R-Bootcamp Assignement

Stefan Berchtold, Gabriel Mullis

March 11, 2020

Used Packages

Following packages are used in this assignement

```r
# install.packages("reshape1")
# install.packages("dplyr")
# install.packages("tidyr")
# install.packages("ggplot2")
# install.packages("ggthemes")
# install.packages("corrplot")
# install.packages("plotly")
# install.packages("tinytex")
# install.packages("knitr")
library(ggthemes)
library(ggplot2)
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(reshape2)
```

```
##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```r
library(plotly)
```

```
## 
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
## 
##     last_plot
```

```
## The following object is masked from 'package:stats':
## 
##     filter
```

```
## The following object is masked from 'package:graphics':
## 
##     layout
```

```r
library(tinytex)
library(knitr)
```

Read in the Data

```r
GDP<-readRDS("../data/GDP.RDS")
LE<-readRDS("../data/LE.RDS")
summary(GDP)
```

```
##         ï..Country.Name                 Country.Code           Indicator.Name
##   Congo         :  2    Rep.                  :  3   GDP (current US$):251
##   Korea         :  2    China                 :  2   BHS              :  1
##   Afghanistan   :  1    The                   :  2   COD              :  1
##   Albania       :  1    Arab Rep.             :  1   COG              :  1
##   Algeria       :  1    Dem. Peopleâ\200\231s Rep.:  1   EGY             :  1
##   American Samoa:  1    Dem. Rep.             :  1   FSM              :  1
##   (Other)       :256    (Other)               :254   (Other)          :  8
##          Indicator.Code          X1960            X1961
##   GDP (current US$): 13                   :126   Min.   :1.159e+07
##   NY.GDP.MKTP.CD   :251    NY.GDP.MKTP.CD  : 13   1st Qu.:4.892e+08
##                           29878659986.2386:  2   Median :2.967e+09
##                           47133390423.8826:  2   Mean   :7.486e+10
##                           1003692374.55789:  1   3rd Qu.:3.105e+10
##                           1042336802982.28:  1   Max.   :1.428e+12
##                           (Other)         :119   NA's   :132
##      X1962              X1963              X1964
##   Min.   :9.123e+06   Min.   :1.084e+07   Min.   :1.271e+07
```

```
##   1st Qu.:4.923e+08   1st Qu.:5.130e+08   1st Qu.:5.526e+08
##   Median :2.464e+09   Median :2.903e+09   Median :3.647e+09
##   Mean   :7.909e+10   Mean   :8.522e+10   Mean   :9.369e+10
##   3rd Qu.:3.281e+10   3rd Qu.:3.772e+10   3rd Qu.:3.635e+10
##   Max.   :1.533e+12   Max.   :1.651e+12   Max.   :1.809e+12
##   NA's   :130         NA's   :130         NA's   :130
##      X1965               X1966               X1967
##   Min.   :1.359e+07   Min.   :1.447e+07   Min.   :1.584e+07
##   1st Qu.:5.798e+08   1st Qu.:6.500e+08   1st Qu.:6.239e+08
##   Median :2.946e+09   Median :3.132e+09   Median :3.357e+09
##   Mean   :9.670e+10   Mean   :1.074e+11   Mean   :1.108e+11
##   3rd Qu.:3.302e+10   3rd Qu.:3.780e+10   3rd Qu.:3.138e+10
##   Max.   :1.969e+12   Max.   :2.137e+12   Max.   :2.275e+12
##   NA's   :121         NA's   :118         NA's   :114
##      X1968               X1969               X1970
##   Min.   :1.460e+07   Min.   :1.585e+07   Min.   :1.430e+07
##   1st Qu.:6.426e+08   1st Qu.:6.757e+08   1st Qu.:5.280e+08
##   Median :3.618e+09   Median :4.083e+09   Median :4.179e+09
##   Mean   :1.163e+11   Mean   :1.286e+11   Mean   :1.339e+11
##   3rd Qu.:3.261e+10   3rd Qu.:3.649e+10   3rd Qu.:4.106e+10
##   Max.   :2.456e+12   Max.   :2.709e+12   Max.   :2.965e+12
##   NA's   :110         NA's   :110         NA's   :100
##      X1971               X1972               X1973
##   Min.   :1.528e+07   Min.   :1.894e+07   Min.   :2.420e+07
##   1st Qu.:5.384e+08   1st Qu.:5.786e+08   1st Qu.:7.228e+08
##   Median :4.357e+09   Median :5.074e+09   Median :6.242e+09
##   Mean   :1.468e+11   Mean   :1.692e+11   Mean   :2.083e+11
##   3rd Qu.:4.515e+10   3rd Qu.:5.324e+10   3rd Qu.:6.439e+10
##   Max.   :3.278e+12   Max.   :3.783e+12   Max.   :4.616e+12
##   NA's   :99          NA's   :99          NA's   :99
##      X1974               X1975               X1976
##   Min.   :3.151e+07   Min.   :3.251e+07   Min.   :3.004e+07
##   1st Qu.:1.044e+09   1st Qu.:1.126e+09   1st Qu.:1.100e+09
##   Median :7.968e+09   Median :9.264e+09   Median :1.005e+10
##   Mean   :2.415e+11   Mean   :2.634e+11   Mean   :2.831e+11
##   3rd Qu.:8.843e+10   3rd Qu.:9.815e+10   3rd Qu.:9.876e+10
##   Max.   :5.324e+12   Max.   :5.929e+12   Max.   :6.448e+12
##   NA's   :98          NA's   :94          NA's   :93
##      X1977               X1978               X1979
##   Min.   :3.414e+07   Min.   :4.157e+07   Min.   :4.262e+07
##   1st Qu.:9.984e+08   1st Qu.:1.260e+09   1st Qu.:1.408e+09
##   Median :9.911e+09   Median :1.242e+10   Median :1.481e+10
##   Mean   :3.125e+11   Mean   :3.676e+11   Mean   :4.276e+11
##   3rd Qu.:1.001e+11   3rd Qu.:1.147e+11   3rd Qu.:1.346e+11
##   Max.   :7.288e+12   Max.   :8.597e+12   Max.   :9.986e+12
##   NA's   :89          NA's   :90          NA's   :89
##      X1980               X1981               X1982
##   Min.   :3.872e+07   Min.   :3.102e+07   Min.   :3.492e+07
##   1st Qu.:1.378e+09   1st Qu.:1.390e+09   1st Qu.:1.389e+09
##   Median :1.333e+10   Median :1.429e+10   Median :1.431e+10
##   Mean   :4.595e+11   Mean   :4.688e+11   Mean   :4.623e+11
##   3rd Qu.:1.409e+11   3rd Qu.:1.642e+11   3rd Qu.:1.533e+11
##   Max.   :1.124e+13   Max.   :1.164e+13   Max.   :1.153e+13
##   NA's   :79          NA's   :76          NA's   :76
```

```
##      X1983                X1984                X1985
## Min.   :3.784e+07   Min.   :4.125e+07   Min.   :3.213e+07
## 1st Qu.:1.336e+09   1st Qu.:1.344e+09   1st Qu.:1.491e+09
## Median :1.365e+10   Median :1.035e+10   Median :1.120e+10
## Mean   :4.625e+11   Mean   :4.679e+11   Mean   :4.911e+11
## 3rd Qu.:1.259e+11   3rd Qu.:1.141e+11   3rd Qu.:1.168e+11
## Max.   :1.176e+13   Max.   :1.220e+13   Max.   :1.281e+13
## NA's   :75          NA's   :72          NA's   :72
##      X1986                X1987                X1988
## Min.   :3.209e+07   Min.   :3.361e+07   Min.   :4.297e+07
## 1st Qu.:1.784e+09   1st Qu.:1.984e+09   1st Qu.:2.253e+09
## Median :1.023e+10   Median :1.094e+10   Median :1.170e+10
## Mean   :5.694e+11   Mean   :6.326e+11   Mean   :6.962e+11
## 3rd Qu.:1.332e+11   3rd Qu.:1.444e+11   3rd Qu.:1.545e+11
## Max.   :1.514e+13   Max.   :1.723e+13   Max.   :1.927e+13
## NA's   :70          NA's   :66          NA's   :60
##      X1989                X1990                X1991
## Min.   :4.112e+07   Min.   :8.824e+06   Min.   :9.365e+06
## 1st Qu.:2.240e+09   1st Qu.:2.578e+09   1st Qu.:2.689e+09
## Median :1.077e+10   Median :1.230e+10   Median :1.138e+10
## Mean   :7.244e+11   Mean   :7.705e+11   Mean   :8.106e+11
## 3rd Qu.:1.692e+11   3rd Qu.:1.653e+11   3rd Qu.:1.683e+11
## Max.   :2.011e+13   Max.   :2.266e+13   Max.   :2.398e+13
## NA's   :60          NA's   :46          NA's   :45
##      X1992                X1993                X1994
## Min.   :9.743e+06   Min.   :9.631e+06   Min.   :1.089e+07
## 1st Qu.:2.541e+09   1st Qu.:2.628e+09   1st Qu.:2.526e+09
## Median :1.140e+10   Median :1.310e+10   Median :1.291e+10
## Mean   :8.550e+11   Mean   :8.530e+11   Mean   :9.026e+11
## 3rd Qu.:1.673e+11   3rd Qu.:1.833e+11   3rd Qu.:1.680e+11
## Max.   :2.546e+13   Max.   :2.587e+13   Max.   :2.778e+13
## NA's   :45          NA's   :41          NA's   :38
##      X1995                X1996                X1997
## Min.   :1.103e+07   Min.   :1.233e+07   Min.   :1.270e+07
## 1st Qu.:2.733e+09   1st Qu.:3.106e+09   1st Qu.:3.303e+09
## Median :1.390e+10   Median :1.435e+10   Median :1.542e+10
## Mean   :9.743e+11   Mean   :1.004e+12   Mean   :1.003e+12
## 3rd Qu.:1.822e+11   3rd Qu.:1.907e+11   3rd Qu.:1.991e+11
## Max.   :3.087e+13   Max.   :3.155e+13   Max.   :3.144e+13
## NA's   :30          NA's   :30          NA's   :30
##      X1998                X1999                X2000
## Min.   :1.276e+07   Min.   :1.369e+07   Min.   :1.374e+07
## 1st Qu.:3.128e+09   1st Qu.:3.212e+09   1st Qu.:2.849e+09
## Median :1.521e+10   Median :1.571e+10   Median :1.357e+10
## Mean   :9.941e+11   Mean   :1.015e+12   Mean   :1.017e+12
## 3rd Qu.:1.915e+11   3rd Qu.:1.878e+11   3rd Qu.:1.977e+11
## Max.   :3.138e+13   Max.   :3.254e+13   Max.   :3.359e+13
## NA's   :28          NA's   :27          NA's   :20
##      X2001                X2002                X2003
## Min.   :1.320e+07   Min.   :1.545e+07   Min.   :1.823e+07
## 1st Qu.:2.828e+09   1st Qu.:3.000e+09   1st Qu.:3.275e+09
## Median :1.491e+10   Median :1.459e+10   Median :1.765e+10
## Mean   :1.016e+12   Mean   :1.035e+12   Mean   :1.169e+12
## 3rd Qu.:1.981e+11   3rd Qu.:1.955e+11   3rd Qu.:2.202e+11
```

```
##   Max.    :3.340e+13   Max.    :3.467e+13   Max.    :3.890e+13
##   NA's    :20          NA's    :15          NA's    :15
##       X2004               X2005               X2006
##   Min.    :2.153e+07   Min.    :2.184e+07   Min.    :2.290e+07
##   1st Qu.:3.814e+09   1st Qu.:4.375e+09   1st Qu.:4.670e+09
##   Median :2.090e+10   Median :2.149e+10   Median :2.583e+10
##   Mean    :1.322e+12   Mean    :1.442e+12   Mean    :1.584e+12
##   3rd Qu.:2.566e+11   3rd Qu.:3.075e+11   3rd Qu.:3.452e+11
##   Max.    :4.382e+13   Max.    :4.746e+13   Max.    :5.145e+13
##   NA's    :14          NA's    :13          NA's    :13
##       X2007               X2008               X2009
##   Min.    :2.044e+07   Min.    :3.029e+07   Min.    :2.710e+07
##   1st Qu.:5.452e+09   1st Qu.:6.082e+09   1st Qu.:5.820e+09
##   Median :3.215e+10   Median :3.590e+10   Median :3.702e+10
##   Mean    :1.809e+12   Mean    :2.026e+12   Mean    :1.922e+12
##   3rd Qu.:4.047e+11   3rd Qu.:4.862e+11   3rd Qu.:4.304e+11
##   Max.    :5.797e+13   Max.    :6.361e+13   Max.    :6.033e+13
##   NA's    :12          NA's    :13          NA's    :13
##       X2010               X2011               X2012
##   Min.    :3.182e+07   Min.    :3.871e+07   Min.    :3.767e+07
##   1st Qu.:6.907e+09   1st Qu.:7.603e+09   1st Qu.:7.997e+09
##   Median :4.134e+10   Median :4.461e+10   Median :4.658e+10
##   Mean    :2.130e+12   Mean    :2.384e+12   Mean    :2.460e+12
##   3rd Qu.:4.782e+11   3rd Qu.:5.292e+11   3rd Qu.:5.485e+11
##   Max.    :6.605e+13   Max.    :7.339e+13   Max.    :7.509e+13
##   NA's    :13          NA's    :12          NA's    :13
##       X2013               X2014               X2015
##   Min.    :3.751e+07   Min.    :3.729e+07   Min.    :3.549e+07
##   1st Qu.:8.402e+09   1st Qu.:9.029e+09   1st Qu.:8.290e+09
##   Median :4.975e+10   Median :5.258e+10   Median :5.164e+10
##   Mean    :2.531e+12   Mean    :2.614e+12   Mean    :2.462e+12
##   3rd Qu.:5.520e+11   3rd Qu.:5.512e+11   3rd Qu.:5.078e+11
##   Max.    :7.724e+13   Max.    :7.933e+13   Max.    :7.505e+13
##   NA's    :11          NA's    :12          NA's    :12
##       X2016               X2017               X2018
##   Min.    :3.655e+07   Min.    :4.062e+07   Min.    :4.259e+07
##   1st Qu.:8.173e+09   1st Qu.:9.470e+09   1st Qu.:1.140e+10
##   Median :5.142e+10   Median :5.649e+10   Median :5.990e+10
##   Mean    :2.508e+12   Mean    :2.715e+12   Mean    :2.953e+12
##   3rd Qu.:5.140e+11   3rd Qu.:5.927e+11   3rd Qu.:5.863e+11
##   Max.    :7.616e+13   Max.    :8.095e+13   Max.    :8.591e+13
##   NA's    :14          NA's    :17          NA's    :24
##       X2019               X
##   Min.    :4.019e+08   Mode:logical
##   1st Qu.:1.155e+10   NA's:264
##   Median :3.707e+10
##   Mean    :2.388e+11
##   3rd Qu.:2.019e+11
##   Max.    :1.619e+12
##   NA's    :254
```

```
summary (LE)
```

```
##               Country          Year              Status    Life.expectancy
```

```
##  Afghanistan         :  16    Min.    :2000    Developed : 512    Min.    :36.30
##  Albania             :  16    1st Qu.:2004    Developing:2426    1st Qu.:63.10
##  Algeria             :  16    Median :2008                        Median :72.10
##  Angola              :  16    Mean    :2008                        Mean    :69.22
##  Antigua and Barbuda:  16    3rd Qu.:2012                        3rd Qu.:75.70
##  Argentina           :  16    Max.    :2015                        Max.    :89.00
##  (Other)             :2842                                         NA's    :10
##  Adult.Mortality infant.deaths      Alcohol       percentage.expenditure
##  Min.    :  1.0    Min.    :    0.0    Min.    : 0.0100    Min.    :     0.000
##  1st Qu.: 74.0    1st Qu.:    0.0    1st Qu.: 0.8775    1st Qu.:     4.685
##  Median :144.0    Median :    3.0    Median : 3.7550    Median :    64.913
##  Mean    :164.8    Mean    :   30.3    Mean    : 4.6029    Mean    :   738.251
##  3rd Qu.:228.0    3rd Qu.:   22.0    3rd Qu.: 7.7025    3rd Qu.:   441.534
##  Max.    :723.0    Max.    :1800.0    Max.    :17.8700    Max.    :19479.912
##  NA's    :10                          NA's    :194
##   Hepatitis.B        Measles              BMI         under.five.deaths
##  Min.    : 1.00    Min.    :     0.0    Min.    : 1.00    Min.    :    0.00
##  1st Qu.:77.00    1st Qu.:     0.0    1st Qu.:19.30    1st Qu.:    0.00
##  Median :92.00    Median :    17.0    Median :43.50    Median :    4.00
##  Mean    :80.94    Mean    :  2419.6    Mean    :38.32    Mean    :   42.04
##  3rd Qu.:97.00    3rd Qu.:   360.2    3rd Qu.:56.20    3rd Qu.:   28.00
##  Max.    :99.00    Max.    :212183.0    Max.    :87.30    Max.    :2500.00
##  NA's    :553                          NA's    :34
##      Polio        Total.expenditure  Diphtheria        HIV.AIDS
##  Min.    : 3.00    Min.    : 0.370    Min.    : 2.00    Min.    : 0.100
##  1st Qu.:78.00    1st Qu.: 4.260    1st Qu.:78.00    1st Qu.: 0.100
##  Median :93.00    Median : 5.755    Median :93.00    Median : 0.100
##  Mean    :82.55    Mean    : 5.938    Mean    :82.32    Mean    : 1.742
##  3rd Qu.:97.00    3rd Qu.: 7.492    3rd Qu.:97.00    3rd Qu.: 0.800
##  Max.    :99.00    Max.    :17.600    Max.    :99.00    Max.    :50.600
##  NA's    :19      NA's    :226      NA's    :19
##      GDP            Population       thinness..1.19.years
##  Min.    :      1.68    Min.    :3.400e+01    Min.    : 0.10
##  1st Qu.:    463.94    1st Qu.:1.958e+05    1st Qu.: 1.60
##  Median :   1766.95    Median :1.387e+06    Median : 3.30
##  Mean    :   7483.16    Mean    :1.275e+07    Mean    : 4.84
##  3rd Qu.:   5910.81    3rd Qu.:7.420e+06    3rd Qu.: 7.20
##  Max.    :119172.74    Max.    :1.294e+09    Max.    :27.70
##  NA's    :448      NA's    :652      NA's    :34
##  thinness.5.9.years Income.composition.of.resources    Schooling
##  Min.    : 0.10    Min.    :0.0000                  Min.    : 0.00
##  1st Qu.: 1.50    1st Qu.:0.4930                  1st Qu.:10.10
##  Median : 3.30    Median :0.6770                  Median :12.30
##  Mean    : 4.87    Mean    :0.6276                  Mean    :11.99
##  3rd Qu.: 7.20    3rd Qu.:0.7790                  3rd Qu.:14.30
##  Max.    :28.60    Max.    :0.9480                  Max.    :20.70
##  NA's    :34      NA's    :167                  NA's    :163
```

Reduce the Datasets to the 6 Countries which are analysed and merge them

```r
# Reduce the Life Expectancy Dataset to the 6 Countries which are analysed
redLE2<-subset(LE, Country %in% c("New Zealand", "Switzerland", "Poland", "Portugal", "Belgium", "Austr
```

```r
redLE<-subset(redLE2, Year %in% 2000:2014)
#summary(redLE)
#Reduce the GDP Data to the same 6 Countries as above and remove Non-binary Values except the Countryna
redGDP2<-subset(GDP[,c(1,45:59)], ï..Country.Name %in% c("New Zealand", "Switzerland", "Poland", "Portug
#Rename the columns to the same Structure as the Life Expectancy dataset
colnames(redGDP2)<-c("Country",2000:2014)
#Change Number Format
options(scipen = 999) #(Reset mit scipen =0)
#Change the format of the GDP dataset to the same structure as the Life Expectancy dataset.
redGDP<-gather(redGDP2, key="Year", value="BIP", "2000", "2001","2002","2003","2004","2005","2006","2007
#merge the datasets to the final dataset
redDB<-merge(redLE, redGDP, by = c("Country", "Year"))
```

```r
## Finding the correlations in the full dataset in the Year 2010
LE10<-subset(LE, Year %in% 2010)
corfullDB<-LE10[,c(2,4:22)]
corfull<-cor(corfullDB)
```

```
## Warning in cor(corfullDB): Standardabweichung ist Null
```

```r
corfulllist <- as.data.frame(as.table(corfull))
subset(subset(corfulllist, abs(Freq) > 0.3),abs(Freq)<1)
```

```
##                         Var1                   Var2       Freq
## 23           Adult.Mortality        Life.expectancy -0.7809818
## 26    percentage.expenditure        Life.expectancy  0.4056123
## 34                  HIV.AIDS        Life.expectancy -0.5887995
## 42           Life.expectancy        Adult.Mortality -0.7809818
## 54                  HIV.AIDS        Adult.Mortality  0.6089740
## 70          under.five.deaths          infant.deaths  0.9970049
## 102          Life.expectancy percentage.expenditure  0.4056123
## 154                 HIV.AIDS                Measles  0.3359007
## 184            infant.deaths      under.five.deaths  0.9970049
## 262          Life.expectancy               HIV.AIDS -0.5887995
## 263          Adult.Mortality               HIV.AIDS  0.6089740
## 268                  Measles               HIV.AIDS  0.3359007
```

```r
#Finding the correlations in the reduced dataset
corDB<-redDB[,c(2,4:23)]
corDBB<-cor(corDB)
```

```
## Warning in cor(corDB): Standardabweichung ist Null
```

```r
corlist <- as.data.frame(as.table(corDBB))
subset(subset(corlist, abs(Freq) > 0.5),abs(Freq)<1)
```

```
##                         Var1                   Var2       Freq
## 2            Life.expectancy                   Year  0.5391805
## 21                       BIP                   Year  0.5574446
## 22                      Year        Life.expectancy  0.5391805
## 25             infant.deaths        Life.expectancy -0.6645879
```

```
## 31              under.five.deaths              Life.expectancy -0.6504706
## 40  Income.composition.of.resources            Life.expectancy  0.6261555
## 65              Life.expectancy                  infant.deaths  -0.6645879
## 73              under.five.deaths                infant.deaths   0.9140977
## 80              thinness..1.19.years             infant.deaths   0.7179805
## 81              thinness.5.9.years               infant.deaths   0.6890031
## 82  Income.composition.of.resources             infant.deaths  -0.5599799
## 120                         GDP       percentage.expenditure    0.9097840
## 124 Income.composition.of.resources  percentage.expenditure    0.5802680
## 191              Life.expectancy              under.five.deaths -0.6504706
## 193              infant.deaths               under.five.deaths  0.9140977
## 206              thinness..1.19.years        under.five.deaths  0.6925761
## 207              thinness.5.9.years          under.five.deaths  0.6651073
## 208 Income.composition.of.resources     under.five.deaths -0.5709618
## 223                   Diphtheria                       Polio    0.6046082
## 263                        Polio                  Diphtheria    0.6046082
## 300       percentage.expenditure                        GDP    0.9097840
## 313 Income.composition.of.resources                       GDP    0.5841879
## 315                          BIP                        GDP    0.5308397
## 340              infant.deaths       thinness..1.19.years    0.7179805
## 346              under.five.deaths   thinness..1.19.years    0.6925761
## 354              thinness.5.9.years  thinness..1.19.years    0.9927666
## 355 Income.composition.of.resources  thinness..1.19.years -0.5304021
## 356                     Schooling   thinness..1.19.years -0.5707932
## 361              infant.deaths         thinness.5.9.years    0.6890031
## 367              under.five.deaths     thinness.5.9.years    0.6651073
## 374              thinness..1.19.years  thinness.5.9.years    0.9927666
## 377                     Schooling     thinness.5.9.years -0.5134080
## 380              Life.expectancy Income.composition.of.resources  0.6261555
## 382              infant.deaths Income.composition.of.resources -0.5599799
## 384       percentage.expenditure Income.composition.of.resources  0.5802680
## 388              under.five.deaths Income.composition.of.resources -0.5709618
## 393                          GDP Income.composition.of.resources  0.5841879
## 395              thinness..1.19.years Income.composition.of.resources -0.5304021
## 416              thinness..1.19.years            Schooling -0.5707932
## 417              thinness.5.9.years               Schooling -0.5134080
## 421                         Year                        BIP    0.5574446
## 435                          GDP                        BIP    0.5308397
```
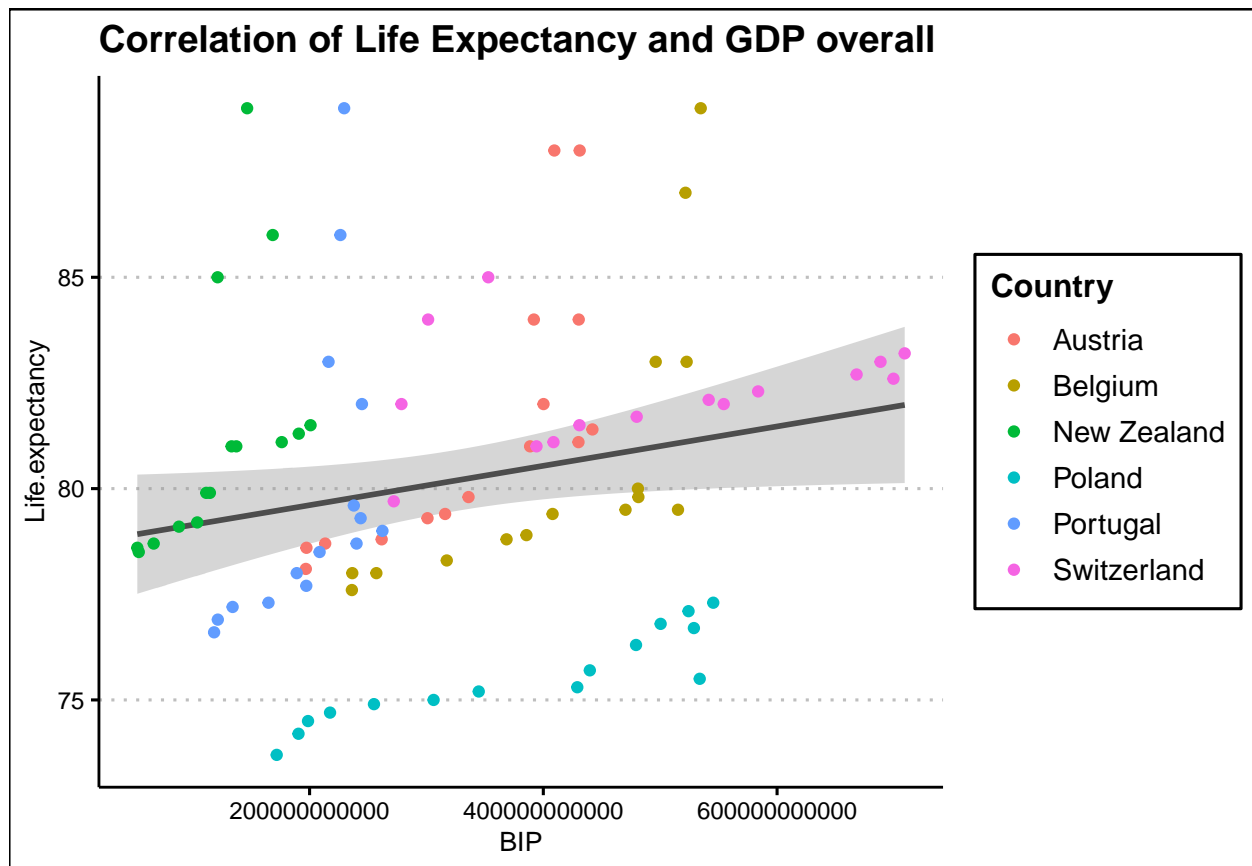
```
ggplot(data = redDB,
   mapping= aes(x=reorder(Country, Life.expectancy), y=Life.expectancy, fill=Country)) +
   geom_boxplot() +
   xlab("Country") +
   theme(legend.position="none") +
   xlab("")+
    ggtitle("Life expectancy in the examined countrys")+
   theme_clean()
```
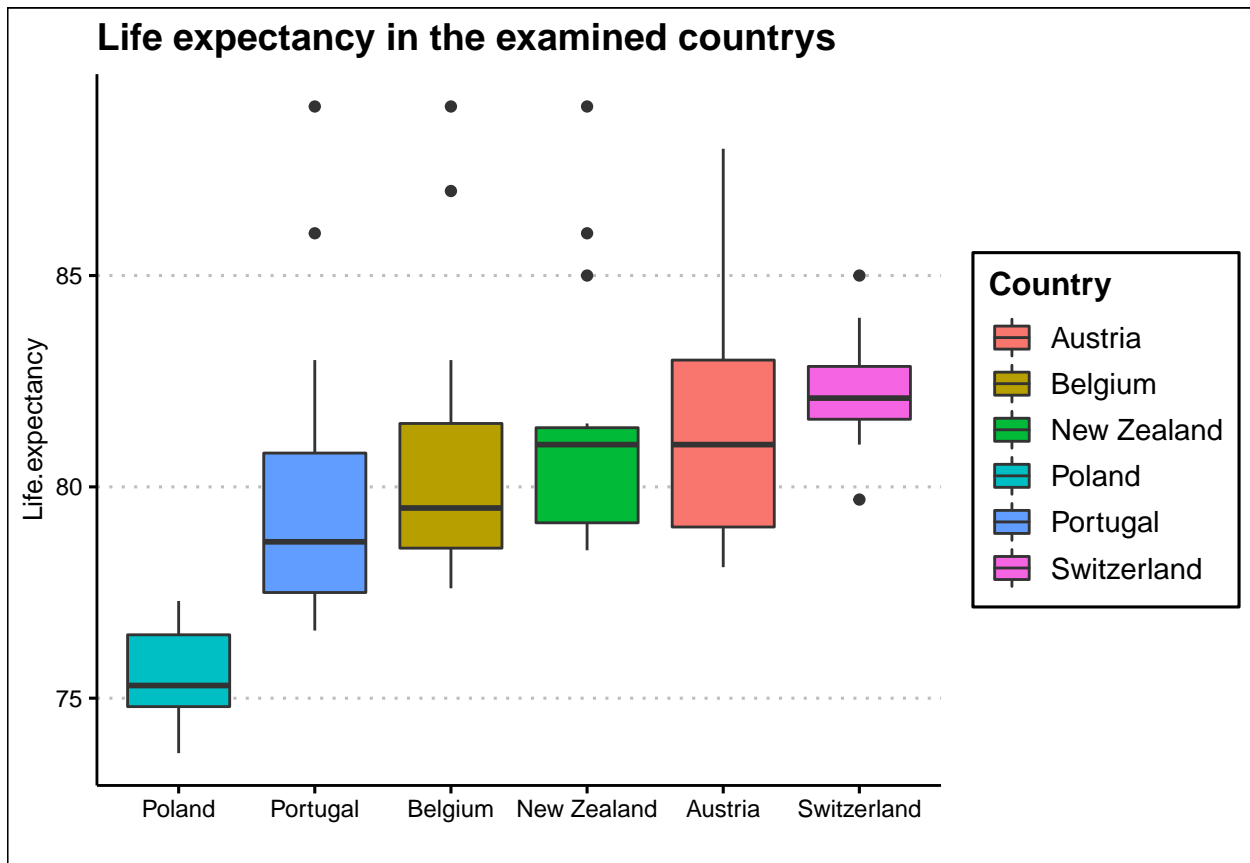


```
#Life Expectancy and BIP (GDP)
ggplot(data = redDB,
      mapping = aes(y = Life.expectancy,
                    x = BIP,
                    color= Country)) +

  geom_smooth(method = "lm", se=T, color="gray30") +
  geom_point()+
  ggtitle("Correlation of Life Expectancy and GDP overall")+
  theme_clean()
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Correlation of Life Expectancy and GDP overall

[1] "This graph shows that there seems to be a correlation between the Life Expectancy and the\nGDP a

```r
#Life Expectancy and BIP (GDP) for each country
ggplot(data = redDB,
       mapping = aes(y = Life.expectancy,
                     x = BIP
       )) +

  geom_smooth(method = "lm", se=T, color="grey30") +
  geom_point(color="darkblue")+
  facet_wrap(. ~ Country) +
  ggtitle("Correlation of Life Expectancy and GDP for each country")+
  theme_clean()
```

`geom_smooth()` using formula 'y ~ x'

Correlation of Life Expectancy and GDP for each country
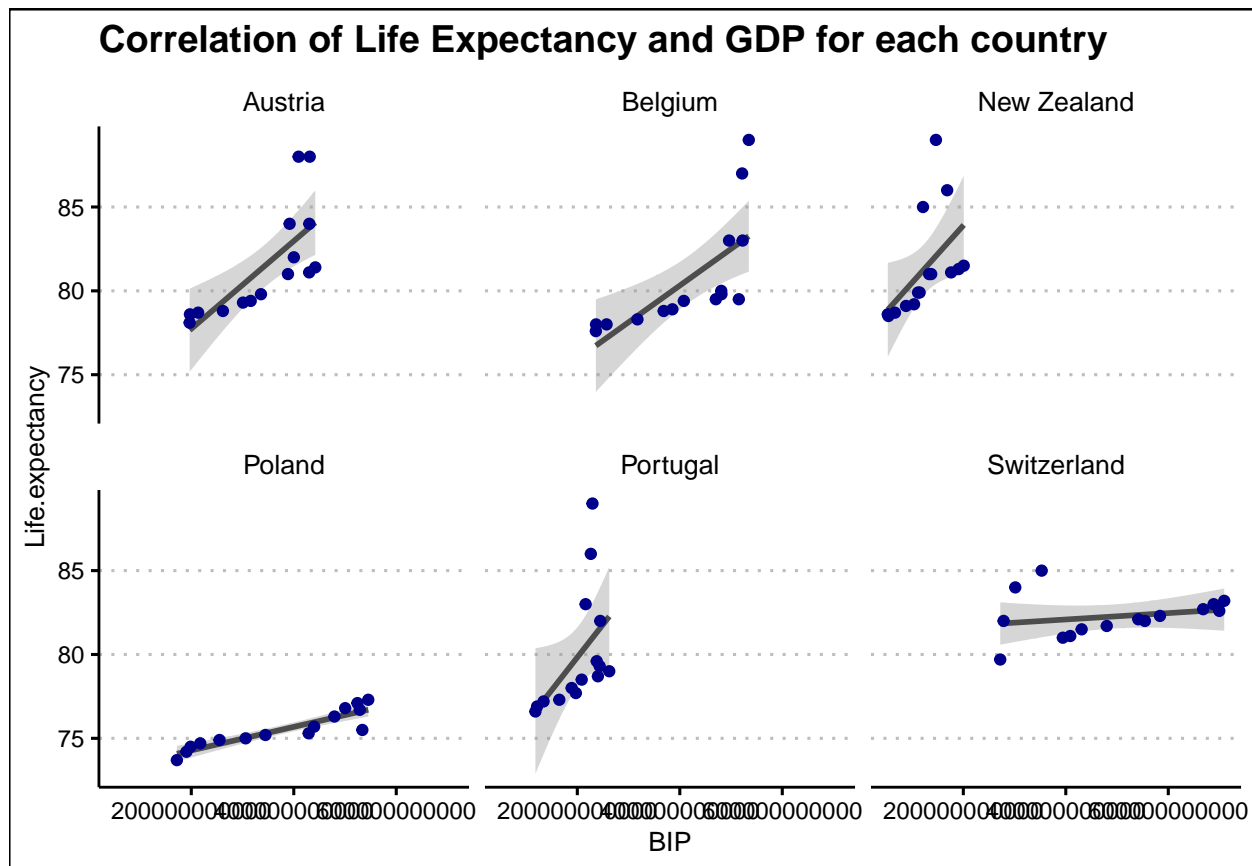
"These graph shows that there seems to be a correlation between the Life Expectancy and the
GDP as the line of the linear model is slightly rising.
The spread in the Data points is big, but each country has a higher Life Expectrancy at a
higher GDP itself "

[1] "These graph shows that there seems to be a correlation between the Life Expectancy and the \nGDI

```
#Life Expectancy and Alcohol consumption
ggplot(data = redDB,
       mapping = aes(y = Life.expectancy,
                     x = Alcohol,
                     color=Country
       )) +

  geom_smooth(method = "lm",formula = y~log(x),  se=T, color="grey30") +
  geom_point()+
  ggtitle("Correlation of Life Expectancy and Alcohol overall")+
  theme_clean()
```

Correlation of Life Expectancy and Alcohol overall
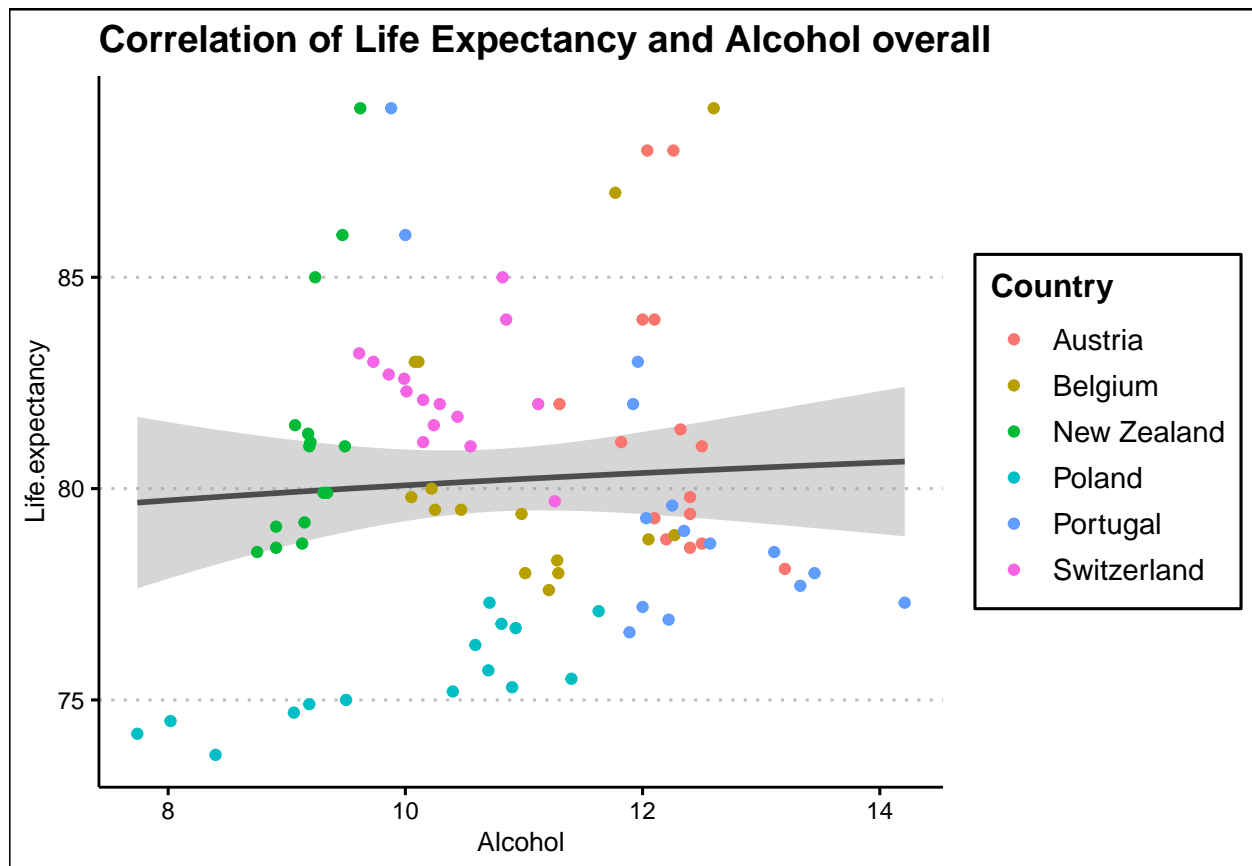
```
"The line of the linear model is almost horizontal, so there doesnt seem to be a real
correlation between the Life expectancy and the alcolhol conumption overall"
```

[1] "The line of the linear model is almost horizontal, so there doesnt seem to be a real \ncorrelati

```r
#Life Expectancy and Alcohol consumption for each country
ggplot(data = redDB,
       mapping = aes(y = Life.expectancy,
                     x = Alcohol
       )) +

  geom_smooth(method = "lm",formula = y~log(x),  se=T, color="grey30") +
  geom_point(color="darkblue")+
  theme_clean() +
  facet_wrap(. ~ Country)+
ggtitle("Correlation of Life Expectancy and GDP in each country")
```

Correlation of Life Expectancy and GDP in each country
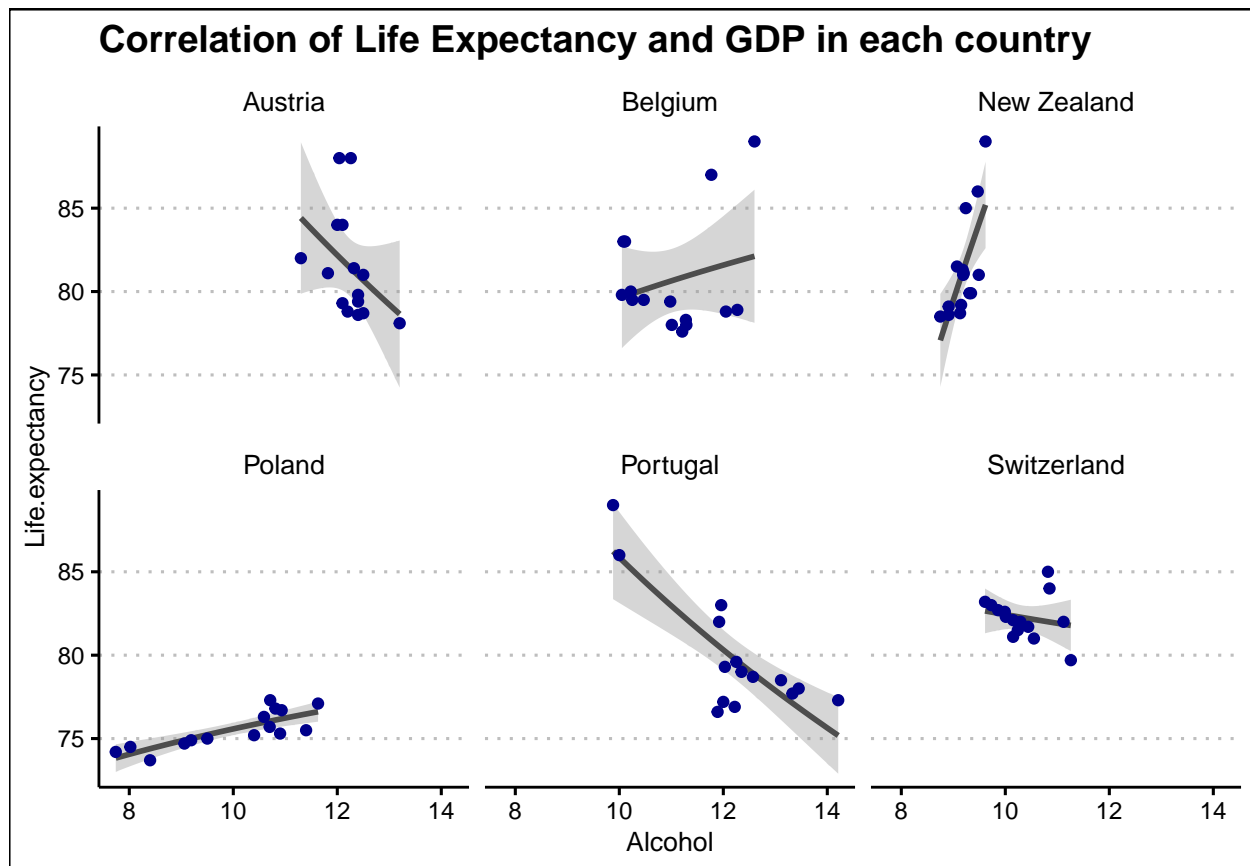
"This Plot is interesting because in half of the countrys (Poland, Belgium and New Zealand),
the Life expectancy is rises with the increase of the alcohol consumption, in the other half
of the countrys, the Life Expectancy falls with higher alcohol consumption"

## [1] "This Plot is interesting because in half of the countrys (Poland, Belgium and New Zealand),\nth

```r
# BIP and Alcohol consumption
ggplot(data = redDB,
       mapping = aes(y = BIP,
                     x = Alcohol,
                     color=Country
       )) +

  geom_smooth(method = "lm",formula = y~log(x),  se=T, color="grey30") +
  geom_point()+
  ggtitle("Correlation of GDP and Alcohol consumption overall")+
  theme_clean()
```

Correlation of GDP and Alcohol consumption overall
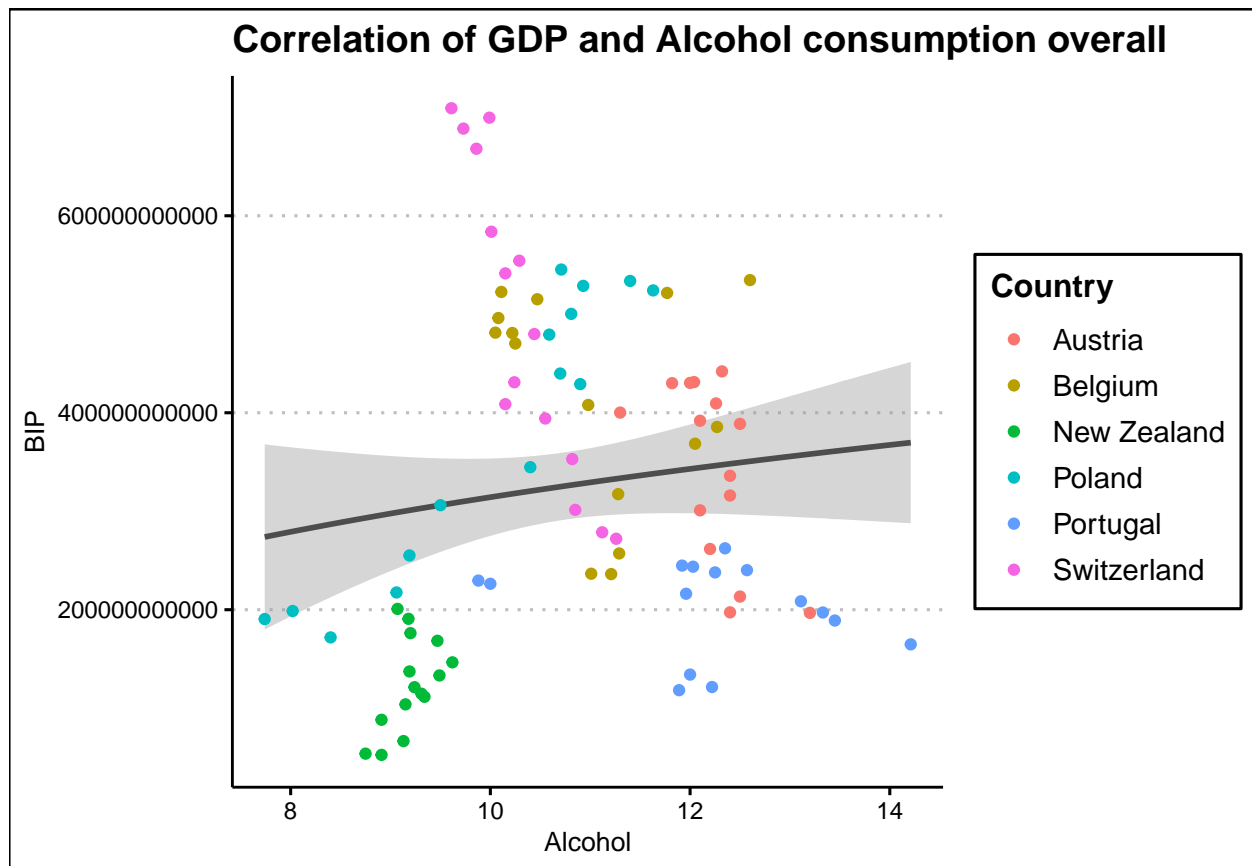
```
#BIP and Alcohol consumption for each country
ggplot(data = redDB,
       mapping = aes(y = BIP,
                     x = Alcohol
       )) +

  geom_smooth(method = "lm",formula = y~log(x),  se=T, color="grey30") +
  geom_point(color="darkblue")+
  theme_clean() +
  facet_wrap(. ~ Country)+
ggtitle("Correlation of GDP and Alcohol consumption in each country")
```

Correlation of GDP and Alcohol consumption in each cou
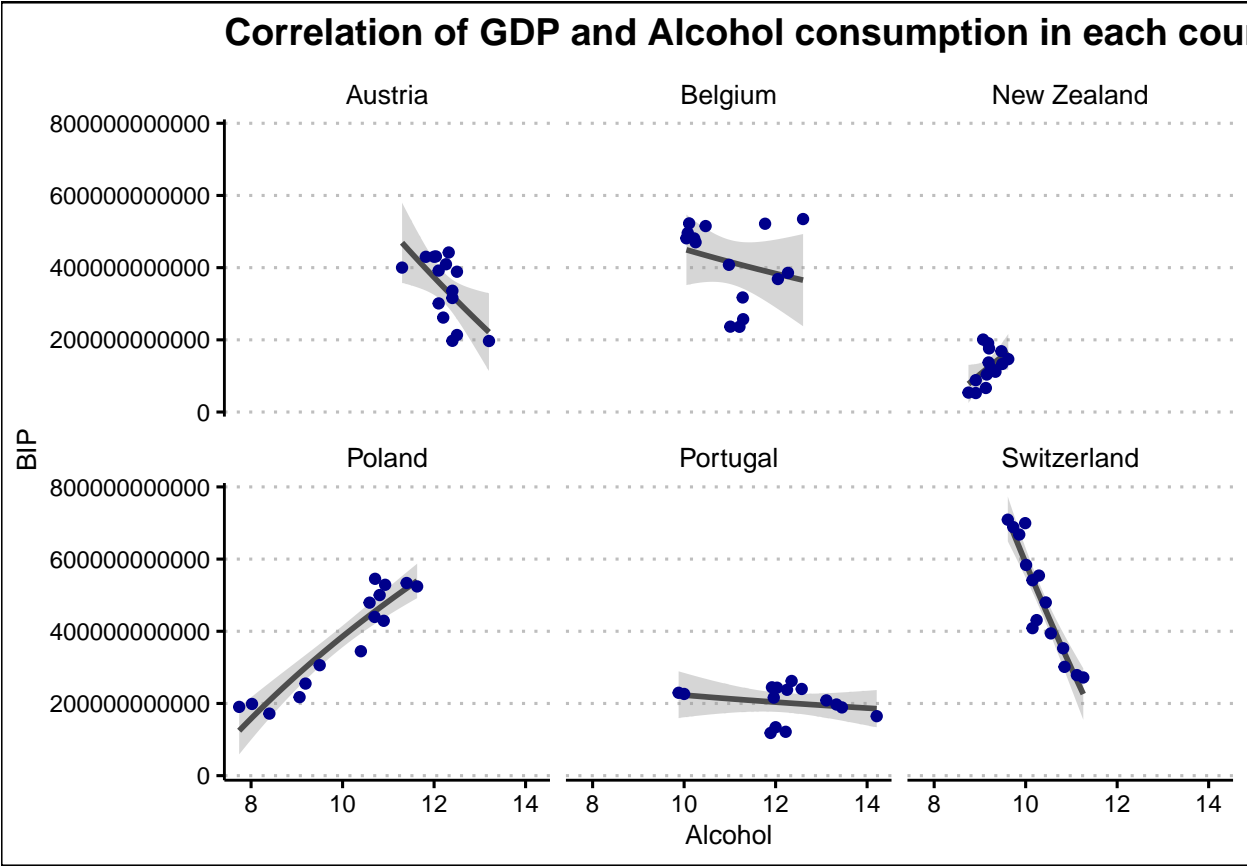
"Like in the Graph above, every countrys shows a different behaviour of the GDP
with raising alcohol consumption "

## [1] "Like in the Graph above, every countrys shows a different behaviour of the GDP \nwith raising al

###

```r
# Linear Model of the GDP and the Life Expectancy

linmod<-lm(redDB$Life.expectancy~redDB$BIP)
summary(linmod)
```

```
##
## Call:
## lm(formula = redDB$Life.expectancy ~ redDB$BIP)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7772 -1.8357 -0.3043  1.5670  9.6407
##
## Coefficients:
##                      Estimate         Std. Error t value           Pr(>|t|)
## (Intercept) 78.676608151128178  0.812565665810661  96.825 <0.0000000000000002
## redDB$BIP    0.000000000004658  0.000000000002232   2.087             0.0398
##
## (Intercept) ***
## redDB$BIP   *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.446 on 88 degrees of freedom
## Multiple R-squared:  0.04715,    Adjusted R-squared:  0.03632
## F-statistic: 4.354 on 1 and 88 DF,  p-value: 0.03981
```

```r
ggplot(data = redDB,
       mapping = aes(y = Life.expectancy,
                     x = BIP,
                     color=Country
       )) +

  geom_abline(intercept = 78.68, slope = 4.658e-12) +
  geom_point()+
   ggtitle("Linear model of live expectancy and GDP")+
  theme_clean()
```

Linear model of live expectancy and GDP
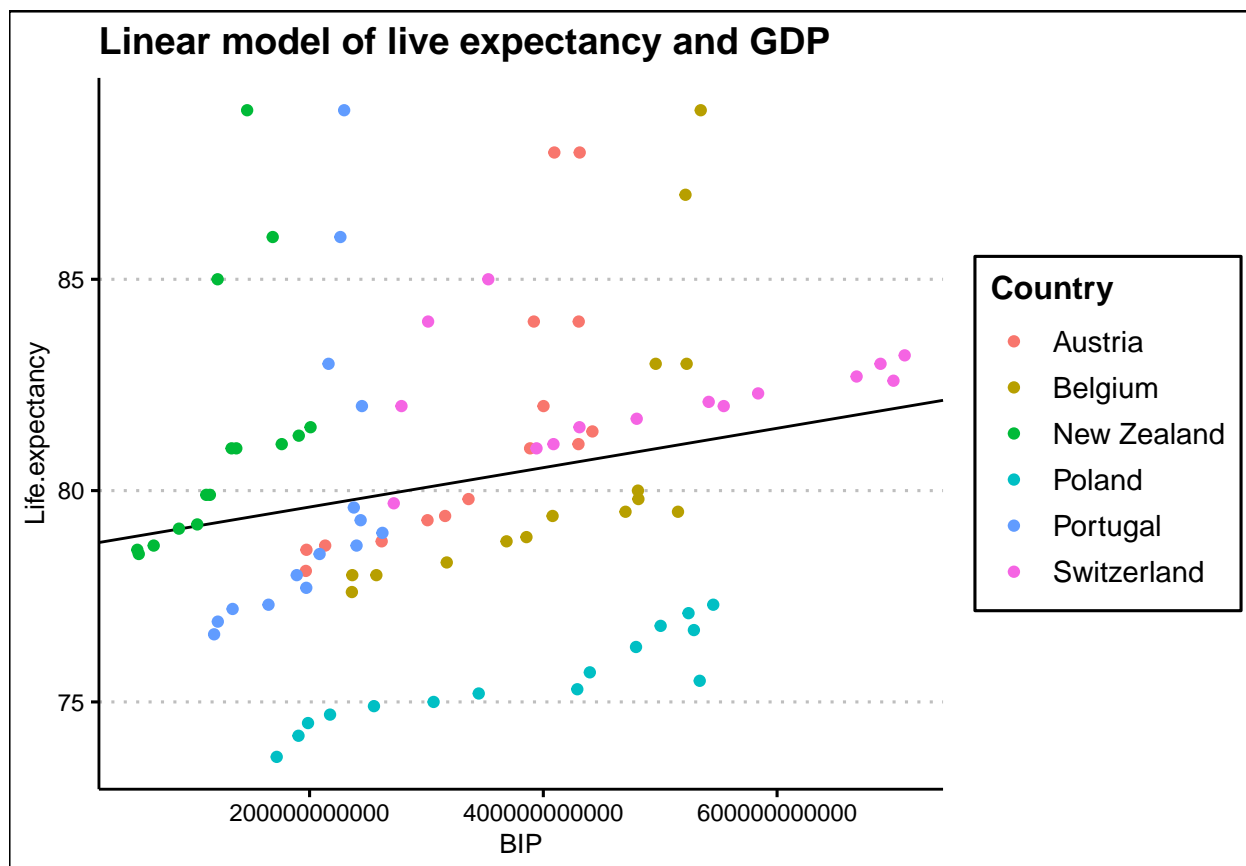
```
#Complex (logarithmic) Model of the GDP and the Life Expectancy

sqmod<-lm(log(redDB$Life.expectancy)~redDB$BIP)
summary(sqmod)
```

```
## 
## Call:
## lm(formula = log(redDB$Life.expectancy) ~ redDB$BIP)
## 
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.074589 -0.022268 -0.003075  0.020354  0.115512
## 
## Coefficients:
##                      Estimate        Std. Error t value          Pr(>|t|)
## (Intercept) 4.36462558541682988 0.01002018267856678 435.583 <0.0000000000000002
## redDB$BIP   0.00000000000005798 0.00000000000002752   2.107             0.038
## 
## (Intercept) ***
## redDB$BIP   *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0425 on 88 degrees of freedom
## Multiple R-squared:  0.048,  Adjusted R-squared:  0.03719
## F-statistic: 4.437 on 1 and 88 DF,  p-value: 0.03801
```

```r
#Plot of the Complex model
ggplot(data = redDB,
       mapping = aes(y = Life.expectancy,
                     x = BIP)) +


  geom_smooth(method = "glm", formula = y~x,
              method.args = list(family = gaussian(link = 'log')),color="grey30") +
  geom_point(color="darkblue")+
  theme_clean()+
  ggtitle("Logarithmic model of live expectancy and GDP")+
  scale_y_continuous(trans = 'log10', limits = c(70,90))
```
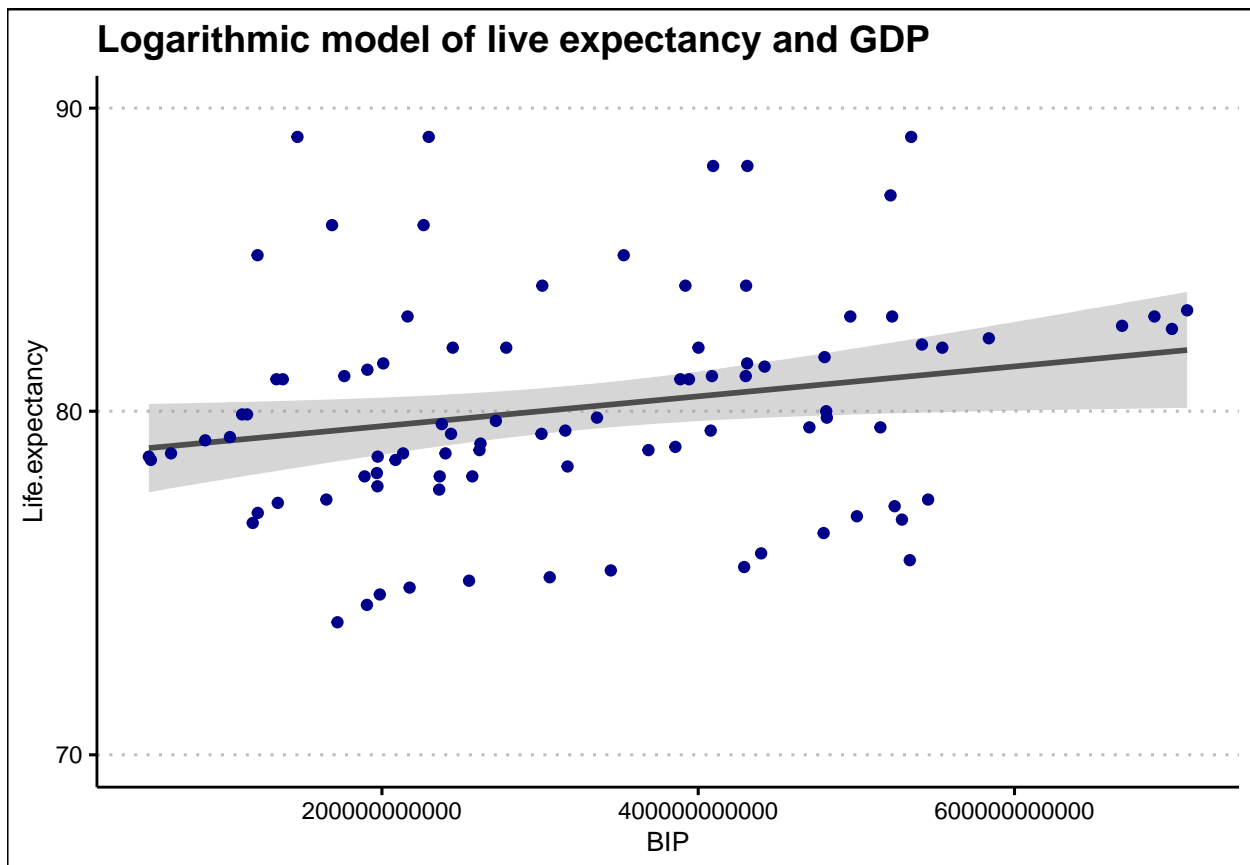


```
"The line of the complex model is almost straight, which shows that the spread of
the datapoints is too big for a representative model"
```

```
## [1] "The line of the complex model is almost straight, which shows that the spread of \nthe datapoin
```

```r
#Plot of the Complex model for each country
ggplot(data = redDB,
       mapping = aes(y = Life.expectancy,
                     x = BIP,
                     color=Country)) +
```

```
geom_smooth(method = "glm", formula = y~x,
                       method.args = list(family = gaussian(link = 'log'))) +
geom_point(color="darkblue")+
theme_clean()   +
ggtitle("Logarithmic model of live expectancy and GDP for each country")+
scale_y_continuous(trans = 'log10', limits = c(70,90))
```
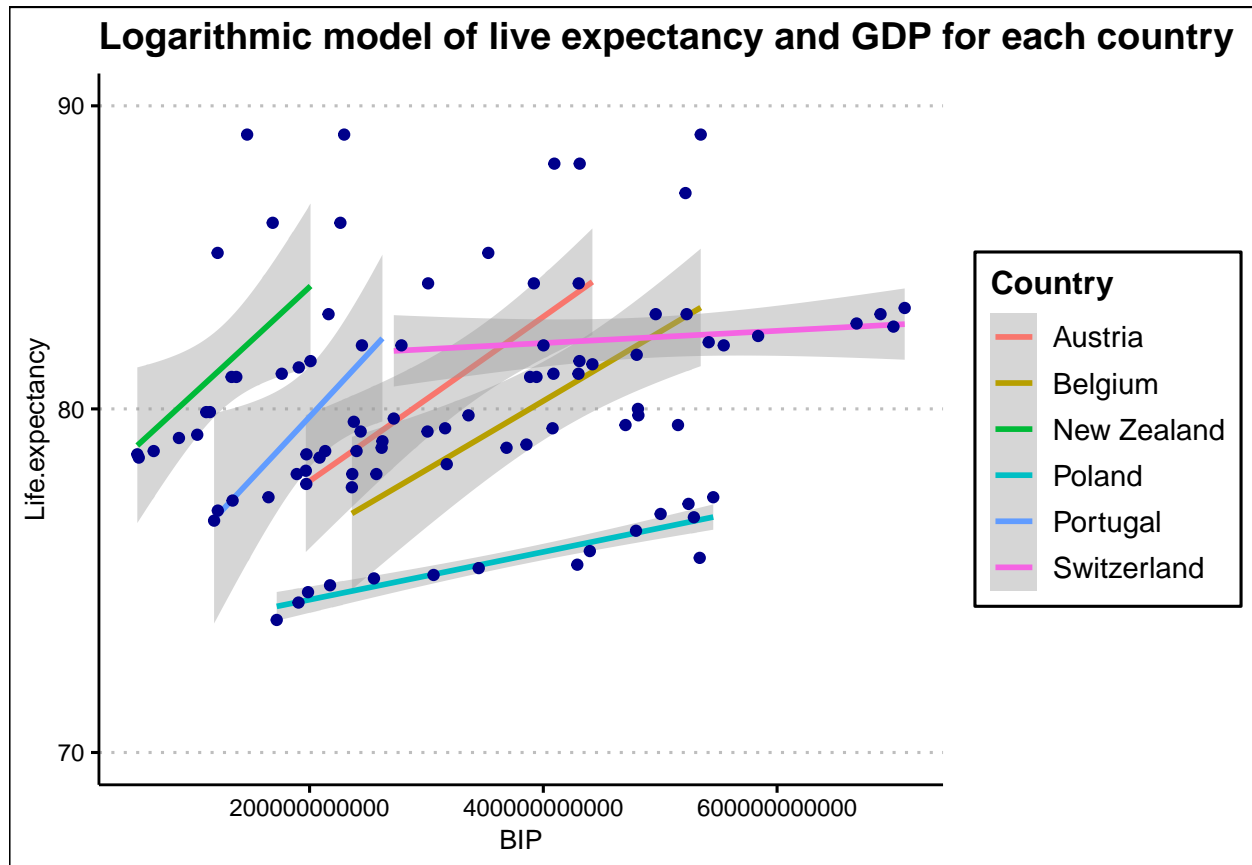


```
"All models seem to be almosts straight, so the spread in data is too
big for a different fitting model "
```

```
## [1] "All models seem to be almosts straight, so the spread in data is too \nbig for a different fitt:
```