte
dIvoire,2003,Developing,48,473,64,3.12,0,63,4770,2.9,92,68,4.65,61,6.7,,,
7.5,7.5,,\nCôte dIvoire"
##  [35] "Côte
dIvoire,2001,Developing,47.8,467,65,3.15,0,1,5790,19.9,94,7,4.85,66,7,,,7
.9,7.9,,\nCôte dIvoire"
##  [36] "Cabo Verde"
##  [37] "Cambodia"
##  [38] "Cameroon"
##  [39] "Canada"
##  [40] "Central African Republic"
##  [41] "Chad"
##  [42] "Chile"
##  [43] "China"
##  [44] "Colombia"
##  [45] "Comoros"
##  [46] "Congo"
##  [47] "Cook Islands"
##  [48] "Costa Rica"
##  [49] "Croatia"
##  [50] "Cuba"
##  [51] "Cyprus"
##  [52] "Czechia"
##  [53] "Democratic Peoples Republic of
Korea,2015,Developing,76,139,6,,0,96,0,32.9,7,99,,96,0.1,,,4.9,4.9,,\nDem
ocratic Peoples Republic of Korea"
##  [54] "Democratic Peoples Republic of
Korea,2013,Developing,71,146,6,3.35,0,93,0,31.8,8,99,,93,0.1,,,5,5,,\nDem
ocratic Peoples Republic of Korea"
##  [55] "Democratic Peoples Republic of
Korea,2011,Developing,69.4,153,8,3.39,0,94,0,3.8,10,99,,94,0.1,,,5.1,5.2,
,\nDemocratic Peoples Republic of Korea"
##  [56] "Democratic Peoples Republic of
Korea,2009,Developing,68.7,161,9,3.35,0,93,0,29.7,11,98,,93,0.1,,,5.3,5.3
,,\nDemocratic Peoples Republic of Korea"
##  [57] "Democratic Peoples Republic of
```

```
Korea,2007,Developing,68.5,166,9,3.13,0,92,3550,28.7,12,99,,92,0.1,,,5.5,
5.5,,\nDemocratic Peoples Republic of Korea"
##  [58] "Democratic Peoples Republic of
Korea,2005,Developing,68.5,165,10,3.21,0,92,0,27.7,13,97,,79,0.1,,,5.7,5.
7,,\nDemocratic Peoples Republic of Korea"
##  [59] "Democratic Peoples Republic of
Korea,2003,Developing,68.1,165,12,3.13,0,27,0,26.7,15,99,,68,0.1,,,5.8,5.
8,,\nDemocratic Peoples Republic of Korea"
##  [60] "Democratic Peoples Republic of
Korea,2001,Developing,66.6,177,16,2.53,0,,0,25.7,21,98,,62,0.1,,,5.9,6,,\
nDemocratic Peoples Republic of Korea"
##  [61] "Democratic Republic of the Congo"
##  [62] "Denmark"
##  [63] "Djibouti"
##  [64] "Dominica"
##  [65] "Dominican Republic"
##  [66] "Ecuador"
##  [67] "Egypt"
##  [68] "El Salvador"
##  [69] "Equatorial Guinea"
##  [70] "Eritrea"
##  [71] "Estonia"
##  [72] "Ethiopia"
##  [73] "Fiji"
##  [74] "Finland"
##  [75] "France"
##  [76] "Gabon"
##  [77] "Gambia"
##  [78] "Georgia"
##  [79] "Germany"
##  [80] "Ghana"
##  [81] "Greece"
##  [82] "Grenada"
##  [83] "Guatemala"
##  [84] "Guinea"
##  [85] "Guinea-Bissau"
##  [86] "Guyana"
##  [87] "Haiti"
##  [88] "Honduras"
##  [89] "Hungary"
##  [90] "Iceland"
##  [91] "India"
##  [92] "Indonesia"
##  [93] "Iran (Islamic Republic of)"
##  [94] "Iraq"
##  [95] "Ireland"
##  [96] "Israel"
##  [97] "Italy"
##  [98] "Jamaica"
##  [99] "Japan"
```

```
## [100] "Jordan"
## [101] "Kazakhstan"
## [102] "Kenya"
## [103] "Kiribati"
## [104] "Kuwait"
## [105] "Kyrgyzstan"
## [106] "Lao Peoples Democratic
Republic,2015,Developing,65.7,194,8,,0,89,56,21.7,11,89,,89,0.2,,,8.8,8.9
,0.582,10.8\nLao Peoples Democratic Republic"
## [107] "Lao Peoples Democratic
Republic,2013,Developing,64.9,23,9,0.01,0,87,71,2.1,12,86,2,87,0.3,,,9,9.
1,0.563,10.4\nLao Peoples Democratic Republic"
## [108] "Lao Peoples Democratic
Republic,2011,Developing,64,213,9,5.39,0,78,113,18.7,13,79,2.2,78,0.3,,,9
.2,9.4,0.542,9.9\nLao Peoples Democratic Republic"
## [109] "Lao Peoples Democratic
Republic,2009,Developing,63.1,223,10,5.18,0,67,78,17.3,14,67,3.77,67,0.2,
,,9.4,9.6,0.525,9.4\nLao Peoples Democratic Republic"
## [110] "Lao Peoples Democratic
Republic,2007,Developing,62.1,234,11,5,0,5,1678,16.1,15,46,4.14,5,0.2,,,9
.7,9.8,0.509,9\nLao Peoples Democratic Republic"
## [111] "Lao Peoples Democratic
Republic,2005,Developing,61,246,11,3.68,0,49,295,14.9,16,5,4.32,49,0.2,,,
1,1.1,0.494,8.7\nLao Peoples Democratic Republic"
## [112] "Lao Peoples Democratic
Republic,2003,Developing,59.8,259,12,3.41,0,5,1810,13.8,17,52,4.91,49,0.1
,,,1.2,1.3,0.477,8.3\nLao Peoples Democratic Republic"
## [113] "Lao Peoples Democratic
Republic,2001,Developing,58.7,271,13,3.13,0,,94,12.7,19,55,4.32,52,0.1,,,
1.4,1.5,0.463,8\nLao Peoples Democratic Republic"
## [114] "Latvia"
## [115] "Lebanon"
## [116] "Lesotho"
## [117] "Liberia"
## [118] "Libya"
## [119] "Lithuania"
## [120] "Luxembourg"
## [121] "Madagascar"
## [122] "Malawi"
## [123] "Malaysia"
## [124] "Maldives"
## [125] "Mali"
## [126] "Malta"
## [127] "Marshall Islands"
## [128] "Mauritania"
## [129] "Mauritius"
## [130] "Mexico"
## [131] "Micronesia (Federated States of)"
## [132] "Monaco"
## [133] "Mongolia"
```

```
## [134] "Montenegro"
## [135] "Morocco"
## [136] "Mozambique"
## [137] "Myanmar"
## [138] "Namibia"
## [139] "Nauru"
## [140] "Nepal"
## [141] "Netherlands"
## [142] "New Zealand"
## [143] "Nicaragua"
## [144] "Niger"
## [145] "Nigeria"
## [146] "Niue"
## [147] "Norway"
## [148] "Oman"
## [149] "Pakistan"
## [150] "Palau"
## [151] "Panama"
## [152] "Papua New Guinea"
## [153] "Paraguay"
## [154] "Peru"
## [155] "Philippines"
## [156] "Poland"
## [157] "Portugal"
## [158] "Qatar"
## [159] "Republic of Korea"
## [160] "Republic of Moldova"
## [161] "Romania"
## [162] "Russian Federation"
## [163] "Rwanda"
## [164] "Saint Kitts and Nevis"
## [165] "Saint Lucia"
## [166] "Saint Vincent and the Grenadines"
## [167] "Samoa"
## [168] "San Marino"
## [169] "Sao Tome and Principe"
## [170] "Saudi Arabia"
## [171] "Senegal"
## [172] "Serbia"
## [173] "Seychelles"
## [174] "Sierra Leone"
## [175] "Singapore"
## [176] "Slovakia"
## [177] "Slovenia"
## [178] "Solomon Islands"
## [179] "Somalia"
## [180] "South Africa"
## [181] "South Sudan"
## [182] "Spain"
## [183] "Sri Lanka"
```

```
## [184] "Sudan"
## [185] "Suriname"
## [186] "Swaziland"
## [187] "Sweden"
## [188] "Switzerland"
## [189] "Syrian Arab Republic"
## [190] "Tajikistan"
## [191] "Thailand"
## [192] "The former Yugoslav republic of Macedonia"
## [193] "Timor-Leste"
## [194] "Togo"
## [195] "Tonga"
## [196] "Trinidad and Tobago"
## [197] "Tunisia"
## [198] "Turkey"
## [199] "Turkmenistan"
## [200] "Tuvalu"
## [201] "Uganda"
## [202] "Ukraine"
## [203] "United Arab Emirates"
## [204] "United Kingdom of Great Britain and Northern Ireland"
## [205] "United Republic of Tanzania"
## [206] "United States of America"
## [207] "Uruguay"
## [208] "Uzbekistan"
## [209] "Vanuatu"
## [210] "Venezuela (Bolivarian Republic of)"
## [211] "Viet Nam"
## [212] "Yemen"
## [213] "Zambia"
## [214] "Zimbabwe"
```

```
#data$Status[unique(data$Country)]

Statuss <-factor(sample(c("Developed", "Developing"),
                        size=length(data$Status),replace = TRUE))
#It is placed as a comment so as not to extend the deployment too much of
codes
#Statuss
Tipodepais<-data$Status
table(Tipodepais)
```

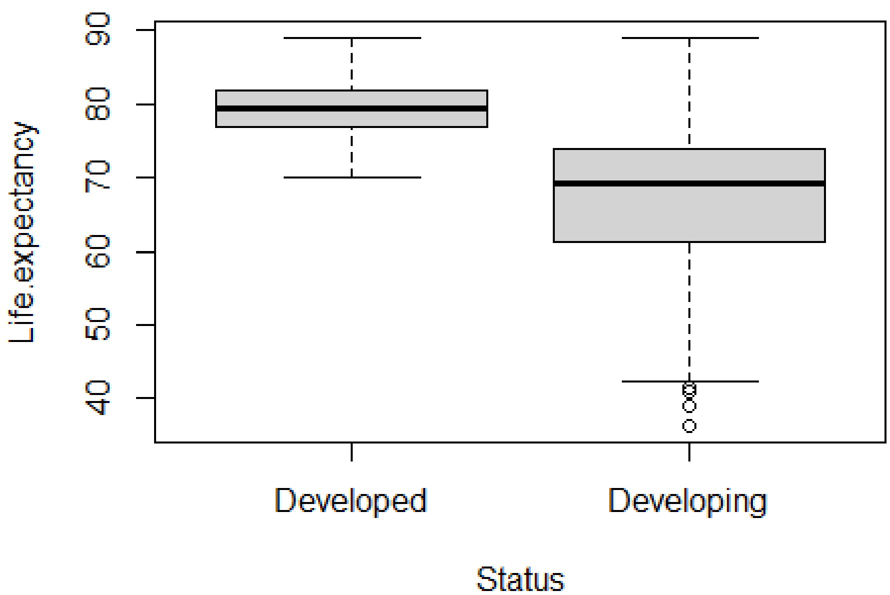
```
## Tipodepais
##  Developed Developing
##        512       2402
```

```
sum(Tipodepais=="Developed")
```

```
## [1] 512
```

```r
Statust<-prop.table(table(Tipodepais))
Statust

## Tipodepais
##  Developed Developing
##  0.1757035  0.8242965

ggplot(data)+
  aes(x=Status,fill="Stated")+
  geom_bar()+
  ggtitle("Status")
```


```r
boxplot(Life.expectancy ~ Status, data = data)
```




```
#Country

length(data$Country[data$Country == "Costa Rica"])#number of times CR
appears

## [1] 16

CostaRicacountryy<-  data %>%
  filter(Country=="Costa Rica",
         Year==2014)#information about a specific variable
CostaRicacountryy

##      Country Year     Status Life.expectancy Adult.Mortality
infant.deaths
## 1 Costa Rica 2014 Developing            79.5              96
1
##   Alcohol percentage.expenditure Hepatitis.B Measles  BMI
under.five.deaths
## 1    3.45               384.5129          91       1 59.5
1
##   Polio Total.expenditure Diphtheria HIV.AIDS      GDP Population
## 1    91              9.31         91      0.1 1647.442    4757575
##   thinness..1.19.years thinness.5.9.years
Income.composition.of.resources
## 1                  1.7                1.7
0.768
```

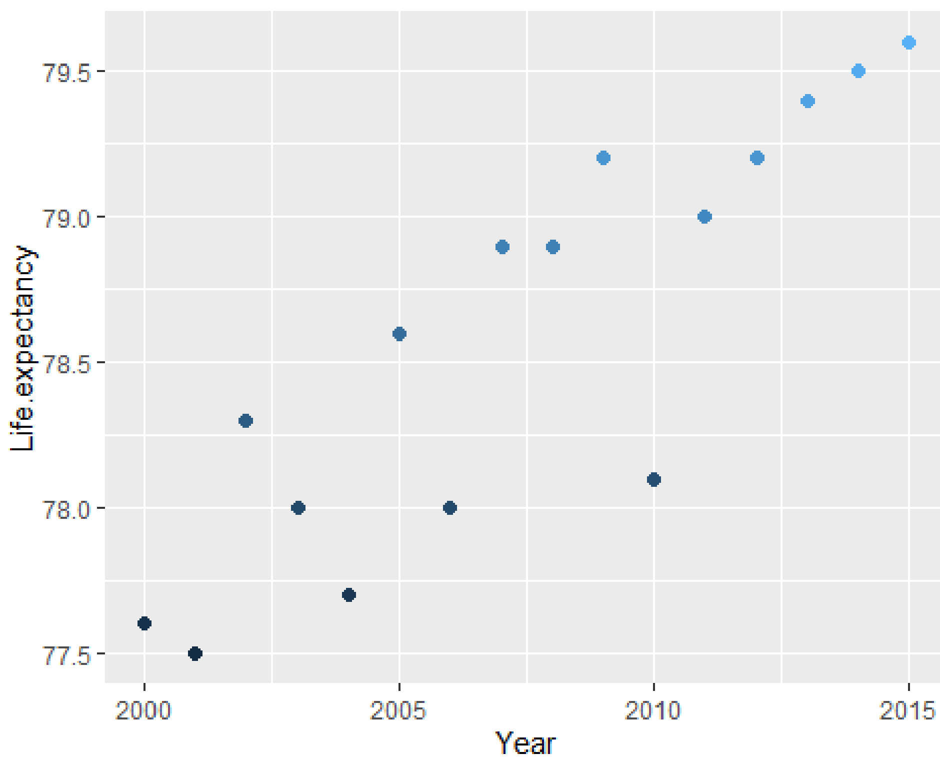
```
##   Schooling
## 1      13.9
```

```r
#View(CostaRicacountryy)
summary(CostaRicacountryy)
```

```
##    Country               Year         Status          Life.expectancy
##  Length:1           Min.   :2014   Length:1           Min.   :79.5   
##  Class :character   1st Qu.:2014   Class :character   1st Qu.:79.5   
##  Mode  :character   Median :2014   Mode  :character   Median :79.5   
##                     Mean   :2014                      Mean   :79.5   
##                     3rd Qu.:2014                      3rd Qu.:79.5   
##                     Max.   :2014                      Max.   :79.5   
##  Adult.Mortality infant.deaths    Alcohol     percentage.expenditure
##  Min.   :96      Min.   :1     Min.   :3.45   Min.   :384.5         
##  1st Qu.:96      1st Qu.:1     1st Qu.:3.45   1st Qu.:384.5         
##  Median :96      Median :1     Median :3.45   Median :384.5         
##  Mean   :96      Mean   :1     Mean   :3.45   Mean   :384.5         
##  3rd Qu.:96      3rd Qu.:1     3rd Qu.:3.45   3rd Qu.:384.5         
##  Max.   :96      Max.   :1     Max.   :3.45   Max.   :384.5         
##   Hepatitis.B    Measles       BMI        under.five.deaths     Polio   
##  Min.   :91   Min.   :1   Min.   :59.5   Min.   :1          Min.   :91  
##  1st Qu.:91   1st Qu.:1   1st Qu.:59.5   1st Qu.:1          1st Qu.:91  
##  Median :91   Median :1   Median :59.5   Median :1          Median :91  
##  Mean   :91   Mean   :1   Mean   :59.5   Mean   :1          Mean   :91  
##  3rd Qu.:91   3rd Qu.:1   3rd Qu.:59.5   3rd Qu.:1          3rd Qu.:91  
##  Max.   :91   Max.   :1   Max.   :59.5   Max.   :1          Max.   :91  
##  Total.expenditure   Diphtheria    HIV.AIDS        GDP        
Population     
##  Min.   :9.31      Min.   :91   Min.   :0.1   Min.   :1647   Min.   
:4757575  
##  1st Qu.:9.31      1st Qu.:91   1st Qu.:0.1   1st Qu.:1647   1st 
Qu.:4757575  
##  Median :9.31      Median :91   Median :0.1   Median :1647   Median 
:4757575  
##  Mean   :9.31      Mean   :91   Mean   :0.1   Mean   :1647   Mean   
:4757575  
##  3rd Qu.:9.31      3rd Qu.:91   3rd Qu.:0.1   3rd Qu.:1647   3rd 
Qu.:4757575  
##  Max.   :9.31      Max.   :91   Max.   :0.1   Max.   :1647   Max.   
:4757575  
##  thinness..1.19.years thinness.5.9.years
Income.composition.of.resources
##  Min.   :1.7          Min.   :1.7         Min.   :0.768                  
##  1st Qu.:1.7          1st Qu.:1.7         1st Qu.:0.768                  
##  Median :1.7          Median :1.7         Median :0.768                  
##  Mean   :1.7          Mean   :1.7         Mean   :0.768                  
##  3rd Qu.:1.7          3rd Qu.:1.7         3rd Qu.:0.768                  
##  Max.   :1.7          Max.   :1.7         Max.   :0.768                  
##    Schooling   
```

```
##  Min.   :13.9
##  1st Qu.:13.9
##  Median :13.9
##  Mean   :13.9
##  3rd Qu.:13.9
##  Max.   :13.9
```

```
data  %>%
  filter(Country == "Costa Rica") %>%
  ggplot(aes(Year, Life.expectancy,col = Life.expectancy)) +
  geom_point(size = 2)+
  theme(legend.position = "none")
```

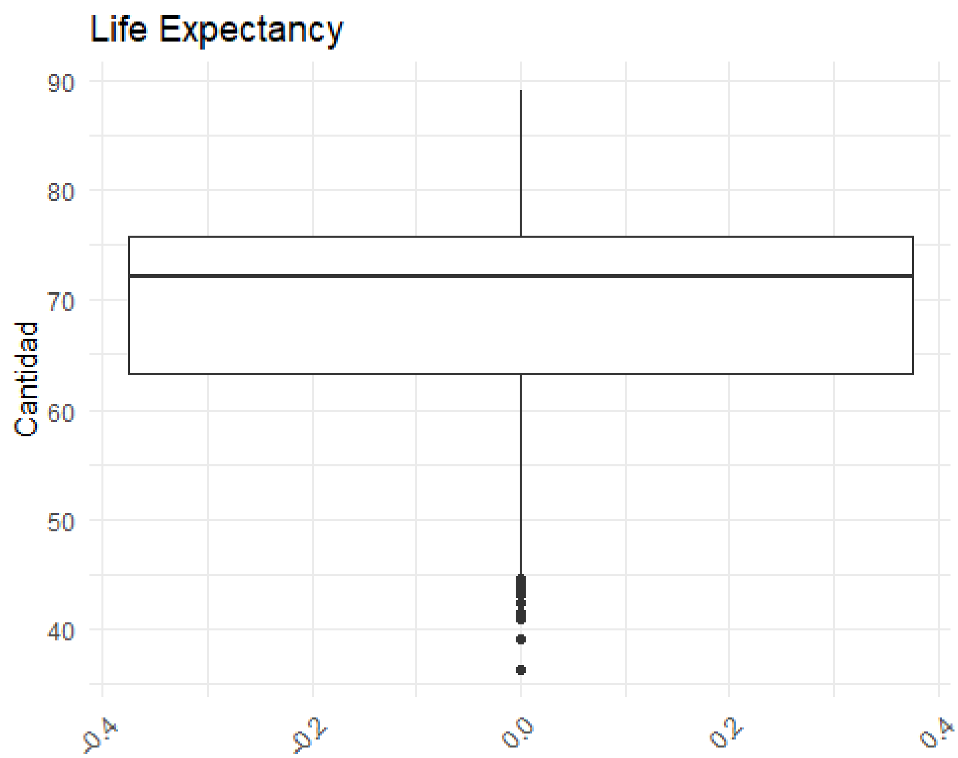



```
#Comparison CR vs 2 countries
data %>%
  filter(Year == 2014 & Country %in% c("Costa Rica", "Canada", "Chile"))
%>%
  select(Life.expectancy, Country, Year)
```

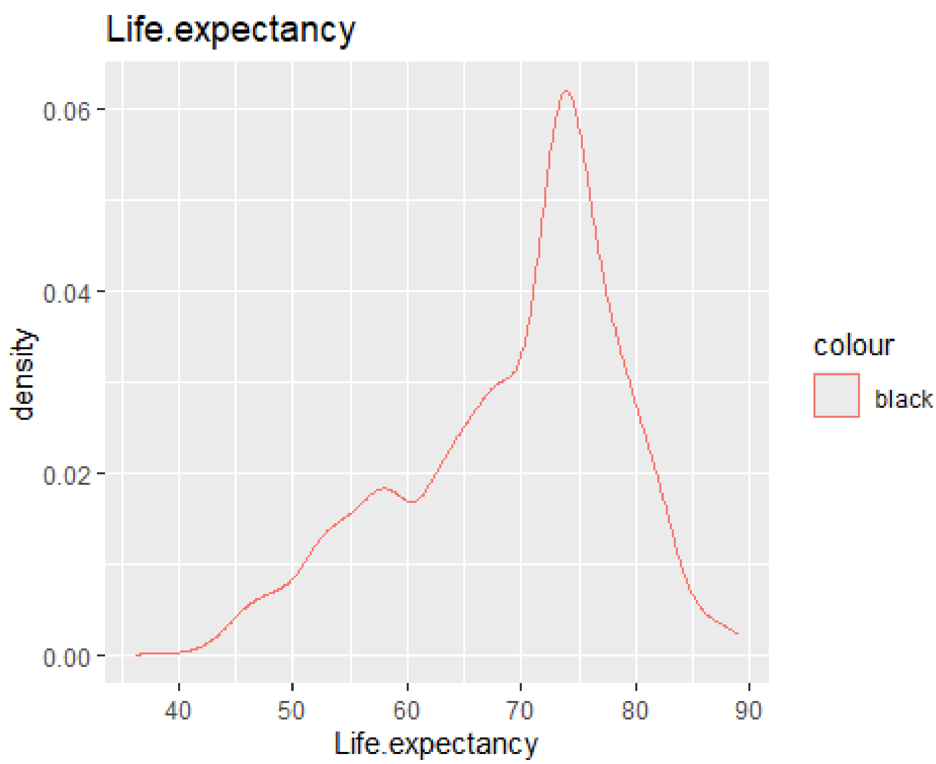
```
##   Life.expectancy    Country Year
## 1            82.0     Canada 2014
## 2            83.0      Chile 2014
## 3            79.5 Costa Rica 2014
```

```
#Life Expectancy

summary(data$Life.expectancy)
```

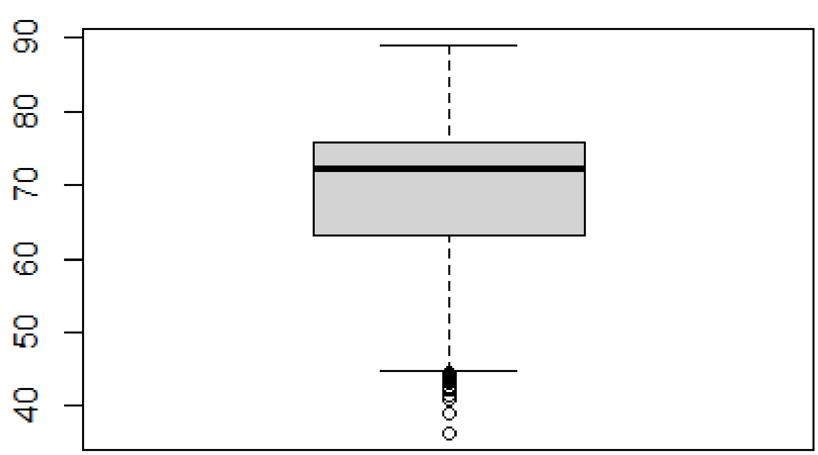
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   36.30   63.30   72.10   69.29   75.70   89.00

#Life_Adult <-data %>% select(Adult.Mortality, Country) %>%
filter(Year==2015 & Country=="Costa Rica")
#Life_Adult

#Life_Adult_world <-data %>% select(Adult.Mortality, Country) %>%
filter(Year==2015)
#Life_Adult_world


data %>%
  group_by(data$Country) %>%
  summarize(min_size = min(Life.expectancy, na.rm = TRUE))

## # A tibble: 214 × 2
##    `data$Country`      min_size
##    <chr>                  <dbl>
##  1 Afghanistan             54.8
##  2 Albania                 72.6
##  3 Algeria                 71.3
##  4 Angola                  45.3
##  5 Antigua and Barbuda     73.6
##  6 Argentina               74
##  7 Armenia                 72
##  8 Australia               79.5
##  9 Austria                 78.1
## 10 Azerbaijan              66.6
## # ℹ 204 more rows

#Life.expectancy

ggplot(data)+
  aes(x=, y=Life.expectancy)+
  geom_boxplot()+
  labs(title = "Life Expectancy", y = "Cantidad") +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



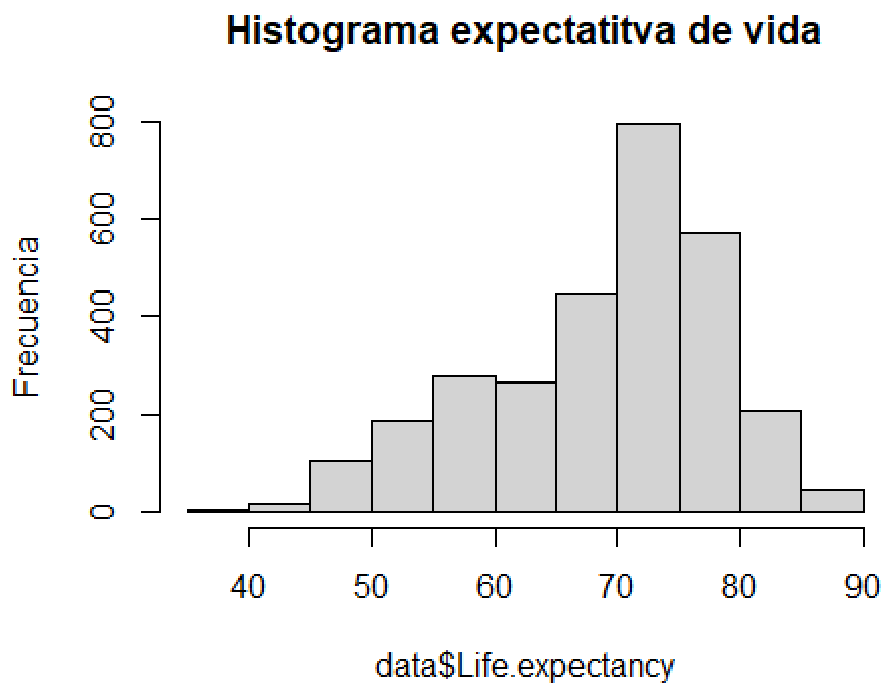

```
#Distribution by age
data %>% ggplot((aes(x=Life.expectancy)))+
  geom_density(aes(x=Life.expectancy,binwidth = 1, color = "black"))+
  ggtitle("Life.expectancy")
```




```r
boxplot(data$Life.expectancy)
```

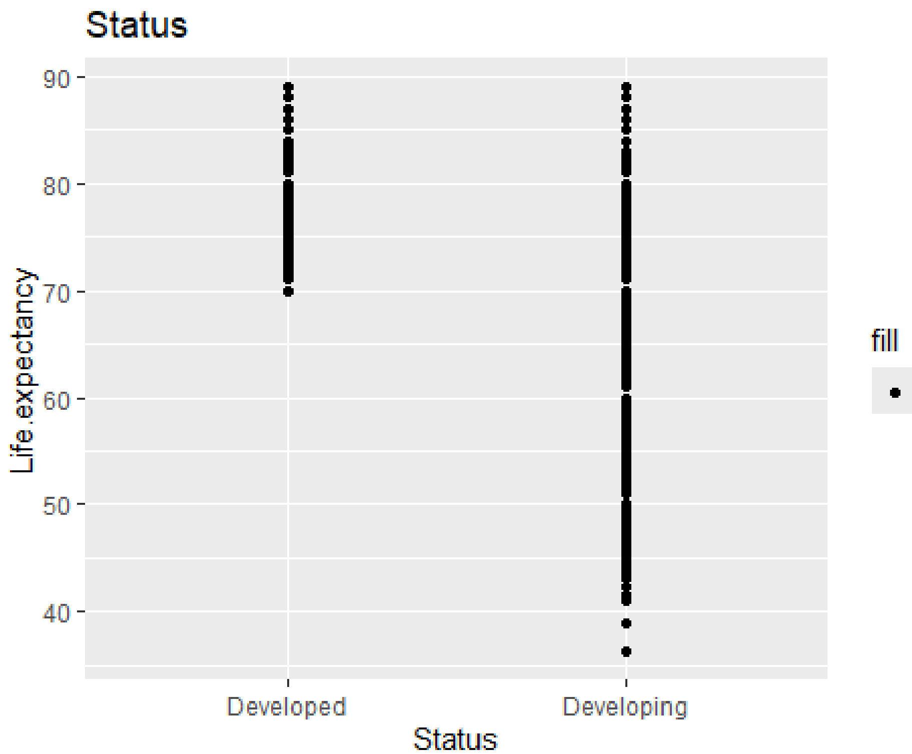



```r
hist(data$Life.expectancy, main="Histograma expectatitva de vida",ylab="Frecuencia")
```

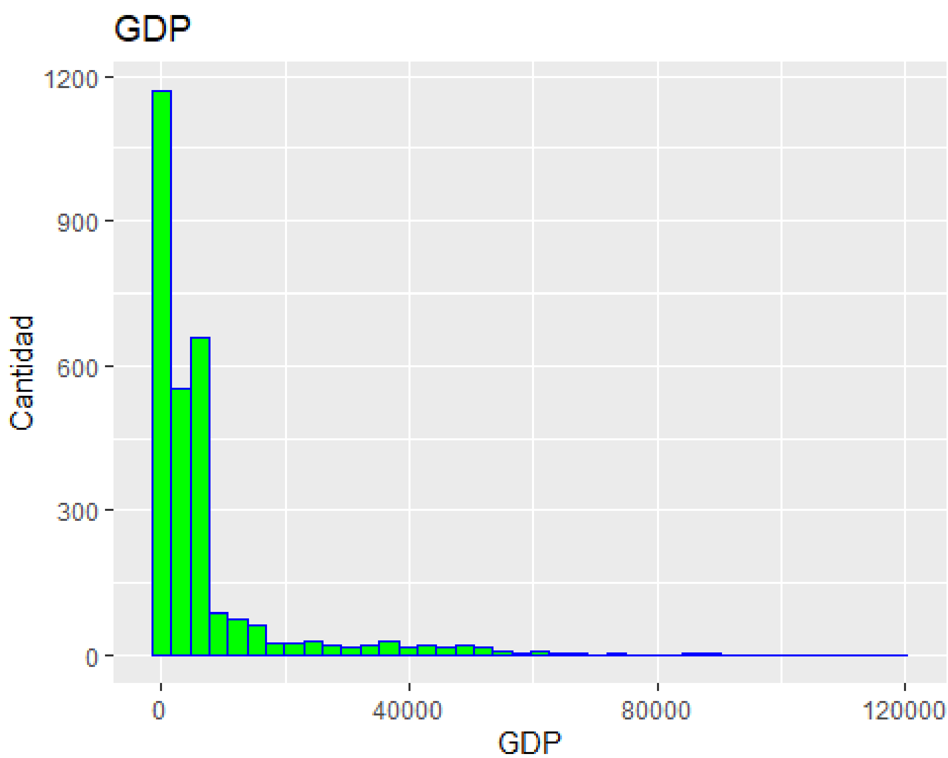



```r
#Years of life expectancy

mean(data$Life.expectancy)

## [1] 69.2947

min(data$Life.expectancy)

## [1] 36.3

max(data$Life.expectancy)

## [1] 89

#It is placed as a comment so as not to extend the deployment too much of
codes

#data %>% select(Life.expectancy, Country) %>% filter(Life.expectancy
<=50)
match(50,data$Life.expectancy)

## [1] 1479

#data %>% select(Life.expectancy, Country) %>% filter(Life.expectancy
>=80)
#data %>% select(Life.expectancy, Country, Year) %>%
filter(Country=="Costa Rica")
```

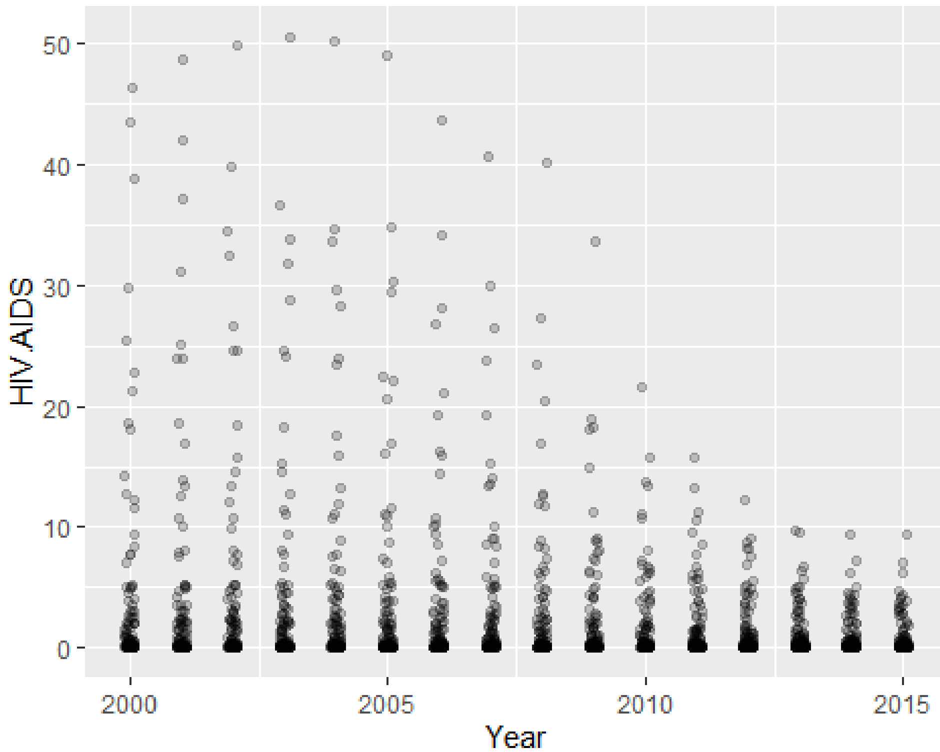
```r
ggplot(data)+
  aes(x=Status,y=Life.expectancy,fill="") +
  geom_point()+
  ggtitle("Status")
```



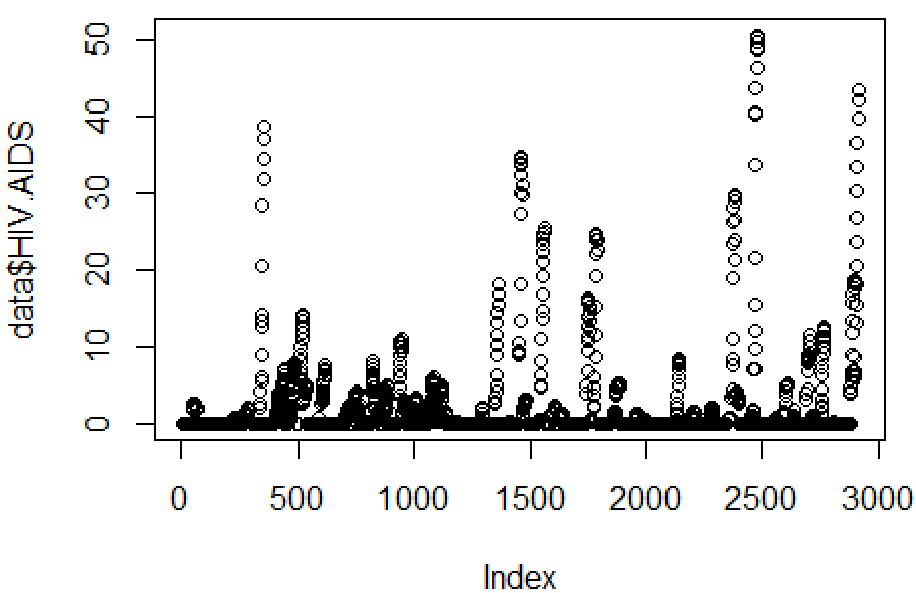

```r
#GDP

summary(data$GDP)

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##      1.68    572.94   2970.19   7483.16   7483.16 119172.74

sd((data$GDP))

## [1] 13190.81

data %>% ggplot((aes(x=GDP)))+
  geom_histogram(color="Blue",fill="green",bins=40)+
  labs(title = "GDP", y = "Cantidad")
```




```
#HIV.AIDS

data  %>% ggplot(aes(Year,HIV.AIDS)) + geom_jitter(width = 0.1, alpha =
0.2)
```




```r
plot(data$HIV.AIDS)
```




```r
summary(data$HIV.AIDS)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.100   0.100   0.100   1.743   0.800  50.600
```

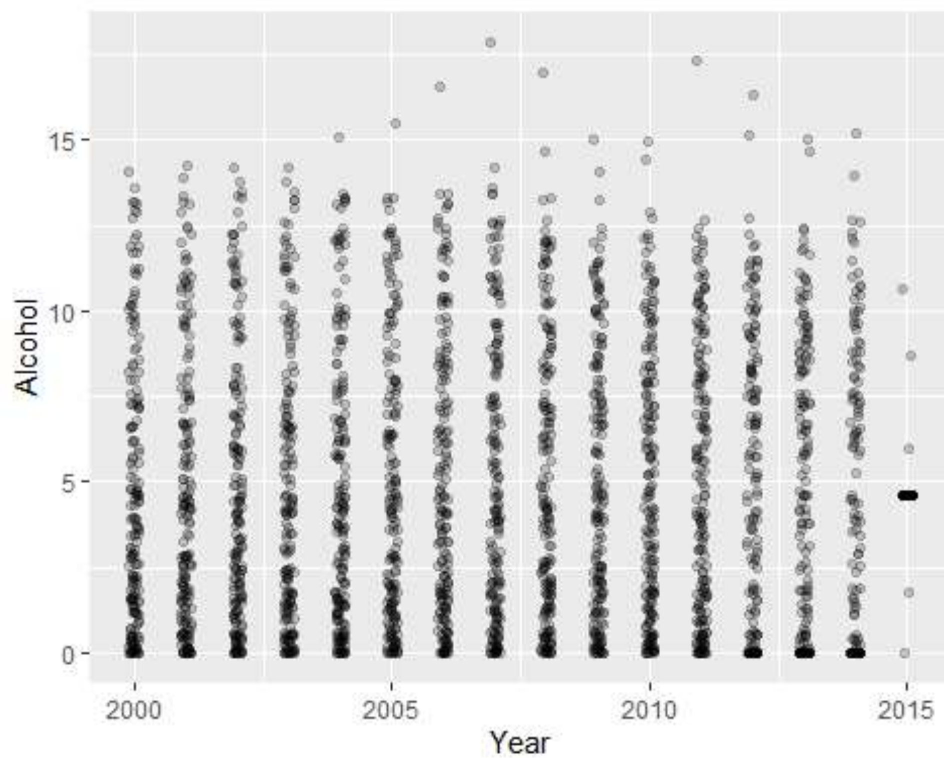
```
sd((data$HIV.AIDS))
```

```
## [1] 5.09424
```

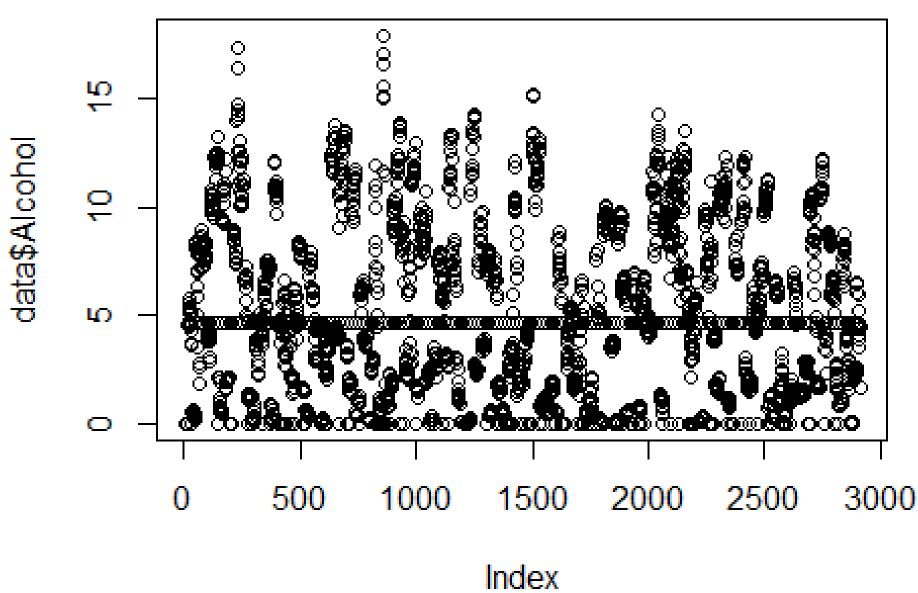
```
data %>%
  group_by(data$Year) %>%
  summarize(max = max(HIV.AIDS, na.rm = TRUE))
```

```
## # A tibble: 16 × 2
##    `data$Year`   max
##          <dbl> <dbl>
##  1        2000  46.4
##  2        2001  48.8
##  3        2002  49.9
##  4        2003  50.6
##  5        2004  50.3
##  6        2005  49.1
##  7        2006  43.7
##  8        2007  40.7
##  9        2008  40.2
## 10        2009  33.7
## 11        2010  21.6
## 12        2011  15.7
## 13        2012  12.2
## 14        2013   9.8
## 15        2014   9.4
## 16        2015   9.3
```

```
#Alcohol
data  %>% ggplot(aes(Year,Alcohol)) + geom_jitter(width = 0.1, alpha =
0.2)
```




```r
plot(data$Alcohol)
```




```r
summary(data$Alcohol)
```

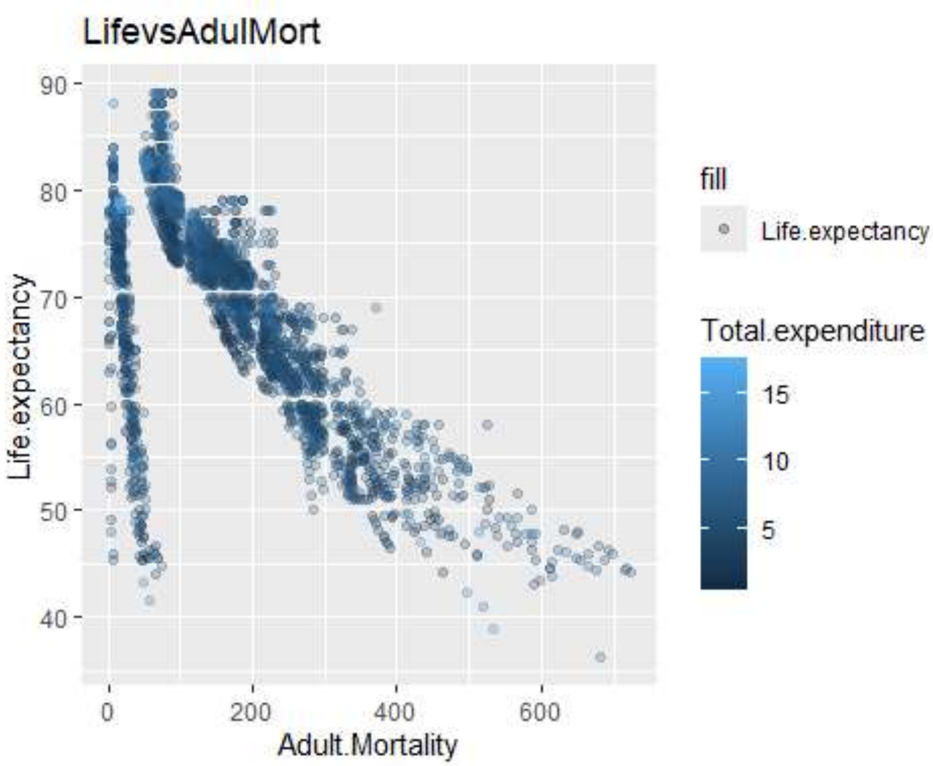
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.010   1.073   4.170   4.613   7.438  17.870
```

```r
sd((data$Alcohol))
```

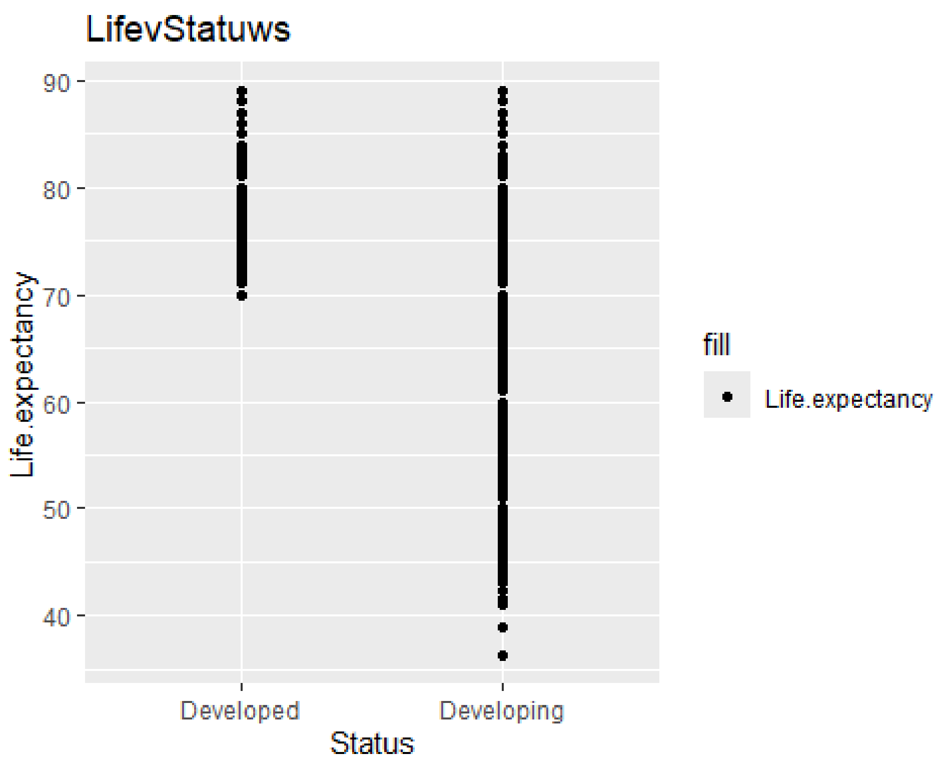
```
## [1] 3.929774
```

Bivariate analysis

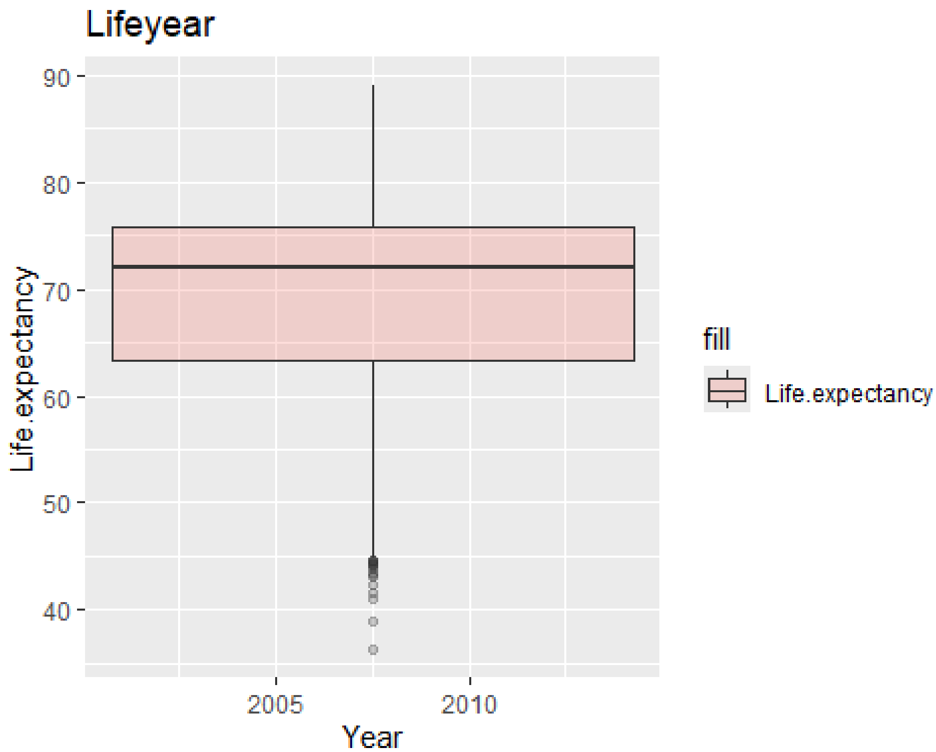
```r
#Life.expectancy vs Adult.Mortality

LifevsAdulMort<-ggplot(data)+aes(x=Adult.Mortality, y=Life.expectancy,
color=Total.expenditure,fill="Life.expectancy")+
  geom_point(alpha=0.25)+
  ggtitle("LifevsAdulMort")
LifevsAdulMort
```



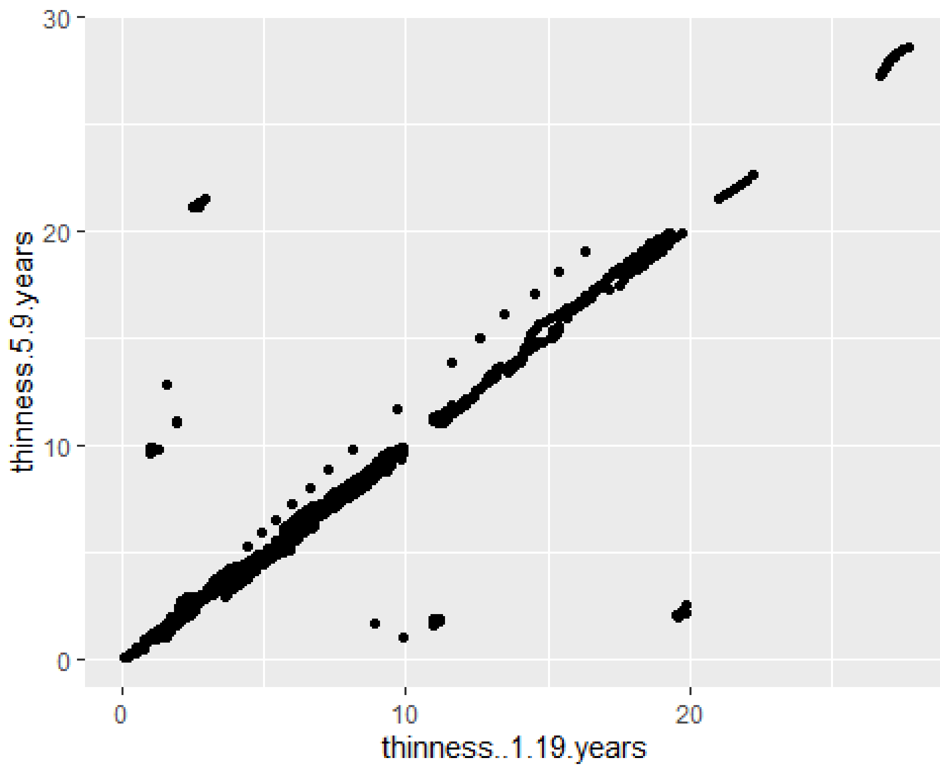

```r
#Life.expectancy vs Status

LifevStatuws<-ggplot(data)+ geom_point (aes(x=Status,
y=Life.expectancy,fill="Life.expectancy")) +
  ggtitle("LifevStatuws")
LifevStatuws
```



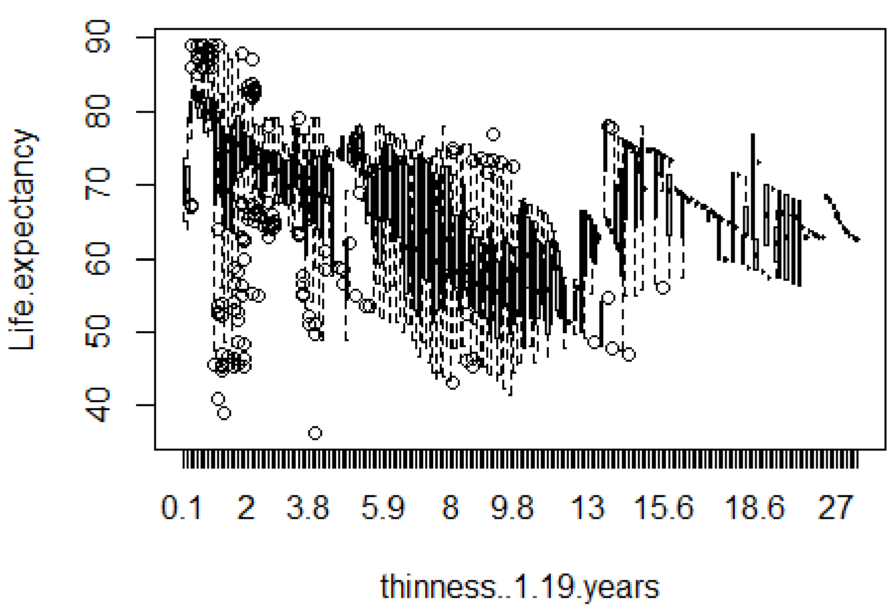

```
#Life.expectancy vs Year

Lifeyear2<-ggplot(data)+aes(x=Year,
y=Life.expectancy,fill="Life.expectancy")+
  geom_boxplot(alpha=0.25)+
  ggtitle("Lifeyear")
Lifeyear2
```



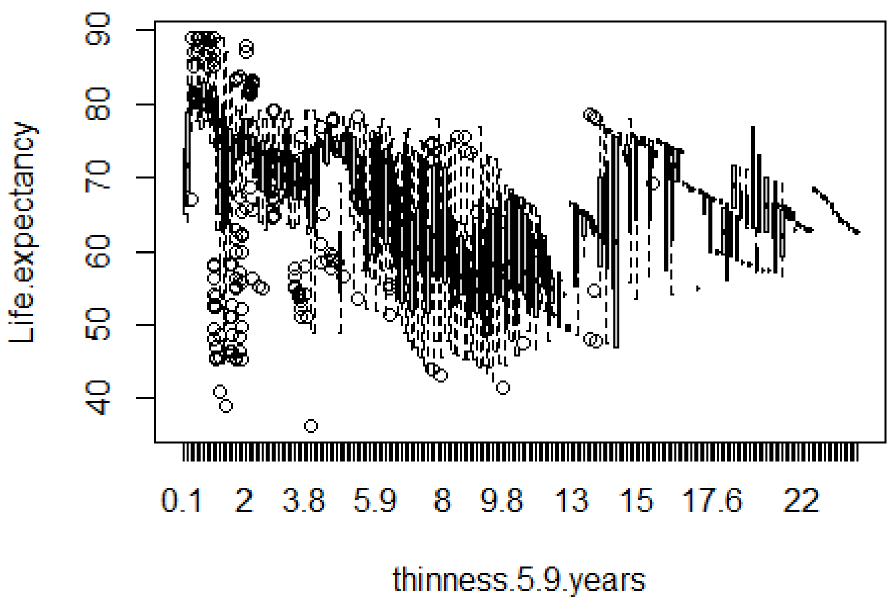

```
#thinness..1.19.years vs thinness.5.9.years

thinness1.19vs..5..9<-ggplot(data,color="red",size="cty")+
  aes(x=thinness..1.19.years, thinness.5.9.years)+
  geom_point()

thinness1.19vs..5..9
```




```r
boxplot(Life.expectancy~thinness..1.19.years, data = data)
```




```r
boxplot(Life.expectancy~ thinness.5.9.years, data = data)
```



Second part: Application of machine learning

Principal Components Analysis is applied as an unsupervised technique and multiple and logistic regression as a supervised analysis.

OECD countries are prioritized for carrying out this second part.

```
## Create the List of OECD Member Countries

ocde_countries <- c("Australia", "Austria", "Belgium", "Canada", "Chile",
"Colombia", "Czech Republic","Denmark", "Estonia", "Finland", "France",
"Germany", "Greece", "Hungary",
                    "Iceland", "Ireland", "Israel", "Italy", "Japan",
"Korea", "Latvia",
                    "Lithuania", "Luxembourg", "Mexico", "Netherlands",
"New Zealand", "Norway",
                    "Poland", "Portugal", "Slovak Republic", "Slovenia",
"Spain", "Sweden",
                    "Switzerland", "Turkey", "United Kingdom", "United
States","Costa Rica")


length(ocde_countries)

## [1] 38
```

```
#Use the list to filter the data before performing PCA
# Filter data to include only OECD countries

data_ocde <- data %>% filter(Country %in% ocde_countries)
length(data_ocde)

## [1] 22

# Show the first rows of the filtered dataframe
head(data_ocde)

##     Country Year    Status Life.expectancy Adult.Mortality
infant.deaths
## 1 Australia 2015 Developed            82.8              59
1
## 2 Australia 2014 Developed            82.7               6
1
## 3 Australia 2013 Developed            82.5              61
1
## 4 Australia 2012 Developed            82.3              61
1
## 5 Australia 2011 Developed            82.0              63
1
## 6 Australia 2010 Developed            81.9              64
1
##   Alcohol percentage.expenditure Hepatitis.B Measles  BMI
under.five.deaths
## 1   4.613                  0.000          93      74 66.6
1
## 2   9.710              10769.363          91     340 66.1
1
## 3   9.870              11734.854          91     158 65.5
1
## 4  10.030              11714.999          91     199 65.0
1
## 5  10.300              10986.265          92     190 64.4
1
## 6  10.520               8875.786          92      70 63.9
1
##   Polio Total.expenditure Diphtheria HIV.AIDS      GDP Population
## 1    93          5.944711         93      0.1 56554.39   23789338
## 2    92          9.420000         92      0.1 62214.69    2346694
## 3    91          9.360000         91      0.1 67792.34   23117353
## 4    92          9.360000         92      0.1 67677.63   22728254
## 5    92          9.200000         92      0.1 62245.13     223424
## 6    92          9.200000         92      0.1 51874.85     223175
##   thinness..1.19.years thinness.5.9.years
Income.composition.of.resources
## 1                  0.6                0.6
0.937
## 2                  0.6                0.6
```

```
0.936
## 3                     0.6                0.6
0.933
## 4                     0.6                0.6
0.930
## 5                     0.6                0.6
0.927
## 6                     0.7                0.6
0.927
##   Schooling
## 1      20.4
## 2      20.4
## 3      20.3
## 4      20.1
## 5      19.8
## 6      19.5
```

```
str(data_ocde)
```

```
## 'data.frame':    528 obs. of  22 variables:
##  $ Country                        : chr  "Australia" "Australia"
"Australia" "Australia" ...
##  $ Year                           : num  2015 2014 2013 2012 2011 ...
##  $ Status                         : chr  "Developed" "Developed"
"Developed" "Developed" ...
##  $ Life.expectancy                : num  82.8 82.7 82.5 82.3 82 81.9
81.7 81.3 81.3 81.2 ...
##  $ Adult.Mortality                : num  59 6 61 61 63 64 66 66 66 66
...
##  $ infant.deaths                  : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Alcohol                        : num  4.61 9.71 9.87 10.03 10.3 ...
##  $ percentage.expenditure         : num  0 10769 11735 11715 10986 ...
##  $ Hepatitis.B                    : num  93 91 91 91 92 92 94 94 94 95
...
##  $ Measles                        : num  74 340 158 199 190 70 104 65
11 0 ...
##  $ BMI                            : num  66.6 66.1 65.5 65 64.4 63.9
63.4 62.9 62.5 62 ...
##  $ under.five.deaths              : num  1 1 1 1 1 1 1 1 2 2 ...
##  $ Polio                          : num  93 92 91 92 92 92 92 92 92 92
...
##  $ Total.expenditure              : num  5.94 9.42 9.36 9.36 9.2 ...
##  $ Diphtheria                     : num  93 92 91 92 92 92 92 92 92 92
...
##  $ HIV.AIDS                       : num  0.1 0.1 0.1 0.1 0.1 0.1 0.1
0.1 0.1 0.1 ...
##  $ GDP                            : num  56554 62215 67792 67678 62245
...
##  $ Population                     : num  23789338 2346694 23117353
22728254 223424 ...
```

```
##  $ thinness..1.19.years          : num  0.6 0.6 0.6 0.6 0.6 0.7 0.7
0.7 0.7 0.7 ...
##  $ thinness.5.9.years            : num  0.6 0.6 0.6 0.6 0.6 0.6 0.6
0.6 0.6 0.6 ...
##  $ Income.composition.of.resources: num  0.937 0.936 0.933 0.93 0.927
0.927 0.925 0.921 0.918 0.915 ...
##  $ Schooling                     : num  20.4 20.4 20.3 20.1 19.8 19.5
19.1 19.1 19 20.3 ...
```

```
colSums(is.na(data_ocde))
```

```
##                          Country                            Year
##                                0                               0
##                           Status                 Life.expectancy
##                                0                               0
##                  Adult.Mortality                   infant.deaths
##                                0                               0
##                          Alcohol          percentage.expenditure
##                                0                               0
##                      Hepatitis.B                         Measles
##                                0                               0
##                              BMI               under.five.deaths
##                                0                               0
##                            Polio               Total.expenditure
##                                0                               0
##                       Diphtheria                        HIV.AIDS
##                                0                               0
##                              GDP                      Population
##                                0                               0
##             thinness..1.19.years              thinness.5.9.years
##                                0                               0
## Income.composition.of.resources                       Schooling
##                                0                               0
```

```
#Add numerical variables by country (filtered)
# Average numerical variables by country for OECD countries

data_agg_ocde <- data_ocde %>%
  group_by(Country) %>%
  summarize(across(where(is.numeric), mean))

## Delete the Year and HIV/AIDS column

data_agg_ocde2 <-data_agg_ocde[,-c(2,15)]
data_agg_ocde2
```

```
## # A tibble: 33 × 19
##    Country    Life.expectancy Adult.Mortality infant.deaths Alcohol
##    <chr>                <dbl>           <dbl>         <dbl>   <dbl>
##  1 Australia             81.8            63.2             1    9.81
##  2 Austria               81.5            65.8             0   11.8
```

```
## 3 Belgium 80.7 70.2 0.25 10.6
## 4 Canada 81.7 64.6 2 7.84
## 5 Chile 79.4 63.6 2 6.83
## 6 Colombia 73.3 124. 13.9 4.43
## 7 Costa Rica 78.6 69.4 1 3.97
## 8 Denmark 79.3 66.1 0 10.3
## 9 Estonia 74.9 170. 0 8.74
## 10 Finland 80.7 68.9 0 9.24
## # ℹ 23 more rows
## # ℹ 14 more variables: percentage.expenditure <dbl>, Hepatitis.B
<dbl>,
## #   Measles <dbl>, BMI <dbl>, under.five.deaths <dbl>, Polio <dbl>,
## #   Total.expenditure <dbl>, Diphtheria <dbl>, GDP <dbl>, Population
<dbl>,
## #   thinness..1.19.years <dbl>, thinness.5.9.years <dbl>,
## #   Income.composition.of.resources <dbl>, Schooling <dbl>

# Show the first rows of the added dataframe
data_agg_ocde

## # A tibble: 33 × 21
##    Country     Year Life.expectancy Adult.Mortality infant.deaths
Alcohol
##    <chr>      <dbl>           <dbl>           <dbl>         <dbl>
<dbl>
##  1 Australia  2008.            81.8            63.2          1
9.81
##  2 Austria    2008.            81.5            65.8          0
11.8
##  3 Belgium    2008.            80.7            70.2          0.25
10.6
##  4 Canada     2008.            81.7            64.6          2
7.84
##  5 Chile      2008.            79.4            63.6          2
6.83
##  6 Colombia   2008.            73.3           124.          13.9
4.43
##  7 Costa Rica 2008.            78.6            69.4          1
3.97
##  8 Denmark    2008.            79.3            66.1          0
10.3
##  9 Estonia    2008.            74.9           170.           0
8.74
## 10 Finland    2008.            80.7            68.9          0
9.24
## # ℹ 23 more rows
## # ℹ 15 more variables: percentage.expenditure <dbl>, Hepatitis.B
<dbl>,
## #   Measles <dbl>, BMI <dbl>, under.five.deaths <dbl>, Polio <dbl>,
## #   Total.expenditure <dbl>, Diphtheria <dbl>, HIV.AIDS <dbl>, GDP
```

```
<dbl>,
## #   Population <dbl>, thinness..1.19.years <dbl>, thinness.5.9.years
<dbl>,
## #   Income.composition.of.resources <dbl>, Schooling <dbl>

# Check for missing values
sum(is.na(data_agg_ocde))

## [1] 0

# Ensure relevant variables are in the correct format
str(data_agg_ocde)

## tibble [33 × 21] (S3: tbl_df/tbl/data.frame)
##  $ Country                        : chr [1:33] "Australia" "Austria"
"Belgium" "Canada" ...
##  $ Year                           : num [1:33] 2008 2008 2008 2008
2008 ...
##  $ Life.expectancy                : num [1:33] 81.8 81.5 80.7 81.7
79.5 ...
##  $ Adult.Mortality                : num [1:33] 63.2 65.8 70.2 64.6
63.6 ...
##  $ infant.deaths                  : num [1:33] 1 0 0.25 2 2 ...
##  $ Alcohol                        : num [1:33] 9.81 11.76 10.64 7.84
6.83 ...
##  $ percentage.expenditure         : num [1:33] 5332 4928 2392 4694 776
...
##  $ Hepatitis.B                    : num [1:33] 92.6 81.1 74.5 38.6
78.7 ...
##  $ Measles                        : num [1:33] 103.94 77.25 81.56
129.5 1.06 ...
##  $ BMI                            : num [1:33] 55.9 48.3 50.9 55.9
55.6 ...
##  $ under.five.deaths              : num [1:33] 1.38 0 1 2 2.19 ...
##  $ Polio                          : num [1:33] 86.8 86 97.8 85.3 83.6
...
##  $ Total.expenditure              : num [1:33] 8.66 4.79 5.96 6.98
6.85 ...
##  $ Diphtheria                     : num [1:33] 86.9 86.8 97.3 91.1
88.8 ...
##  $ HIV.AIDS                       : num [1:33] 0.1 0.1 0.1 0.1 0.1 0.1
0.1 0.1 0.1 0.1 ...
##  $ GDP                            : num [1:33] 34638 33827 16915 29383
6202 ...
##  $ Population                     : num [1:33] 4587010 6474880 2884043
11364054 14671764 ...
##  $ thinness..1.19.years           : num [1:33] 0.669 1.731 0.863 0.506
0.888 ...
##  $ thinness.5.9.years             : num [1:33] 0.625 1.938 0.856 0.438
0.912 ...
##  $ Income.composition.of.resources: num [1:33] 0.918 0.862 0.878 0.892
```

```
0.801 ...
##  $ Schooling                      : num [1:33] 20 15.4 16.8 15.9 14.9
...

#It is placed as a comment so as not to extend the deployment too much of
codes
#View(data_agg_ocde)

## Save the countries variable

paises <- data_agg_ocde2$Country
paises

##  [1] "Australia"   "Austria"     "Belgium"     "Canada"      "Chile"
##  [6] "Colombia"    "Costa Rica"  "Denmark"     "Estonia"     "Finland"
## [11] "France"      "Germany"     "Greece"      "Hungary"     "Iceland"
## [16] "Ireland"     "Israel"      "Italy"       "Japan"       "Latvia"
## [21] "Lithuania"   "Luxembourg"  "Mexico"      "Netherlands" "New
Zealand"
## [26] "Norway"      "Poland"      "Portugal"    "Slovenia"    "Spain"
## [31] "Sweden"      "Switzerland" "Turkey"

# Convert 'Country' column to row names

rownames(data_agg_ocde2) <- data_agg_ocde2$Country


# Remove the 'Country' column as it is now in the row names
data_agg_ocde2 <- data_agg_ocde2[, -which(names(data_agg_ocde2) ==
"Country")]
data_agg_ocde2

## # A tibble: 33 × 18
##    Life.expectancy Adult.Mortality infant.deaths Alcohol
percentage.expenditure
##              <dbl>           <dbl>         <dbl>   <dbl>
<dbl>
##  1            81.8            63.2          1       9.81
5332.
##  2            81.5            65.8          0      11.8
4928.
##  3            80.7            70.2          0.25   10.6
2392.
##  4            81.7            64.6          2       7.84
4694.
##  5            79.4            63.6          2       6.83
776.
##  6            73.3           124.          13.9     4.43
517.
##  7            78.6            69.4          1       3.97
972.
```

```
## 8          79.3          66.1         0      10.3
5313.
## 9          74.9         170.          0       8.74
860.
## 10         80.7          68.9         0       9.24
2889.
## # ℹ 23 more rows
## # ℹ 13 more variables: Hepatitis.B <dbl>, Measles <dbl>, BMI <dbl>,
## #   under.five.deaths <dbl>, Polio <dbl>, Total.expenditure <dbl>,
## #   Diphtheria <dbl>, GDP <dbl>, Population <dbl>,
thinness..1.19.years <dbl>,
## #   thinness.5.9.years <dbl>, Income.composition.of.resources <dbl>,
## #   Schooling <dbl>

#It is placed as a comment so as not to extend the deployment too much of
codes
#View(data_agg_ocde2)

## Convert Country column to rows
#the country becomes an index in order to reference the variables to a
specific observation
#defining its correlative value, based on the average of the data

data_agg_ocde3 <- data.frame(data_agg_ocde2)
rownames(data_agg_ocde3) <- paises
data_agg_ocde3

##              Life.expectancy Adult.Mortality infant.deaths   Alcohol
## Australia           81.81250         63.1875        1.0000  9.808938
## Austria             81.48125         65.7500        0.0000 11.759563
## Belgium             80.68125         70.1875        0.2500 10.640813
## Canada              81.68750         64.6250        2.0000  7.838313
## Chile               79.45000         63.6250        2.0000  6.830188
## Colombia            73.28750        124.2500       13.8750  4.431438
## Costa Rica          78.59375         69.3750        1.0000  3.967688
## Denmark             79.25625         66.0625        0.0000 10.327063
## Estonia             74.94375        169.6875        0.0000  8.739563
## Finland             80.71250         68.8750        0.0000  9.243938
## France              82.21875         73.1250        2.9375 11.917688
## Germany             81.17500         71.2500        2.5000 11.190188
## Greece              81.21875         73.6250        0.1875  8.541438
## Hungary             73.82500        147.0625        0.5625 11.001438
## Iceland             82.44375         49.3750        0.0000  7.287688
## Ireland             80.15000         72.4375        0.0000 12.151438
## Israel              81.30000         59.5000        0.9375  2.629563
## Italy               82.18750         54.1875        2.0000  8.038313
## Japan               82.53750         57.1250        2.8750  6.888938
## Latvia              73.73125        161.8125        0.0000  8.598313
## Lithuania           72.80625        117.2500        0.0000 12.131438
## Luxembourg          80.78125         67.5625        0.0000 11.465188
```

```
## Mexico             75.71875         111.0625        39.6250  5.082688
## Netherlands        81.13125          61.6250         1.0000  8.582063
## New Zealand        81.33750          71.5000         0.0000  9.166250
## Norway             81.79375          66.2500         0.0000  6.234375
## Poland             75.65000         107.5625         2.2500  9.662063
## Portugal           79.99375          58.8750         0.1875 11.736438
## Slovenia           79.73125          76.4375         0.0000 10.370813
## Spain              82.06875          63.6250         1.6250  9.685813
## Sweden             82.51875          59.1875         0.0000  6.782063
## Switzerland        82.33125          55.7500         0.0000  9.980188
## Turkey             73.91250          98.3750        26.9375  1.620813
##               percentage.expenditure Hepatitis.B   Measles      BMI
## Australia                  5332.2265    92.63090  103.9375 55.86250
## Austria                    4928.4392    81.06250   77.2500 48.28750
## Belgium                    2392.4327    74.50000   81.5625 50.89375
## Canada                     4694.0790    38.64269  129.5000 55.86250
## Chile                       775.5408    78.66039    1.0625 55.58750
## Colombia                    516.6402    79.00000    9.5000 49.54375
## Costa Rica                  972.3591    79.31250    0.1875 46.72500
## Denmark                    5313.3358    81.09437   15.7500 55.82500
## Estonia                     859.5750    85.39859    3.5000 56.68125
## Finland                    2889.3155    81.09437    2.8125 52.30000
## France                     3751.5066    51.25000 2661.6250 51.98125
## Germany                    3900.8903    77.43750 1497.1875 51.99375
## Greece                     1759.2468    88.93750   24.8750 58.68125
## Hungary                     376.8311    81.09437    2.0625 56.93125
## Iceland                    4991.5953    81.09437    0.0000 51.07500
## Ireland                    4867.3126    83.24648  174.5000 53.68750
## Israel                     1467.5358    97.43750  138.9375 54.98750
## Italy                      2937.1377    95.50000 1961.3750 56.15000
## Japan                      3923.0503    81.09437 6875.8125 25.60625
## Latvia                      530.6052    88.93750    3.3125 51.30625
## Lithuania                  1015.7538    94.75000   14.9375 49.23125
## Luxembourg                 8177.5763    92.37500    1.1250 47.82500
## Mexico                      465.3275    94.43750   11.0000 51.41875
## Netherlands                3805.6870    70.81488  251.9375 53.96875
## New Zealand                2922.1478    70.18750   94.9375 56.62500
## Norway                     4658.8139    81.09437    6.3125 50.81250
## Poland                      310.6823    97.37500   47.4375 53.73125
## Portugal                   1614.8970    87.62500    7.2500 43.63125
## Slovenia                   1556.3345    81.09437    6.0625 52.20000
## Spain                      2332.6802    92.50000  449.8750 58.66875
## Sweden                     4438.1632    74.25238   18.9375 56.25000
## Switzerland                9801.8104    81.09437  397.5000 51.43750
## Turkey                      253.4172    87.06250 5272.9375 56.41250
##               under.five.deaths   Polio Total.expenditure Diphtheria
GDP
## Australia                1.3750 86.7500          8.655919    86.8750
34637.565
## Austria                  0.0000 86.0000          4.792169    86.7500
```

```
33827.476
## Belgium                    1.0000 97.7500          5.960919    97.3125
16915.306
## Canada                     2.0000 85.3125          6.982794    91.0625
29382.908
## Chile                      2.1875 83.6250          6.850919    88.8125
6202.344
## Colombia                  16.3750 84.2500          6.315294    79.1875
3321.661
## Costa Rica                 1.0000 69.9375          8.452794    79.7500
3957.227
## Denmark                    0.0000 88.0000          7.705294    88.0000
33067.408
## Estonia                    0.0000 94.2500          5.703419    94.1250
8340.433
## Finland                    0.0000 97.1875          8.447169    98.2500
25268.650
## France                     3.4375 98.2500          6.517794    98.0625
26465.551
## Germany                    3.2500 94.8125          4.190294    89.8750
24337.749
## Greece                     0.4375 91.4375          8.845294    96.6250
16454.236
## Hungary                    0.8750 98.9375          7.654669    99.0000
8513.642
## Iceland                    0.0000 89.7500          8.383419    89.7500
30159.503
## Ireland                    0.0000 86.1250          7.617169    86.1250
33835.272
## Israel                     1.0000 94.2500          7.382794    89.0000
18860.476
## Italy                      2.3125 96.1250          8.600294    94.7500
21234.782
## Japan                      4.0000 96.0000          6.439669    96.2500
24892.545
## Latvia                     0.0000 95.0625          6.307169    95.0000
7951.825
## Lithuania                  0.0000 88.8125          6.509669    94.2500
9007.459
## Luxembourg                 0.0000 98.0625          7.447169    98.9375
53257.013
## Mexico                    46.5000 95.0625          6.065919    95.0000
5179.331
## Netherlands                1.0000 96.8125          7.066544    96.8125
34964.720
## New Zealand                0.0000 89.1250          8.692794    70.6250
14775.555
## Norway                     0.0000 93.0625          8.889669    87.9375
27434.947
## Poland                     2.6875 96.7500          6.352169    98.7500
```

```
6792.564
## Portugal             0.3750 96.5000          7.827169    97.0000
11598.626
## Slovenia             0.0000 95.1875          8.673419    95.0000
11441.044
## Spain                1.8750 96.7500          8.374669    96.7500
17093.460
## Sweden               0.0000 98.3125          9.683419    98.3125
29334.991
## Switzerland          0.0000 95.3750          6.078419    94.5625
57362.875
## Turkey              32.2500 80.8125          5.638419    80.7500
3983.918
##                Population thinness..1.19.years thinness.5.9.years
## Australia    4587009.88              0.66875            0.62500
## Austria      6474879.88              1.73125            1.93750
## Belgium      2884042.56              0.86250            0.85625
## Canada      11364053.81              0.50625            0.43750
## Chile       14671763.94              0.88750            0.91250
## Colombia    31767432.62              2.30000            2.07500
## Costa Rica   2309299.44              1.96250            1.90625
## Denmark      4260081.38              1.16250            0.93750
## Estonia       791848.69              2.07500            2.13750
## Finland      3493082.31              0.90000            0.80625
## France      27581733.12              0.62500            0.60000
## Germany     38757347.44              1.11875            1.10625
## Greece       1550208.44              0.81250            0.73125
## Hungary      1604902.25              1.91875            1.91250
## Iceland       186177.62              0.95625            0.90000
## Ireland      3599794.56              0.30000            0.21875
## Israel         27862.88              1.14375            1.10000
## Italy       27643788.94              0.51250            0.52500
## Japan          97384.06              1.81250            1.54375
## Latvia       1174562.69              2.40625            2.43125
## Lithuania    1926212.12              2.93750            2.96875
## Luxembourg    265276.38              0.95000            0.91250
## Mexico      27585265.19              1.73125            1.66875
## Netherlands  9775704.38              1.02500            0.96250
## New Zealand 12753375.12              0.31250            0.30000
## Norway       2614432.31              0.76250            0.70000
## Poland      16053249.62              2.20625            2.36250
## Portugal     1032225.38              0.71875            0.53750
## Slovenia      401279.06              1.76875            1.79375
## Spain       26542854.12              0.60000            0.50000
## Sweden       5514868.31              1.35000            1.30625
## Switzerland  5913241.81              0.53750            0.39375
## Turkey      33501352.81              5.01250            4.85000
##             Income.composition.of.resources Schooling
## Australia                         0.9181250  20.03750
## Austria                           0.8623750  15.38750
```

```
## Belgium                              0.8777500  16.78750
## Canada                               0.8921875  15.87500
## Chile                                0.8013125  14.90000
## Colombia                             0.6818750  12.23125
## Costa Rica                           0.7369375  12.83750
## Denmark                              0.8998750  17.19375
## Estonia                              0.8233125  15.93750
## Finland                              0.8729375  17.29375
## France                               0.8705625  15.90000
## Germany                              0.8945000  16.60000
## Greece                               0.8423125  15.93750
## Hungary                              0.8043125  15.11875
## Iceland                              0.8853125  18.15625
## Ireland                              0.8915000  17.65625
## Israel                               0.8731875  15.71250
## Italy                                0.8580625  15.93125
## Japan                                0.8765625  14.97500
## Latvia                               0.7925000  15.56875
## Lithuania                            0.8066875  16.10000
## Luxembourg                           0.8781250  13.63750
## Mexico                               0.7288750  12.32500
## Netherlands                          0.8997500  17.05625
## New Zealand                          0.8911875  18.86875
## Norway                               0.9314375  17.46875
## Poland                               0.8131875  15.25000
## Portugal                             0.8050000  15.93750
## Slovenia                             0.8604375  16.47500
## Spain                                0.8505625  16.35625
## Sweden                               0.8931250  15.86875
## Switzerland                          0.9110625  15.39375
## Turkey                               0.7033125  12.67500
```

Principal Component Analysis (PCA)

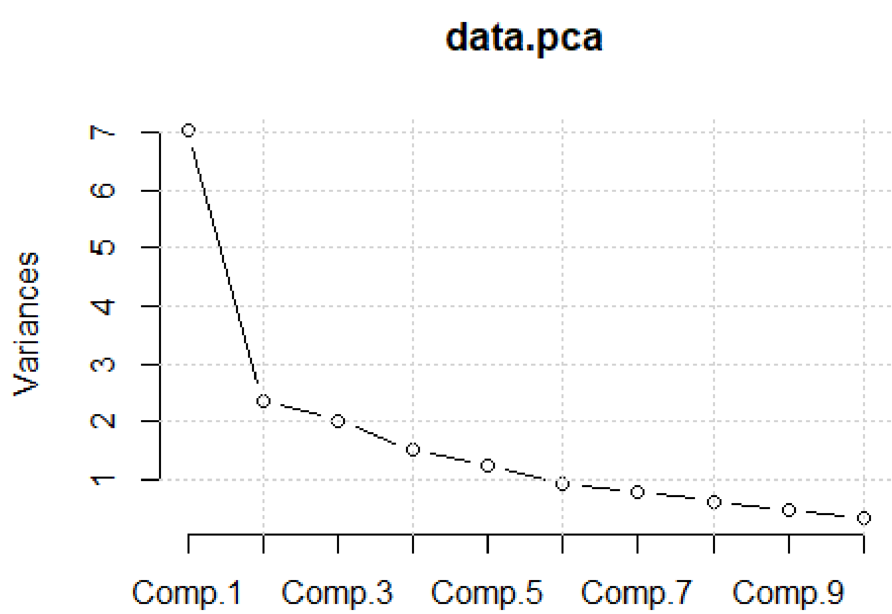
```
# Scale the data (it is one of the steps of PCA)
#Ensures that each attribute has the same level of contribution, so that one variable does not dominate the others

# Normalize data
data_normalized <- scale(data_agg_ocde3)

# Perform PCA on the normalized data
data.pca <- princomp(data_normalized, cor = TRUE, scores = TRUE)

## Print the standard deviations of each component and its correlation

data.pca

## Call:
## princomp(x = data_normalized, cor = TRUE, scores = TRUE)
##
```

```
## Standard deviations:
##      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
Comp.6
## 2.653748289 1.538642002 1.420537577 1.233506976 1.115127205
0.971717258
##      Comp.7      Comp.8      Comp.9     Comp.10     Comp.11
Comp.12
## 0.886284474 0.798038565 0.687892835 0.588585389 0.531472189
0.390820767
##     Comp.13     Comp.14     Comp.15     Comp.16     Comp.17
Comp.18
## 0.297963269 0.215800678 0.186996081 0.117030825 0.041077224
0.009460971
##
##  18  variables and  33 observations.
```

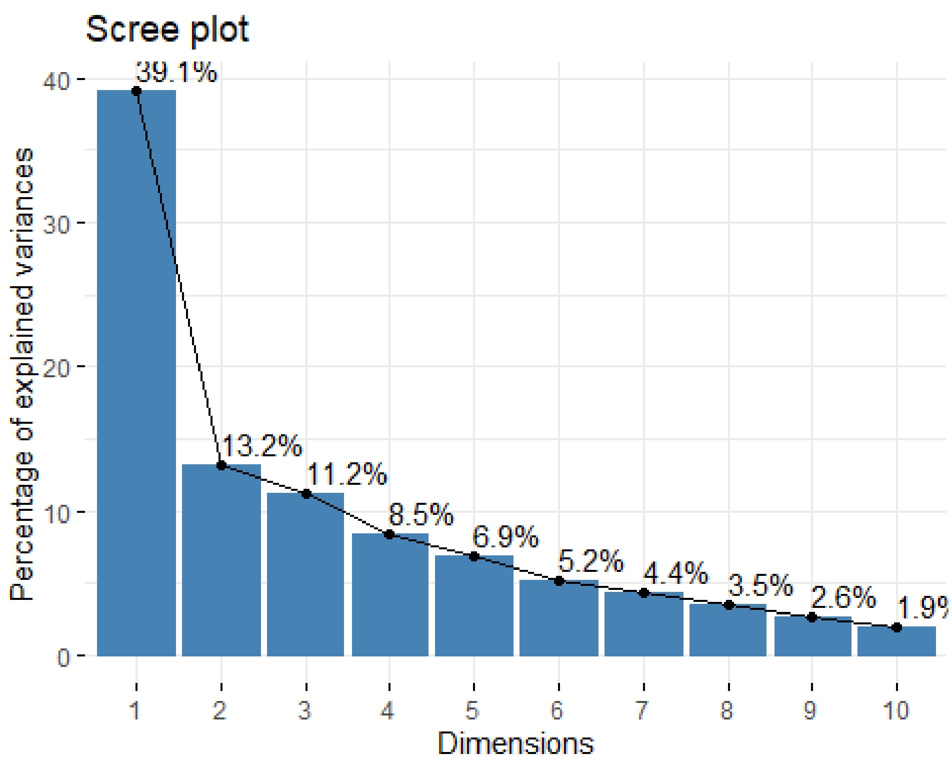
```r
## PCA Summary, function to display a summary of the data.pca object
#PCA is a dimensionality reduction and machine learning method used to
simplify a large data set into a smaller set while still maintaining
significant patterns and trends.
#describe a data set in terms of new uncorrelated variables

summary(data.pca)
```

```
## Importance of components:
##                           Comp.1    Comp.2    Comp.3     Comp.4
Comp.5
## Standard deviation     2.6537483 1.5386420 1.4205376 1.23350698
1.11512720
## Proportion of Variance 0.3912433 0.1315233 0.1121071 0.08452997
0.06908382
## Cumulative Proportion  0.3912433 0.5227666 0.6348737 0.71940365
0.78848746
##                            Comp.6    Comp.7     Comp.8    Comp.9
Comp.10
## Standard deviation     0.97171726 0.8862845 0.79803856 0.6878928
0.58858539
## Proportion of Variance 0.05245747 0.0436389 0.03538142 0.0262887
0.01924626
## Cumulative Proportion  0.84094493 0.8845838 0.91996525 0.9462539
0.96550021
##                           Comp.11     Comp.12     Comp.13     Comp.14
## Standard deviation     0.53147219 0.390820767 0.297963269 0.215800678
## Proportion of Variance 0.01569237 0.008485604 0.004932339 0.002587218
## Cumulative Proportion  0.98119258 0.989678187 0.994610526 0.997197745
##                            Comp.15      Comp.16      Comp.17
Comp.18
## Standard deviation     0.186996081 0.1170308253 4.107722e-02
9.460971e-03
## Proportion of Variance 0.001942641 0.0007609008 9.374102e-05
```

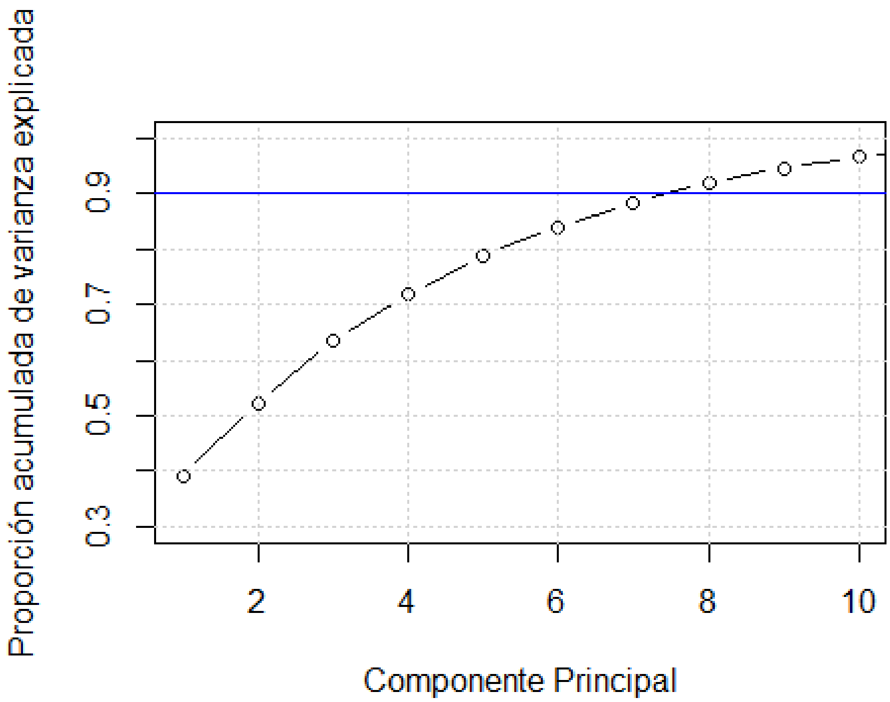
```
4.972777e-06
## Cumulative Proportion  0.999140385 0.9999012862 9.999950e-01
1.000000e+00

## Scree plot, PCA tries to put maximum possible information in the first
component, then maximum remaining information in the second and so on,
until having something like shown in the scree plot below

screeplot(data.pca, type = "lines")
grid()
```




```
##Scree Plot
#use the fviz_eig() function from the R package factoextra to display the
eigenvalues of a #principal component analysis (PCA)


fviz_eig(data.pca, addlabels = TRUE)
```




```
#The resulting graph can help determine the number of principal
components to retain for later analysis.

##Another method to select components through cumulative variances and
set a cumulative variance value

# Explained variance
pc.var <- data.pca$sdev^2

# Proportion of variation explained
pc.pvar <- pc.var / sum(pc.var)

# cumulative ratio
#shows the cumulative 90% of data variability, better representing the
data
#to decide by size or shape of the data
plot(cumsum(pc.pvar),xlab = "Componente Principal", ylab = "Proporción
acumulada de varianza explicada",xlim = c(1,10), ylim = c(0.3,1), type =
'b')
grid()

abline(h = 0.9, col = "blue")
```




```
## Attributes or variables of data.pca

attributes(data.pca)

## $names
## [1] "sdev"     "loadings" "center"   "scale"    "n.obs"    "scores"
"call"
##
## $class
## [1] "princomp"

## Interpretation of loadings

data.pca$loadings

##
## Loadings:
##                                   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Comp.6
## Life.expectancy                    0.310  0.299                0.165
0.131
## Adult.Mortality                   -0.239 -0.388               -0.236
## infant.deaths                     -0.277  0.268         0.313  0.123 -
0.167
## Alcohol                            0.204 -0.270 -0.243        -0.359
## percentage.expenditure             0.275  0.249 -0.143        -0.143 -
0.463
```

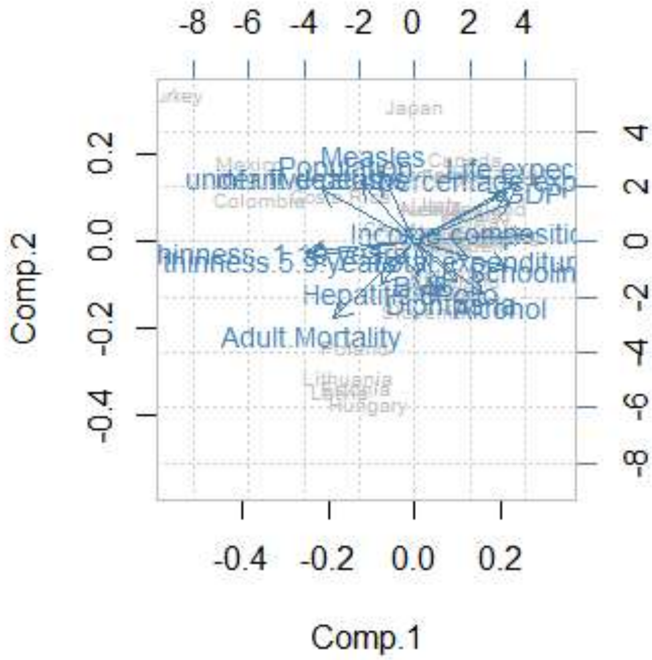
```
## Hepatitis.B                          -0.101 -0.214                0.493 -
0.492
## Measles                                     0.358 -0.327 -0.243
0.332
## BMI                                        -0.175  0.366  0.500 -0.159 -
0.161
## under.five.deaths                    -0.278  0.268         0.308  0.124 -
0.163
## Polio                                 0.126 -0.207 -0.462  0.342  0.202
0.146
## Total.expenditure                     0.159         0.352         0.536
## Diphtheria                                 -0.257 -0.489  0.205  0.240
0.111
## GDP                                   0.283  0.202 -0.185        -0.108 -
0.469
## Population                           -0.160  0.303         0.457 -0.218
0.179
## thinness..1.19.years                 -0.321        -0.103 -0.250        -
0.118
## thinness.5.9.years                   -0.318               -0.238        -
0.123
## Income.composition.of.resources  0.347
## Schooling                             0.275 -0.133  0.152
##                                      Comp.7 Comp.8 Comp.9 Comp.10 Comp.11
Comp.12
## Life.expectancy                                     0.138  0.128   0.213
## Adult.Mortality                             -0.119 -0.301  0.160  -0.502
0.261
## infant.deaths                                      -0.383 -0.137   0.169
## Alcohol                                      0.366        -0.657
## percentage.expenditure                      -0.149        -0.105  -0.248
## Hepatitis.B                           0.315  0.523  0.192  0.121
0.137
## Measles                               0.427               -0.128  -0.262
0.486
## BMI                                   0.311 -0.355  0.278
0.311
## under.five.deaths                                  -0.386 -0.138   0.171
## Polio                                 0.119 -0.109 -0.136  0.225  -0.224  -
0.482
## Total.expenditure                           -0.262        -0.546  -0.354  -
0.139
## Diphtheria                           -0.163 -0.273  0.234 -0.114   0.314
0.316
## GDP                                         -0.210                -0.156
## Population                            0.234  0.240  0.372         -0.271  -
0.248
## thinness..1.19.years                  0.273 -0.268        -0.143   0.152  -
0.268
## thinness.5.9.years                    0.283 -0.258  0.121 -0.116   0.210  -
```

```
0.290
## Income.composition.of.resources  0.285 -0.129 -0.176  0.205   0.114
## Schooling                        0.530  0.117 -0.456          0.208
##                                 Comp.13 Comp.14 Comp.15 Comp.16
Comp.17 Comp.18
## Life.expectancy                  0.238   0.639   0.454           0.109
## Adult.Mortality                 -0.103   0.484   0.146
## infant.deaths
0.703
## Alcohol                          0.312   0.119
## percentage.expenditure          -0.111                  -0.690
## Hepatitis.B
## Measles                          0.211  -0.153
## BMI                              0.327  -0.118
## under.five.deaths
-0.705
## Polio                            0.335  -0.202   0.122
## Total.expenditure                        0.121
## Diphtheria                      -0.449
## GDP                                     -0.196   0.154   0.685
## Population                      -0.417   0.133
## thinness..1.19.years                                            0.723
## thinness.5.9.years                       0.231                 -0.662
## Income.composition.of.resources          0.311  -0.760
## Schooling                       -0.406  -0.167   0.354
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
Comp.9
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
1.000
## Proportion Var  0.056  0.056  0.056  0.056  0.056  0.056  0.056  0.056
0.056
## Cumulative Var  0.056  0.111  0.167  0.222  0.278  0.333  0.389  0.444
0.500
##                Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16
Comp.17
## SS loadings      1.000   1.000   1.000   1.000   1.000   1.000   1.000
1.000
## Proportion Var   0.056   0.056   0.056   0.056   0.056   0.056   0.056
0.056
## Cumulative Var   0.556   0.611   0.667   0.722   0.778   0.833   0.889
0.944
##                Comp.18
## SS loadings      1.000
## Proportion Var   0.056
## Cumulative Var   1.000

#The loadings represent the correlation between the original variables
and the principal components. Loadings tell you how much each original
variable contributes to a specific principal component. High loadings
```

```r
(positive or negative) indicate that the original variable has a strong relationship with the principal component.
#If a variable has a high loading in the first principal component, it means that this variable is important to define that component.
#The higher the loading, the better the representation of the correlation between the original data and the CPA. The value depends on the researcher

## Visualization of the components of interest

data.pca$loadings[ , 1:2]
```

```
##                                      Comp.1      Comp.2
## Life.expectancy                  0.31006892  0.29857763
## Adult.Mortality                 -0.23929625 -0.38788538
## infant.deaths                   -0.27697737  0.26758879
## Alcohol                          0.20409274 -0.27000270
## percentage.expenditure           0.27531904  0.24893102
## Hepatitis.B                     -0.10061981 -0.21402693
## Measles                         -0.09825740  0.35776505
## BMI                              0.01002165 -0.17505361
## under.five.deaths               -0.27768405  0.26828194
## Polio                            0.12645178 -0.20734510
## Total.expenditure                0.15881663 -0.08481128
## Diphtheria                       0.08768470 -0.25695316
## GDP                              0.28318938  0.20184587
## Population                      -0.15988994  0.30343500
## thinness..1.19.years            -0.32131310 -0.05318009
## thinness.5.9.years              -0.31781635 -0.08331136
## Income.composition.of.resources  0.34737868  0.03362721
## Schooling                        0.27541291 -0.13258207
```

```r
## Visualization of the loadings and scores of the first two components
#It is possible to visualize the similarities and dissimilarities of the samples, and also shows the impact of each attribute on each of the main components.

biplot(data.pca, col = c("gray","steelblue"), cex = c(0.6, 0.9))
grid()
```




```
#Higher life expectancy, GDP,Percentage expenditure in countries such as
Canada, Germany, Switzerland, France
#Inversely predominate Measles infections and infant deaths in countries
like Mexico, Colombia, Costa Rica
#Diseases such as Polio, Dipheria, Alcohol are related to the countries
Belgium, Finland, Slovenia and Portugal.
#Adult mortality, thinness..1.19.years, thinness..519.years in aging
countries such as Poland, Lithuania, Hungary, Latvia, Estonia



#The first component explains the greatest variability of the data and
the others with little significance accumulate proportion
#includes information on the proportion of variance explained by each
principal component, the # loadings (i.e., the weights assigned to each
variable in each principal component), and other relevant # statistics.

## visualization of the first 2 components
data.pca$loadings[, 1:2]
```

```
##                                      Comp.1      Comp.2
## Life.expectancy                  0.31006892  0.29857763
## Adult.Mortality                 -0.23929625 -0.38788538
## infant.deaths                   -0.27697737  0.26758879
## Alcohol                          0.20409274 -0.27000270
## percentage.expenditure           0.27531904  0.24893102
## Hepatitis.B                     -0.10061981 -0.21402693
```

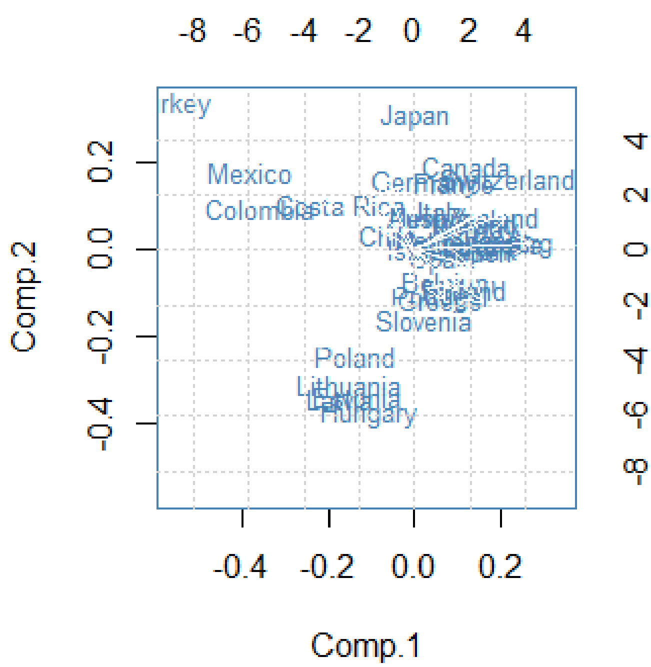
```
## Measles                          -0.09825740  0.35776505
## BMI                               0.01002165 -0.17505361
## under.five.deaths                -0.27768405  0.26828194
## Polio                             0.12645178 -0.20734510
## Total.expenditure                 0.15881663 -0.08481128
## Diphtheria                        0.08768470 -0.25695316
## GDP                               0.28318938  0.20184587
## Population                       -0.15988994  0.30343500
## thinness..1.19.years             -0.32131310 -0.05318009
## thinness.5.9.years               -0.31781635 -0.08331136
## Income.composition.of.resources  0.34737868  0.03362721
## Schooling                         0.27541291 -0.13258207

sort(data.pca$loadings[, 1:2])

##  [1] -0.38788538 -0.32131310 -0.31781635 -0.27768405 -0.27697737 -
0.27000270
##  [7] -0.25695316 -0.23929625 -0.21402693 -0.20734510 -0.17505361 -
0.15988994
## [13] -0.13258207 -0.10061981 -0.09825740 -0.08481128 -0.08331136 -
0.05318009
## [19]  0.01002165  0.03362721  0.08768470  0.12645178  0.15881663
0.20184587
## [25]  0.20409274  0.24893102  0.26758879  0.26828194  0.27531904
0.27541291
## [31]  0.28318938  0.29857763  0.30343500  0.31006892  0.34737868
0.35776505

#In this case we can select the first two main components, since they
explain 90% of the total variability. The main decreases occur in the
first two, then the decrease is practically the same for the other
components - equal size
# The first component is a linear combination with the original variables
centered, distinguishes between the scenario of populations with good
income composition, good life expectancy
#high GDP, percentage of spending, schooling, level of alcohol in
relation to a scenario of deaths of children under five years of age,
thinness.5.9.years and 1 to 19 years, child deaths and adult mortality
and hepatitis B.
# The second component distinguishes between the scenario of populations
with muscle mass, good life expectancy, infant deaths, GDP, percentage of
expenditure, good income composition
#as opposed to adult mortality,alcohol,diphtheria,infant death,Polio

## PCA Scores
#a higher value means a greater relationship with the axis

head(data.pca$scores)

##                 Comp.1     Comp.2     Comp.3      Comp.4     Comp.5
Comp.6
```

```
## Australia  2.7332300  0.1664999  1.4841451 -0.07848003  0.2668857 -
1.218078704
## Austria    0.5326316  0.6933611 -0.6374236 -1.36602815 -1.8570869 -
1.235998597
## Belgium    1.2395229 -0.6924777 -0.9496106  0.30469732 -0.4624215
0.808888105
## Canada     1.8096212  1.7110811  1.0999305  0.41521666 -2.0825754
1.128829184
## Chile     -0.8562513  0.3037329  1.6901082  0.17498554 -0.1580283
0.823689781
## Colombia  -5.4251681  0.8690074  1.3061389  0.24828619 -0.8470037 -
0.003724787
##               Comp.7     Comp.8     Comp.9     Comp.10    Comp.11
Comp.12
## Australia  1.5119412  0.5502826 -0.9097370 -0.46423697  0.1311191
0.46082194
## Austria   -0.2178594  0.7740036  0.5060416  0.11425708  0.9459290 -
0.30333134
## Belgium   -0.3821053  0.2533384 -0.3471436  0.53490412  1.0024677 -
0.06497806
## Canada    -1.4271207 -1.7537997 -0.2669738  0.02259291  0.4126496
0.32412446
## Chile     -0.7543341  0.5895778  0.8868338  0.42027996  0.5326102
0.50527337
## Colombia  -1.3743974  0.6069034  0.1904596  0.21207468 -1.2628762 -
0.66266952
##                Comp.13      Comp.14     Comp.15     Comp.16
Comp.17
## Australia -0.560353239  0.104016369  0.22654010  0.04334422 -
0.017586700
## Austria    0.460256050  0.435412351  0.21947475  0.19074616 -
0.028985606
## Belgium    0.273133966 -0.186268646  0.04628877 -0.17957729
0.001499979
## Canada    -0.269383591 -0.008163006 -0.20503210 -0.02994772 -
0.009179346
## Chile     -0.001553621 -0.210728194 -0.12474272 -0.11868485 -
0.063366423
## Colombia  -0.356727593 -0.254264030  0.17237838  0.06969228
0.043208567
##                Comp.18
## Australia -0.013416764
## Austria    0.002178636
## Belgium   -0.033769343
## Canada     0.006084420
## Chile      0.005097716
## Colombia  -0.011109090

#The scores are the values of the new variables (principal components)
for each observation in the original data set. In other words, they are
```

```
the coordinates of the observations in the principal component space. The
scores tell you how the original observations are projected in the
principal component space. If the observations have similar scores on the
first two components, that means that those observations are similar in
terms of the variations captured by those components.


## PCA Scores of the first two components

head(data.pca$scores[ ,1:2])
```

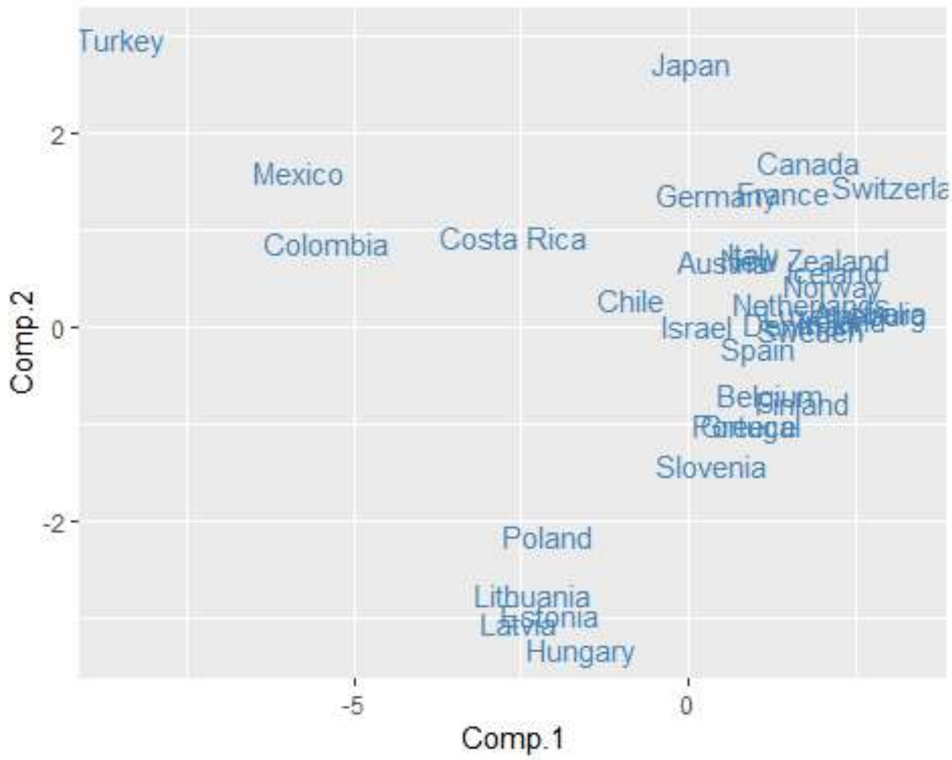
```
##                Comp.1     Comp.2
## Australia  2.7332300  0.1664999
## Austria    0.5326316  0.6933611
## Belgium    1.2395229 -0.6924777
## Canada     1.8096212  1.7110811
## Chile     -0.8562513  0.3037329
## Colombia  -5.4251681  0.8690074
```

```
## Viewing PCA Scores

biplot(data.pca, col = c("steelblue","White"), cex = c(0.8, 0.01))
grid()
```

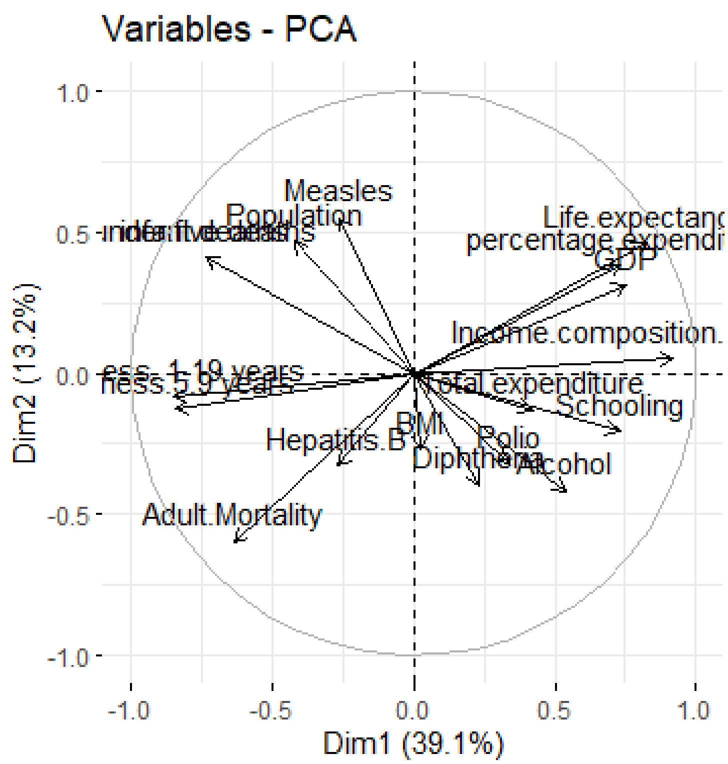



```
## Visualization of PCA Scores with GGPLOT

scores <- data.frame(data.pca$scores)
```

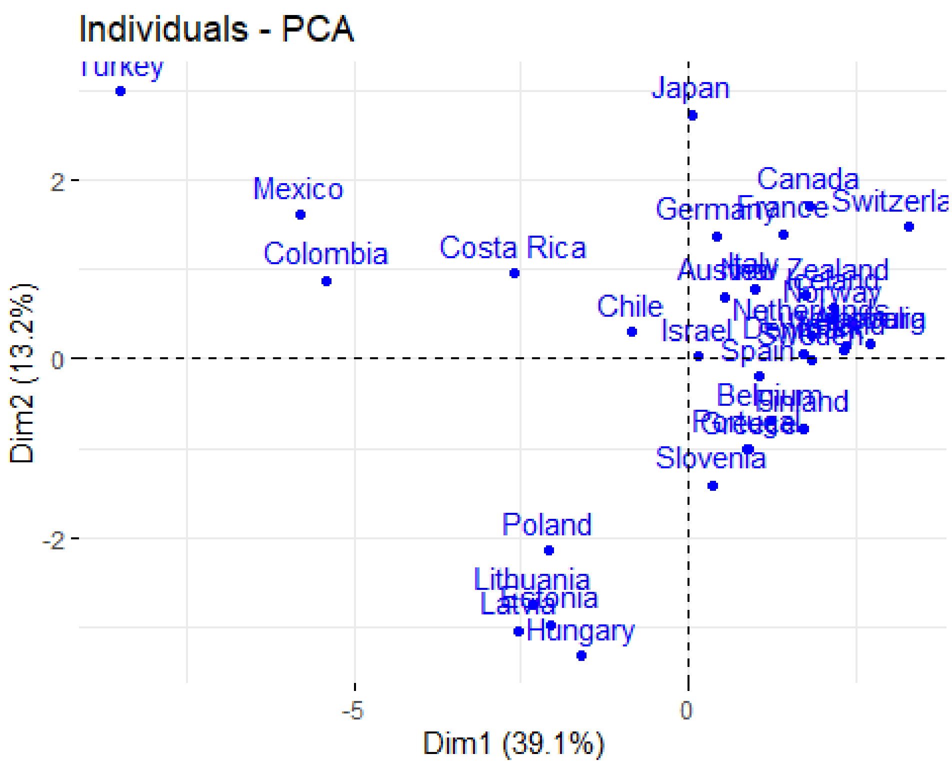
```r
ggplot(data = scores, aes(x = Comp.1, y = Comp.2, label =
rownames(scores))) +
  geom_text(size = 4, col = "steelblue")
```



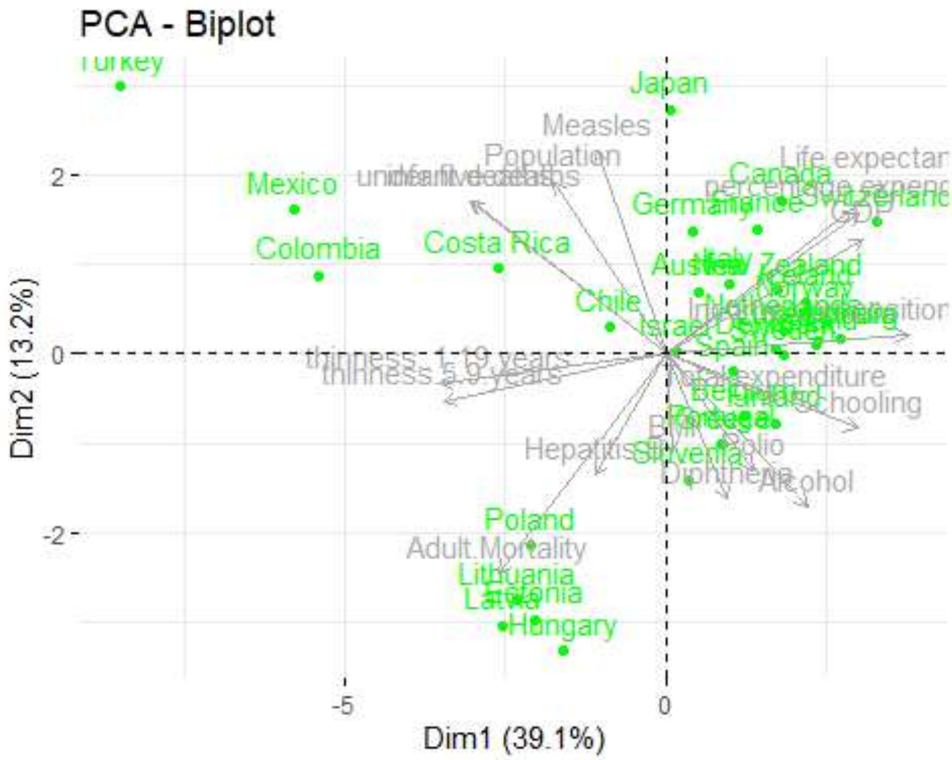

```r
## Loading graph

# Graph of the variables
#Correlation circle in range from 0 to 1
fviz_pca_var(data.pca, col.var = "black")
```



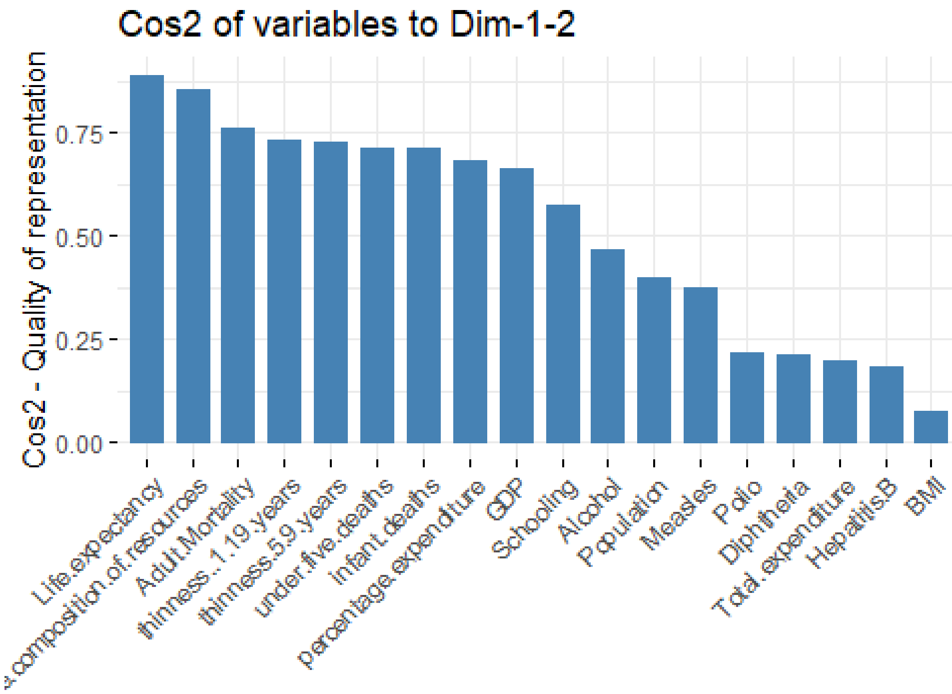

```r
#maintain similar distances or distribution in the quadrants

#shows the impact of each attribute on each of the main components.

#1.All the variables that are grouped together are positively correlated
with each other.
#2.The greater the distance between the variable and the origin, the
better represented that variable will be.
#3.Negatively correlated variables are shown on opposite sides of the
biplot origin

##Score graph


# Individuals Chart
fviz_pca_ind(data.pca, col.ind = "blue")
```



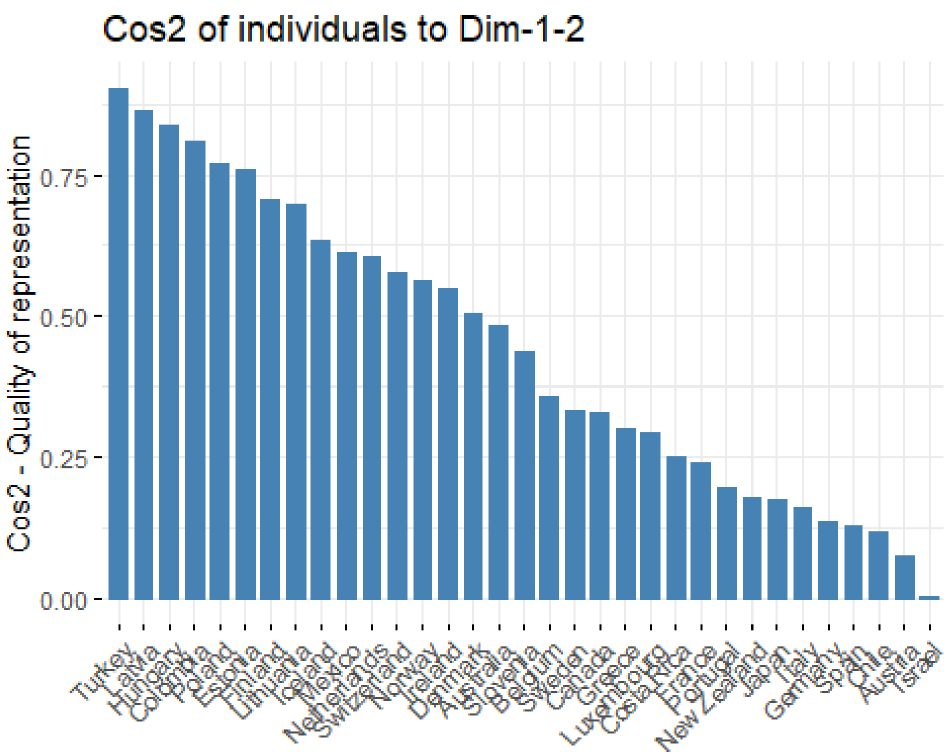

```
##Biplot of Loading and scores

# Individuals Chart
fviz_pca_biplot(data.pca, col.var = "darkgrey",col.ind = "green")
```



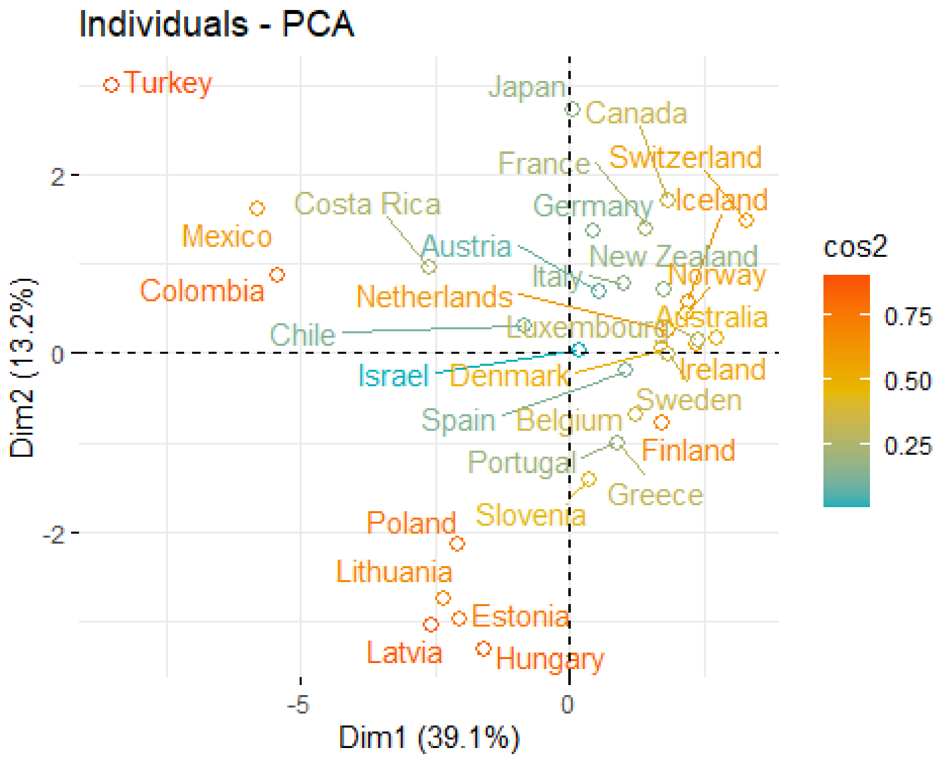

```r
## Contribution of each variable
#The goal of the third visualization is to determine the extent to which each variable is represented in a given component. This quality of representation is called Cos2, it corresponds to the #squared cosine and is computed with the function fviz_cos2.
fviz_cos2(data.pca, choice = "var", axes = 1:2)
```



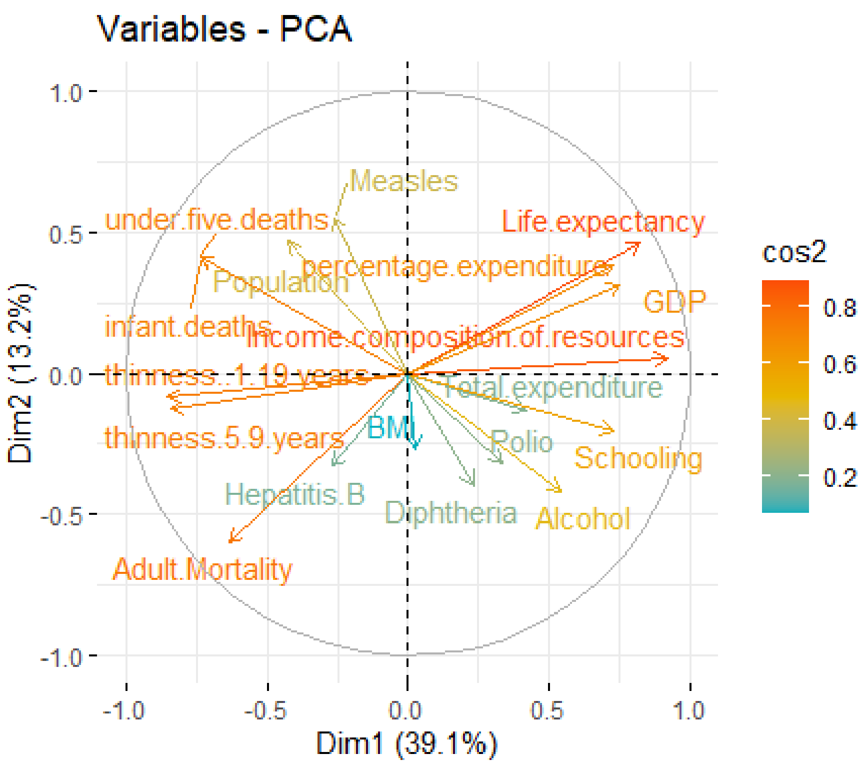

```
#A low value means that the variable is not perfectly represented by that
component. but considering more variables can enrich the analysis

## Contribution of each individual
#a stronger value better representation, but considering more countries
can enrich the analysis

fviz_cos2(data.pca, choice = "ind", axes = 1:2)
```




```r
##Biplot of individuals


# Biplot of individuals in components 1 and 2
#a stronger value better representation,
fviz_pca_ind(data.pca,
             axes = c(1, 2),          # Especifica las dos primeras 
componentes
             col.ind = "cos2",        # Color según los valores de cos2: 
calidad de representación
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE,            # Evita que se solape el texto de 
las etiquetas
             label = "ind",           # Agrega etiquetas a los individuos
             habillage = "none",      # No agrupa los individuos por 
ningún factor
             pointshape = 21,         # Forma del punto (círculo)
             pointsize = 2.5)         # Tamaño del punto
```




```
##Biplot of variables

# Biplot of the variables in components 1 and 2
#a stronger value better representation,

fviz_pca_var(data.pca,
             axes = c(1, 2),          # Especifica las dos primeras
componentes
             col.var = "cos2",        # Color según los valores de cos2:
calidad de representación
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE,            # Evita que se solape el texto de
las etiquetas
             label = "var",           # Agrega etiquetas a los individuos
             habillage = "none",      # No agrupa los individuos por
ningún factor
             pointshape = 21,         # Forma del punto (círculo)
             pointsize = 0.8)         # Tamaño del punto
```




```r
## Square cosines of the variables
#The analysis by cosines is clearer than by loadings to determine the
variables by components

# Cos2 of variables7
#a stronger value better representation,add up to 1
variables_cos2 <- data.pca$loadings[, 1:2]^2
print("Cos2 de las variables:")

## [1] "Cos2 de las variables:"

print(variables_cos2)

##                                          Comp.1      Comp.2
## Life.expectancy                    0.0961427356 0.089148603
## Adult.Mortality                    0.0572626966 0.150455068
## infant.deaths                      0.0767164654 0.071603761
## Alcohol                            0.0416538461 0.072901460
## percentage.expenditure             0.0758005720 0.061966652
## Hepatitis.B                        0.0101243454 0.045807528
## Measles                            0.0096545161 0.127995834
## BMI                                0.0001004335 0.030643767
## under.five.deaths                  0.0771084308 0.071975202
## Polio                              0.0159900523 0.042991991
## Total.expenditure                  0.0252227218 0.007192953
## Diphtheria                         0.0076886070 0.066024928
## GDP                                0.0801962225 0.040741755
```

```
## Population                      0.0255647917 0.092072799
## thinness..1.19.years            0.1032421053 0.002828122
## thinness.5.9.years              0.1010072350 0.006940784
## Income.composition.of.resources 0.1206719502 0.001130789
## Schooling                       0.0758522727 0.017578005
```

```
## Square cosines of individuals

# Calculate the scores of the individuals (coordinates) and then square
and normalize

individuos_cos2 <- data.pca$scores[, 1:2]^2 / rowSums(data.pca$scores^2)
print("Cos2 de los individuos (Primera y Segunda Componente):")
```

```
## [1] "Cos2 de los individuos (Primera y Segunda Componente):"
```

```
print(individuos_cos2[, 1:2])
```

```
##                   Comp.1       Comp.2
## Australia   0.4819077578 1.788296e-03
## Austria     0.0272718443 4.621464e-02
## Belgium     0.2719568579 8.487945e-02
## Canada      0.1744525943 1.559708e-01
## Chile       0.1060600646 1.334545e-02
## Colombia    0.7896304425 2.026023e-02
## Costa Rica  0.2225399621 2.950139e-02
## Denmark     0.5053679570 4.169603e-04
## Estonia     0.2460457328 5.139060e-01
## Finland     0.5865794513 1.177566e-01
## France      0.1222183896 1.161238e-01
## Germany     0.0120545821 1.252692e-01
## Greece      0.1341486818 1.681254e-01
## Hungary     0.1584657715 6.781140e-01
## Iceland     0.5915395674 4.128671e-02
## Ireland     0.5474078709 8.838376e-04
## Israel      0.0027150401 9.255903e-05
## Italy       0.1010239450 6.072266e-02
## Japan       0.0001293057 1.748780e-01
## Latvia      0.3584608813 5.048292e-01
## Lithuania   0.2923330909 4.078603e-01
## Luxembourg  0.2934961608 1.217347e-03
## Mexico      0.5690532228 4.366102e-02
## Netherlands 0.5941541671 1.135123e-02
## New Zealand 0.1535185291 2.488430e-02
## Norway      0.5399544543 2.208238e-02
## Poland      0.3782003889 3.920412e-01
## Portugal    0.0859250878 1.116914e-01
## Slovenia    0.0268833038 4.110635e-01
## Spain       0.1254864487 4.783021e-03
## Sweden      0.3313065089 3.123429e-05
```

```
## Switzerland 0.4808076076 9.539548e-02
## Turkey      0.8047474203 9.897674e-02
```

Regression

Regression Multiple

```
#Valuing 7 and 2 components at the regression level
## Regression model
#y=Bo+B1X1+B2X2+...+BpXp+E
#y=XB+E (vector matricial),Regression shows n, n nuples reg multiple
(xi,yi) define most relevant variables
#Intercept, beta for each variable, p value,  income p is less than 0.005
but almost 0..
# p value other than 0,

# Extract the first 7 main components
pca_components <- data.pca$scores[, 1:7]

# Convert to data.frame and add 'Life.expectancy' variable
pca_df <- as.data.frame(pca_components)

pca_df$Life.expectancy <- data_agg_ocde3$Life.expectancy

# Fit the multiple linear regression model
model <- lm(Life.expectancy ~ ., data = pca_df)

# Model Summary
summary(model)

##
## Call:
## lm(formula = Life.expectancy ~ ., data = pca_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21981 -0.43644 -0.06252  0.60590  2.49779
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 79.46875    0.15283 519.982  < 2e-16 ***
## Comp.1       0.97554    0.05759  16.939 3.26e-15 ***
## Comp.2       0.93939    0.09933   9.457 9.76e-10 ***
## Comp.3       0.03132    0.10759   0.291 0.773367
## Comp.4       0.10138    0.12390   0.818 0.420926
## Comp.5       0.51775    0.13705   3.778 0.000875 ***
## Comp.6       0.41159    0.15728   2.617 0.014840 *
## Comp.7       0.12643    0.17244   0.733 0.470258
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8779 on 25 degrees of freedom
## Multiple R-squared:  0.941,  Adjusted R-squared:  0.9245
## F-statistic: 56.97 on 7 and 25 DF,  p-value: 8.622e-14
```

#Residuals: These values indicate the distribution of the model's residuals, which should ideally be centered around zero.

#Coefficients: #(Intercept): The intercept is 79.255082 with an extremely small p-value (< 2e-16), indicating that it is significantly different from zero. #Comp.1: Estimated coefficient of 1.036428 with a p-value of 1.71e-10 (***), indicating that this component is highly significant. #Comp.2: Estimated coefficient of 1.106976 with a p-value of 2.18e-07 (***)***, also highly significant.

#Residual standard error: 0.9161, indicating that the standard deviation of the residuals is relatively small. #Multiple R-squared: 0.9547, meaning the model explains 95.47% of the variability in Life.expectancy. #Adjusted R-squared: 0.9336, adjusted for the number of predictors, still indicating a good fit. #F-statistic: 45.19 with a p-value of 5.829e-09, indicating that the model as a whole is highly significant.

`{r}

Extract the first 2 main components

pca_components2 <- data.pca$scores[, 1:2]

Convert to data.frame and add 'Life.expectancy' variable

pca_df2 <- as.data.frame(pca_components2)

pca_df2$Life.expectancy < -data_{a}gg_{o}cde$3Life.expectancy

Fit the multiple linear regression model

model2 <- lm(Life.expectancy ~ ., data = pca_df2)

Model Summary

summary(model2)

````
```r
# Assuming data_new is your new observations
# data_new must be in the same format as data_normalized_dep
````
````

```
# Normalize the new data (the same way as the original data)
#data_new_normalized <- scale(data_new, center = TRUE, scale = TRUE)

# Apply the PCA transformation to the new data
#pca_new_scores <- predict(data.pca, newdata = data_new_normalized)

# Extract the first 2 principal components from the new data
#pca_new_components2 <- pca_new_scores[, 1:2]

# Convert to data.frame
#pca_new_df2 <- as.data.frame(pca_new_components2)
#colnames(pca_new_df2) <- c("PC1", "PC2")

# Use the fitted model to make predictions
#predicted_life_expectancy <- predict(model2, newdata = pca_new_df2)

# Show predictions
#print(predicted_life_expectancy)

#Test

#Heteroscedasticity Test

#bptest(model2)
```

#Multicollinearity Test #VIF Test (Variance Inflation Factor)

```
#vif(model2)
#The VIF test value is less than 5. Therefore, there is no
multicollinearity.
```

#Autocorrelation Test

```
#dwtest(model2)
#bgtest(model2)
#2 to 4 indicate negative autocorrelation
#best value, most meaningful, best effect
#p value greater than 0.05 (depends on our confidence level) and r
squared is

#Regression shows n, n nuples reg multiple (xi,yi) define most relevant
variables
#y=alfa+betax+epsilon lineal simple}
#Ordinary least squares methods-minimize the errors in the line and
return estimates of the parameters
#valuation of atypicals-analyze them, there are assumptions of linearity,
homoesthecity, normality of the residuals and independence,
multicollinearity (correlation=they explain the same thing)
```

Logistic regression

Average numerical variables by country for OECD countries

```
data_agg_ocde_logistic <- data_ocde %>%
  group_by(Country, Status) %>%
  summarize(across(where(is.numeric), mean))

data_agg_ocde_logistic

## # A tibble: 33 × 22
## # Groups:   Country [33]
##    Country    Status  Year Life.expectancy Adult.Mortality
infant.deaths Alcohol
##    <chr>      <chr>  <dbl>           <dbl>           <dbl>
<dbl>   <dbl>
##  1 Australia  Devel… 2008.            81.8            63.2           1
9.81
##  2 Austria    Devel… 2008.            81.5            65.8           0
11.8
##  3 Belgium    Devel… 2008.            80.7            70.2
0.25   10.6
##  4 Canada     Devel… 2008.            81.7            64.6           2
7.84
##  5 Chile      Devel… 2008.            79.4            63.6           2
6.83
##  6 Colombia   Devel… 2008.            73.3           124.
13.9     4.43
##  7 Costa Rica Devel… 2008.            78.6            69.4           1
3.97
##  8 Denmark    Devel… 2008.            79.3            66.1           0
10.3
##  9 Estonia    Devel… 2008.            74.9           170.            0
8.74
## 10 Finland    Devel… 2008.            80.7            68.9           0
9.24
## # ℹ 23 more rows
## # ℹ 15 more variables: percentage.expenditure <dbl>, Hepatitis.B
<dbl>,
## #   Measles <dbl>, BMI <dbl>, under.five.deaths <dbl>, Polio <dbl>,
## #   Total.expenditure <dbl>, Diphtheria <dbl>, HIV.AIDS <dbl>, GDP
<dbl>,
## #   Population <dbl>, thinness..1.19.years <dbl>, thinness.5.9.years
<dbl>,
## #   Income.composition.of.resources <dbl>, Schooling <dbl>
```

Delete the Year and HIV/AIDS column

```
data_agg_ocde_logistic <-data_agg_ocde_logistic[,-c(3,16)]
data_agg_ocde_logistic
```

```
## # A tibble: 33 × 20
## # Groups:   Country [33]
##    Country     Status     Life.expectancy Adult.Mortality infant.deaths
Alcohol
##    <chr>       <chr>                <dbl>           <dbl>         <dbl>
<dbl>
##  1 Australia   Developed             81.8            63.2          1
9.81
##  2 Austria     Developed             81.5            65.8          0
11.8
##  3 Belgium     Developed             80.7            70.2          0.25
10.6
##  4 Canada      Developing            81.7            64.6          2
7.84
##  5 Chile       Developing            79.4            63.6          2
6.83
##  6 Colombia    Developing            73.3           124.          13.9
4.43
##  7 Costa Rica Developing             78.6            69.4          1
3.97
##  8 Denmark     Developed             79.3            66.1          0
10.3
##  9 Estonia     Developing            74.9           170.           0
8.74
## 10 Finland     Developing            80.7            68.9          0
9.24
## # ℹ 23 more rows
## # ℹ 14 more variables: percentage.expenditure <dbl>, Hepatitis.B
<dbl>,
## #   Measles <dbl>, BMI <dbl>, under.five.deaths <dbl>, Polio <dbl>,
## #   Total.expenditure <dbl>, Diphtheria <dbl>, GDP <dbl>, Population
<dbl>,
## #   thinness..1.19.years <dbl>, thinness.5.9.years <dbl>,
## #   Income.composition.of.resources <dbl>, Schooling <dbl>
```

Convert Country column to rows

```
data_agg_ocde_logistic_df <- data.frame(data_agg_ocde_logistic)
rownames(data_agg_ocde_logistic_df) <- data_agg_ocde_logistic$Country
data_agg_ocde_logistic_df <-data_agg_ocde_logistic_df[,-c(1)]
data_agg_ocde_logistic_df
```

```
##                   Status Life.expectancy Adult.Mortality infant.deaths
Alcohol
## Australia     Developed        81.81250         63.1875        1.0000
9.808938
## Austria       Developed        81.48125         65.7500        0.0000
11.759563
## Belgium       Developed        80.68125         70.1875        0.2500
10.640813
## Canada       Developing        81.68750         64.6250        2.0000
```

```
7.838313
## Chile       Developing        79.45000        63.6250        2.0000
6.830188
## Colombia    Developing        73.28750       124.2500       13.8750
4.431438
## Costa Rica  Developing        78.59375        69.3750        1.0000
3.967688
## Denmark      Developed        79.25625        66.0625        0.0000
10.327063
## Estonia     Developing        74.94375       169.6875        0.0000
8.739563
## Finland     Developing        80.71250        68.8750        0.0000
9.243938
## France      Developing        82.21875        73.1250        2.9375
11.917688
## Germany      Developed        81.17500        71.2500        2.5000
11.190188
## Greece      Developing        81.21875        73.6250        0.1875
8.541438
## Hungary      Developed        73.82500       147.0625        0.5625
11.001438
## Iceland      Developed        82.44375        49.3750        0.0000
7.287688
## Ireland      Developed        80.15000        72.4375        0.0000
12.151438
## Israel      Developing        81.30000        59.5000        0.9375
2.629563
## Italy        Developed        82.18750        54.1875        2.0000
8.038313
## Japan        Developed        82.53750        57.1250        2.8750
6.888938
## Latvia       Developed        73.73125       161.8125        0.0000
8.598313
## Lithuania    Developed        72.80625       117.2500        0.0000
12.131438
## Luxembourg   Developed        80.78125        67.5625        0.0000
11.465188
## Mexico      Developing        75.71875       111.0625       39.6250
5.082688
## Netherlands  Developed        81.13125        61.6250        1.0000
8.582063
## New Zealand  Developed        81.33750        71.5000        0.0000
9.166250
## Norway       Developed        81.79375        66.2500        0.0000
6.234375
## Poland       Developed        75.65000       107.5625        2.2500
9.662063
## Portugal     Developed        79.99375        58.8750        0.1875
11.736438
## Slovenia     Developed        79.73125        76.4375        0.0000
```

```
10.370813
## Spain        Developed        82.06875         63.6250         1.6250
9.685813
## Sweden       Developed        82.51875         59.1875         0.0000
6.782063
## Switzerland  Developed        82.33125         55.7500         0.0000
9.980188
## Turkey      Developing        73.91250         98.3750        26.9375
1.620813
##             percentage.expenditure Hepatitis.B   Measles      BMI
## Australia                5332.2265    92.63090  103.9375 55.86250
## Austria                  4928.4392    81.06250   77.2500 48.28750
## Belgium                  2392.4327    74.50000   81.5625 50.89375
## Canada                   4694.0790    38.64269  129.5000 55.86250
## Chile                     775.5408    78.66039    1.0625 55.58750
## Colombia                  516.6402    79.00000    9.5000 49.54375
## Costa Rica                972.3591    79.31250    0.1875 46.72500
## Denmark                  5313.3358    81.09437   15.7500 55.82500
## Estonia                   859.5750    85.39859    3.5000 56.68125
## Finland                  2889.3155    81.09437    2.8125 52.30000
## France                   3751.5066    51.25000 2661.6250 51.98125
## Germany                  3900.8903    77.43750 1497.1875 51.99375
## Greece                   1759.2468    88.93750   24.8750 58.68125
## Hungary                   376.8311    81.09437    2.0625 56.93125
## Iceland                  4991.5953    81.09437    0.0000 51.07500
## Ireland                  4867.3126    83.24648  174.5000 53.68750
## Israel                   1467.5358    97.43750  138.9375 54.98750
## Italy                    2937.1377    95.50000 1961.3750 56.15000
## Japan                    3923.0503    81.09437 6875.8125 25.60625
## Latvia                    530.6052    88.93750    3.3125 51.30625
## Lithuania                1015.7538    94.75000   14.9375 49.23125
## Luxembourg               8177.5763    92.37500    1.1250 47.82500
## Mexico                    465.3275    94.43750   11.0000 51.41875
## Netherlands              3805.6870    70.81488  251.9375 53.96875
## New Zealand              2922.1478    70.18750   94.9375 56.62500
## Norway                   4658.8139    81.09437    6.3125 50.81250
## Poland                    310.6823    97.37500   47.4375 53.73125
## Portugal                 1614.8970    87.62500    7.2500 43.63125
## Slovenia                 1556.3345    81.09437    6.0625 52.20000
## Spain                    2332.6802    92.50000  449.8750 58.66875
## Sweden                   4438.1632    74.25238   18.9375 56.25000
## Switzerland              9801.8104    81.09437  397.5000 51.43750
## Turkey                    253.4172    87.06250 5272.9375 56.41250
##             under.five.deaths   Polio Total.expenditure Diphtheria
GDP
## Australia              1.3750 86.7500          8.655919    86.8750
34637.565
## Austria                0.0000 86.0000          4.792169    86.7500
33827.476
## Belgium                1.0000 97.7500          5.960919    97.3125
```

```
16915.306
## Canada                    2.0000 85.3125          6.982794    91.0625
29382.908
## Chile                     2.1875 83.6250          6.850919    88.8125
6202.344
## Colombia                 16.3750 84.2500          6.315294    79.1875
3321.661
## Costa Rica                1.0000 69.9375          8.452794    79.7500
3957.227
## Denmark                   0.0000 88.0000          7.705294    88.0000
33067.408
## Estonia                   0.0000 94.2500          5.703419    94.1250
8340.433
## Finland                   0.0000 97.1875          8.447169    98.2500
25268.650
## France                    3.4375 98.2500          6.517794    98.0625
26465.551
## Germany                   3.2500 94.8125          4.190294    89.8750
24337.749
## Greece                    0.4375 91.4375          8.845294    96.6250
16454.236
## Hungary                   0.8750 98.9375          7.654669    99.0000
8513.642
## Iceland                   0.0000 89.7500          8.383419    89.7500
30159.503
## Ireland                   0.0000 86.1250          7.617169    86.1250
33835.272
## Israel                    1.0000 94.2500          7.382794    89.0000
18860.476
## Italy                     2.3125 96.1250          8.600294    94.7500
21234.782
## Japan                     4.0000 96.0000          6.439669    96.2500
24892.545
## Latvia                    0.0000 95.0625          6.307169    95.0000
7951.825
## Lithuania                 0.0000 88.8125          6.509669    94.2500
9007.459
## Luxembourg                0.0000 98.0625          7.447169    98.9375
53257.013
## Mexico                   46.5000 95.0625          6.065919    95.0000
5179.331
## Netherlands               1.0000 96.8125          7.066544    96.8125
34964.720
## New Zealand               0.0000 89.1250          8.692794    70.6250
14775.555
## Norway                    0.0000 93.0625          8.889669    87.9375
27434.947
## Poland                    2.6875 96.7500          6.352169    98.7500
6792.564
## Portugal                  0.3750 96.5000          7.827169    97.0000
```

```
11598.626
## Slovenia               0.0000 95.1875          8.673419    95.0000
11441.044
## Spain                  1.8750 96.7500          8.374669    96.7500
17093.460
## Sweden                 0.0000 98.3125          9.683419    98.3125
29334.991
## Switzerland            0.0000 95.3750          6.078419    94.5625
57362.875
## Turkey                32.2500 80.8125          5.638419    80.7500
3983.918
##               Population thinness..1.19.years thinness.5.9.years
## Australia    4587009.88               0.66875            0.62500
## Austria      6474879.88               1.73125            1.93750
## Belgium      2884042.56               0.86250            0.85625
## Canada      11364053.81               0.50625            0.43750
## Chile       14671763.94               0.88750            0.91250
## Colombia    31767432.62               2.30000            2.07500
## Costa Rica   2309299.44               1.96250            1.90625
## Denmark      4260081.38               1.16250            0.93750
## Estonia       791848.69               2.07500            2.13750
## Finland      3493082.31               0.90000            0.80625
## France      27581733.12               0.62500            0.60000
## Germany     38757347.44               1.11875            1.10625
## Greece       1550208.44               0.81250            0.73125
## Hungary      1604902.25               1.91875            1.91250
## Iceland       186177.62               0.95625            0.90000
## Ireland      3599794.56               0.30000            0.21875
## Israel         27862.88               1.14375            1.10000
## Italy       27643788.94               0.51250            0.52500
## Japan          97384.06               1.81250            1.54375
## Latvia       1174562.69               2.40625            2.43125
## Lithuania    1926212.12               2.93750            2.96875
## Luxembourg    265276.38               0.95000            0.91250
## Mexico      27585265.19               1.73125            1.66875
## Netherlands  9775704.38               1.02500            0.96250
## New Zealand 12753375.12               0.31250            0.30000
## Norway       2614432.31               0.76250            0.70000
## Poland      16053249.62               2.20625            2.36250
## Portugal     1032225.38               0.71875            0.53750
## Slovenia      401279.06               1.76875            1.79375
## Spain       26542854.12               0.60000            0.50000
## Sweden       5514868.31               1.35000            1.30625
## Switzerland  5913241.81               0.53750            0.39375
## Turkey      33501352.81               5.01250            4.85000
##             Income.composition.of.resources Schooling
## Australia                         0.9181250  20.03750
## Austria                           0.8623750  15.38750
## Belgium                           0.8777500  16.78750
## Canada                            0.8921875  15.87500
```

```
## Chile                               0.8013125  14.90000
## Colombia                            0.6818750  12.23125
## Costa Rica                          0.7369375  12.83750
## Denmark                             0.8998750  17.19375
## Estonia                             0.8233125  15.93750
## Finland                             0.8729375  17.29375
## France                              0.8705625  15.90000
## Germany                             0.8945000  16.60000
## Greece                              0.8423125  15.93750
## Hungary                             0.8043125  15.11875
## Iceland                             0.8853125  18.15625
## Ireland                             0.8915000  17.65625
## Israel                              0.8731875  15.71250
## Italy                               0.8580625  15.93125
## Japan                               0.8765625  14.97500
## Latvia                              0.7925000  15.56875
## Lithuania                           0.8066875  16.10000
## Luxembourg                          0.8781250  13.63750
## Mexico                              0.7288750  12.32500
## Netherlands                         0.8997500  17.05625
## New Zealand                         0.8911875  18.86875
## Norway                              0.9314375  17.46875
## Poland                              0.8131875  15.25000
## Portugal                            0.8050000  15.93750
## Slovenia                            0.8604375  16.47500
## Spain                               0.8505625  16.35625
## Sweden                              0.8931250  15.86875
## Switzerland                         0.9110625  15.39375
## Turkey                              0.7033125  12.67500
```

Load necessary libraries

```
library(dplyr)
library(caret)
library(lattice)
```

Convert Status variable to a factor

```
data_agg_ocde_logistic_df$Status <-
factor(data_agg_ocde_logistic_df$Status, levels = c("Developing",
"Developed"))

str(data_agg_ocde_logistic_df)

## 'data.frame':    33 obs. of  19 variables:
##  $ Status                          : Factor w/ 2 levels
"Developing","Developed": 2 2 2 1 1 1 1 2 1 1 ...
##  $ Life.expectancy                 : num  81.8 81.5 80.7 81.7 79.5 ...
##  $ Adult.Mortality                 : num  63.2 65.8 70.2 64.6 63.6 ...
```

```
##  $ infant.deaths                   : num  1 0 0.25 2 2 ...
##  $ Alcohol                         : num  9.81 11.76 10.64 7.84 6.83
...
##  $ percentage.expenditure          : num  5332 4928 2392 4694 776 ...
##  $ Hepatitis.B                     : num  92.6 81.1 74.5 38.6 78.7 ...
##  $ Measles                         : num  103.94 77.25 81.56 129.5 1.06
...
##  $ BMI                             : num  55.9 48.3 50.9 55.9 55.6 ...
##  $ under.five.deaths               : num  1.38 0 1 2 2.19 ...
##  $ Polio                           : num  86.8 86 97.8 85.3 83.6 ...
##  $ Total.expenditure               : num  8.66 4.79 5.96 6.98 6.85 ...
##  $ Diphtheria                      : num  86.9 86.8 97.3 91.1 88.8 ...
##  $ GDP                             : num  34638 33827 16915 29383 6202
...
##  $ Population                      : num  4587010 6474880 2884043
11364054 14671764 ...
##  $ thinness..1.19.years            : num  0.669 1.731 0.863 0.506 0.888
...
##  $ thinness.5.9.years              : num  0.625 1.938 0.856 0.438 0.912
...
##  $ Income.composition.of.resources: num  0.918 0.862 0.878 0.892 0.801
...
##  $ Schooling                       : num  20 15.4 16.8 15.9 14.9 ...
```

Split the data into training and test sets (70% training, 30% testing)

```
set.seed(123)
trainIndex <- createDataPartition(data_agg_ocde_logistic_df$Status, p =
.7,
                                  list = FALSE,
                                  times = 1)
dfTrain <- data_agg_ocde_logistic_df[trainIndex,]
dfTest <- data_agg_ocde_logistic_df[-trainIndex,]
```

Fit the logistic regression model

```
logistic_model <- glm(Status ~ Measles + BMI + Polio + Alcohol +
Life.expectancy,
                      data = dfTrain,
                      family = binomial)
```

Model Summary

```
summary(logistic_model)

##
## Call:
## glm(formula = Status ~ Measles + BMI + Polio + Alcohol +
```

```
Life.expectancy,
##     family = binomial, data = dfTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5336  -0.2333   0.1249   0.4187   1.2470
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -3.234e+01  2.348e+01  -1.377   0.1684
## Measles          8.732e-04  1.331e-03   0.656   0.5119
## BMI             -2.438e-01  3.078e-01  -0.792   0.4282
## Polio            7.448e-02  1.336e-01   0.557   0.5772
## Alcohol          8.621e-01  4.550e-01   1.895   0.0582 .
## Life.expectancy  4.095e-01  2.346e-01   1.746   0.0809 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 30.553  on 23  degrees of freedom
## Residual deviance: 12.131  on 18  degrees of freedom
## AIC: 24.131
##
## Number of Fisher Scoring iterations: 7
```

Make predictions on the test set

```
predictions <- predict(logistic_model, newdata = dfTest, type =
"response")
```

Convert probabilities into classes

```
predicted_classes <- ifelse(predictions > 0.5, "Developed", "Developing")
```

Convert predictions and actual labels to factors with the same levels

```
predicted_classes <- factor(predicted_classes, levels = c("Developing",
"Developed"))
dfTest$Status <- factor(dfTest$Status, levels = c("Developing",
"Developed"))
```

Create a confusion table

```
confusionMatrix(predicted_classes, dfTest$Status)
```

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction   Developing Developed
##   Developing          0         0
##   Developed           3         6
##
##                Accuracy : 0.6667
##                  95% CI : (0.2993, 0.9251)
##     No Information Rate : 0.6667
##     P-Value [Acc > NIR] : 0.6503
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : 0.2482
##
##             Sensitivity : 0.0000
##             Specificity : 1.0000
##          Pos Pred Value :    NaN
##          Neg Pred Value : 0.6667
##              Prevalence : 0.3333
##          Detection Rate : 0.0000
##    Detection Prevalence : 0.0000
##       Balanced Accuracy : 0.5000
##
##        'Positive' Class : Developing
##
```

```
dfTrain
```

```
##                 Status Life.expectancy Adult.Mortality infant.deaths
Alcohol
## Australia    Developed        81.81250         63.1875        1.0000
9.808938
## Belgium      Developed        80.68125         70.1875        0.2500
10.640813
## Chile       Developing        79.45000         63.6250        2.0000
6.830188
## Colombia    Developing        73.28750        124.2500       13.8750
4.431438
## Costa Rica  Developing        78.59375         69.3750        1.0000
3.967688
## Denmark      Developed        79.25625         66.0625        0.0000
10.327063
## Estonia     Developing        74.94375        169.6875        0.0000
8.739563
## Finland     Developing        80.71250         68.8750        0.0000
9.243938
## Germany      Developed        81.17500         71.2500        2.5000
11.190188
```

```
## Iceland      Developed        82.44375         49.3750         0.0000
7.287688
## Ireland      Developed        80.15000         72.4375         0.0000
12.151438
## Israel      Developing        81.30000         59.5000         0.9375
2.629563
## Italy        Developed        82.18750         54.1875         2.0000
8.038313
## Japan        Developed        82.53750         57.1250         2.8750
6.888938
## Latvia       Developed        73.73125        161.8125         0.0000
8.598313
## Mexico      Developing        75.71875        111.0625        39.6250
5.082688
## New Zealand  Developed        81.33750         71.5000         0.0000
9.166250
## Norway       Developed        81.79375         66.2500         0.0000
6.234375
## Portugal     Developed        79.99375         58.8750         0.1875
11.736438
## Slovenia     Developed        79.73125         76.4375         0.0000
10.370813
## Spain        Developed        82.06875         63.6250         1.6250
9.685813
## Sweden       Developed        82.51875         59.1875         0.0000
6.782063
## Switzerland  Developed        82.33125         55.7500         0.0000
9.980188
## Turkey      Developing        73.91250         98.3750        26.9375
1.620813
##             percentage.expenditure Hepatitis.B   Measles      BMI
## Australia                5332.2265    92.63090  103.9375 55.86250
## Belgium                  2392.4327    74.50000   81.5625 50.89375
## Chile                     775.5408    78.66039    1.0625 55.58750
## Colombia                  516.6402    79.00000    9.5000 49.54375
## Costa Rica                972.3591    79.31250    0.1875 46.72500
## Denmark                  5313.3358    81.09437   15.7500 55.82500
## Estonia                   859.5750    85.39859    3.5000 56.68125
## Finland                  2889.3155    81.09437    2.8125 52.30000
## Germany                  3900.8903    77.43750 1497.1875 51.99375
## Iceland                  4991.5953    81.09437    0.0000 51.07500
## Ireland                  4867.3126    83.24648  174.5000 53.68750
## Israel                   1467.5358    97.43750  138.9375 54.98750
## Italy                    2937.1377    95.50000 1961.3750 56.15000
## Japan                    3923.0503    81.09437 6875.8125 25.60625
## Latvia                    530.6052    88.93750    3.3125 51.30625
## Mexico                    465.3275    94.43750   11.0000 51.41875
## New Zealand              2922.1478    70.18750   94.9375 56.62500
## Norway                   4658.8139    81.09437    6.3125 50.81250
## Portugal                 1614.8970    87.62500    7.2500 43.63125
```

```
## Slovenia                  1556.3345    81.09437    6.0625 52.20000
## Spain                     2332.6802    92.50000  449.8750 58.66875
## Sweden                    4438.1632    74.25238   18.9375 56.25000
## Switzerland               9801.8104    81.09437  397.5000 51.43750
## Turkey                     253.4172    87.06250 5272.9375 56.41250
##              under.five.deaths   Polio Total.expenditure Diphtheria
GDP
## Australia                1.3750 86.7500          8.655919    86.8750
34637.565
## Belgium                  1.0000 97.7500          5.960919    97.3125
16915.306
## Chile                    2.1875 83.6250          6.850919    88.8125
6202.344
## Colombia                16.3750 84.2500          6.315294    79.1875
3321.661
## Costa Rica               1.0000 69.9375          8.452794    79.7500
3957.227
## Denmark                  0.0000 88.0000          7.705294    88.0000
33067.408
## Estonia                  0.0000 94.2500          5.703419    94.1250
8340.433
## Finland                  0.0000 97.1875          8.447169    98.2500
25268.650
## Germany                  3.2500 94.8125          4.190294    89.8750
24337.749
## Iceland                  0.0000 89.7500          8.383419    89.7500
30159.503
## Ireland                  0.0000 86.1250          7.617169    86.1250
33835.272
## Israel                   1.0000 94.2500          7.382794    89.0000
18860.476
## Italy                    2.3125 96.1250          8.600294    94.7500
21234.782
## Japan                    4.0000 96.0000          6.439669    96.2500
24892.545
## Latvia                   0.0000 95.0625          6.307169    95.0000
7951.825
## Mexico                  46.5000 95.0625          6.065919    95.0000
5179.331
## New Zealand              0.0000 89.1250          8.692794    70.6250
14775.555
## Norway                   0.0000 93.0625          8.889669    87.9375
27434.947
## Portugal                 0.3750 96.5000          7.827169    97.0000
11598.626
## Slovenia                 0.0000 95.1875          8.673419    95.0000
11441.044
## Spain                    1.8750 96.7500          8.374669    96.7500
17093.460
## Sweden                   0.0000 98.3125          9.683419    98.3125
```

```
29334.991
## Switzerland            0.0000 95.3750          6.078419    94.5625
57362.875
## Turkey                32.2500 80.8125          5.638419    80.7500
3983.918
##               Population thinness..1.19.years thinness.5.9.years
## Australia    4587009.88               0.66875            0.62500
## Belgium      2884042.56               0.86250            0.85625
## Chile       14671763.94               0.88750            0.91250
## Colombia    31767432.62               2.30000            2.07500
## Costa Rica   2309299.44               1.96250            1.90625
## Denmark      4260081.38               1.16250            0.93750
## Estonia       791848.69               2.07500            2.13750
## Finland      3493082.31               0.90000            0.80625
## Germany     38757347.44               1.11875            1.10625
## Iceland       186177.62               0.95625            0.90000
## Ireland      3599794.56               0.30000            0.21875
## Israel         27862.88               1.14375            1.10000
## Italy       27643788.94               0.51250            0.52500
## Japan          97384.06               1.81250            1.54375
## Latvia       1174562.69               2.40625            2.43125
## Mexico      27585265.19               1.73125            1.66875
## New Zealand 12753375.12               0.31250            0.30000
## Norway       2614432.31               0.76250            0.70000
## Portugal     1032225.38               0.71875            0.53750
## Slovenia      401279.06               1.76875            1.79375
## Spain       26542854.12               0.60000            0.50000
## Sweden       5514868.31               1.35000            1.30625
## Switzerland  5913241.81               0.53750            0.39375
## Turkey      33501352.81               5.01250            4.85000
##             Income.composition.of.resources Schooling
## Australia                         0.9181250  20.03750
## Belgium                           0.8777500  16.78750
## Chile                             0.8013125  14.90000
## Colombia                          0.6818750  12.23125
## Costa Rica                        0.7369375  12.83750
## Denmark                           0.8998750  17.19375
## Estonia                           0.8233125  15.93750
## Finland                           0.8729375  17.29375
## Germany                           0.8945000  16.60000
## Iceland                           0.8853125  18.15625
## Ireland                           0.8915000  17.65625
## Israel                            0.8731875  15.71250
## Italy                             0.8580625  15.93125
## Japan                             0.8765625  14.97500
## Latvia                            0.7925000  15.56875
## Mexico                            0.7288750  12.32500
## New Zealand                       0.8911875  18.86875
## Norway                            0.9314375  17.46875
## Portugal                          0.8050000  15.93750
```

```
## Slovenia                          0.8604375  16.47500
## Spain                             0.8505625  16.35625
## Sweden                            0.8931250  15.86875
## Switzerland                       0.9110625  15.39375
## Turkey                            0.7033125  12.67500
```

dfTest

```
##                   Status Life.expectancy Adult.Mortality infant.deaths
Alcohol
## Austria      Developed        81.48125         65.7500        0.0000
11.759563
## Canada      Developing        81.68750         64.6250        2.0000
7.838313
## France      Developing        82.21875         73.1250        2.9375
11.917688
## Greece      Developing        81.21875         73.6250        0.1875
8.541438
## Hungary      Developed        73.82500        147.0625        0.5625
11.001438
## Lithuania    Developed        72.80625        117.2500        0.0000
12.131438
## Luxembourg   Developed        80.78125         67.5625        0.0000
11.465188
## Netherlands  Developed        81.13125         61.6250        1.0000
8.582063
## Poland       Developed        75.65000        107.5625        2.2500
9.662063
##             percentage.expenditure Hepatitis.B   Measles      BMI
## Austria                  4928.4392    81.06250   77.2500 48.28750
## Canada                   4694.0790    38.64269  129.5000 55.86250
## France                   3751.5066    51.25000 2661.6250 51.98125
## Greece                   1759.2468    88.93750   24.8750 58.68125
## Hungary                   376.8311    81.09437    2.0625 56.93125
## Lithuania                1015.7538    94.75000   14.9375 49.23125
## Luxembourg               8177.5763    92.37500    1.1250 47.82500
## Netherlands              3805.6870    70.81488  251.9375 53.96875
## Poland                    310.6823    97.37500   47.4375 53.73125
##             under.five.deaths   Polio Total.expenditure Diphtheria
GDP
## Austria                0.0000 86.0000          4.792169    86.7500
33827.476
## Canada                 2.0000 85.3125          6.982794    91.0625
29382.908
## France                 3.4375 98.2500          6.517794    98.0625
26465.551
## Greece                 0.4375 91.4375          8.845294    96.6250
16454.236
## Hungary                0.8750 98.9375          7.654669    99.0000
8513.642
```

```
## Lithuania              0.0000 88.8125          6.509669    94.2500
9007.459
## Luxembourg             0.0000 98.0625          7.447169    98.9375
53257.013
## Netherlands            1.0000 96.8125          7.066544    96.8125
34964.720
## Poland                 2.6875 96.7500          6.352169    98.7500
6792.564
##              Population thinness..1.19.years thinness.5.9.years
## Austria       6474879.9              1.73125            1.93750
## Canada       11364053.8              0.50625            0.43750
## France       27581733.1              0.62500            0.60000
## Greece        1550208.4              0.81250            0.73125
## Hungary       1604902.2              1.91875            1.91250
## Lithuania     1926212.1              2.93750            2.96875
## Luxembourg     265276.4              0.95000            0.91250
## Netherlands   9775704.4              1.02500            0.96250
## Poland       16053249.6              2.20625            2.36250
##              Income.composition.of.resources Schooling
## Austria                            0.8623750  15.38750
## Canada                             0.8921875  15.87500
## France                             0.8705625  15.90000
## Greece                             0.8423125  15.93750
## Hungary                            0.8043125  15.11875
## Lithuania                          0.8066875  16.10000
## Luxembourg                         0.8781250  13.63750
## Netherlands                        0.8997500  17.05625
## Poland                             0.8131875  15.25000
```

```
predictions
```

```
##     Austria      Canada      France      Greece     Hungary
Lithuania
##   0.9971624   0.6711379   0.9998101   0.6910442   0.7034198
0.9279360
##  Luxembourg Netherlands      Poland
##   0.9981042   0.9277019   0.7527416
```

```
predicted_classes
```

```
##     Austria      Canada      France      Greece     Hungary
Lithuania
##   Developed   Developed   Developed   Developed   Developed
Developed
##  Luxembourg Netherlands      Poland
##   Developed   Developed   Developed
## Levels: Developing Developed
```

Conclusions

#Life expectancy in the study period reveals results in two segments for the population and variables under study from the comprehensive perspective, which explain 90% of the variability of the data and with the highest level of significance.

#Determining different variables or factors in common for groups of countries such as the following: #Greater life expectancy, GDP, percentage of spending related to countries such as Canada, Germany, Switzerland, France.

#Inversely, measles infections and child deaths predominate in countries such as Mexico, Colombia, Costa Rica.

#Diseases like Polio, Diferia, Alcohol are related to the countries Belgium, Finland, Slovenia and Portugal.

#Adult mortality, thinness...1.19.years, thinness...519.years in countries where their population ages such as Poland, Lithuania, Hungary, Latvia, Estonia

#Although the study period is limited to 5 years and responds to previous years, some of these factors may currently prevail or have been overcome by the countries involved due to development from public and private actions carried out recently or changes in the environmental condition, political or others factors.

#Now, to counteract the prevailing and future challenges on this issue of longevity among the studied countries or others, it is necessary to detect opportunities and strategies in public policies in order to overcome multifactorial poverty with social investment on issues such as health, employment, housing, education and social protection, which improve the income and living conditions of the population.

Bibliographic reference

#Giuseppe Ciaburro.Regression Analysis with R.2018.(https://www.google.co.cr/books/edition/Regression_Analysis_with_R/nyZKDwAAQBAJ?hl=es&gbpv=0)

#Databricks.Machine Learning Models. 2024.(https://www.databricks.com/glossary/machine-learning-models)

#IBM.What is principal component analysis (PCA).(https://www.ibm.com/topics/principal-component-analysis)

#Geeksforgeeks.Principal Component Analysis(PCA).2023.(https://www.geeksforgeeks.org/principal-component-analysis-pca/)