

# 2353 Final Project Life Exp

Yi Yang

2023-04-14

```
knitr::opts_chunk$set(echo = TRUE)
library(haven)
library(psych)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr 1.0.1
## v tibble 3.1.8       v dplyr 1.0.10
## v tidyr 1.2.1        v stringr 1.5.0
## v readr 2.1.3        v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x ggplot2::%+%( ) masks psych::%+%( )
## x ggplot2::alpha( ) masks psych::alpha( )
## x dplyr::filter( ) masks stats::filter( )
## x dplyr::lag( ) masks stats::lag( )
```

```
library(ggplot2)
library(lmtest)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##      select
```

```
library(psych)
library(lars)
```

```
## Loaded lars 1.3
##
##
## Attaching package: 'lars'
##
## The following object is masked from 'package:psych':
##
##     error.bars
```

```
library(leaps)
library(glmnet)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
## Loaded glmnet 4.1-6
```

```
library(pander)
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
##
```

```
## The following object is masked from 'package:psych':
##
##      logit
```

```
library(faraway)
```

```
##
## Attaching package: 'faraway'
##
## The following objects are masked from 'package:car':
##
##      logit, vif
##
## The following object is masked from 'package:lattice':
##
##      melanoma
##
## The following object is masked from 'package:psych':
##
##      logit
```

```
library(readr)
library(corrplot)
library(r02pro)
```

## HEAD Data Preparation

```
my_data <- read.csv("Life Expectancy Data.csv")
head(my_data)
```

```
##      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
## 6 Afghanistan 2010 Developing           58.8             279             74
##      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
## 6      0.01              79.679367           66    1989 16.7            102
##      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
## 6     66              9.20           66      0.1 553.32894  2883167
##      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
## 6      18.4      18.4      0.448
##   Schooling
## 1      10.1
## 2      10.0
## 3       9.9
## 4       9.8
## 5       9.5
## 6       9.2
```

```
dim(my_data)
```

```
## [1] 2938  22
```

```
for (i in colnames(my_data)) {
  print(i)
  print(sum(is.na(my_data$i)))
}
```

```
## [1] "Country"
## [1] 0
## [1] "Year"
## [1] 0
## [1] "Status"
## [1] 0
## [1] "Life.expectancy"
## [1] 0
## [1] "Adult.Mortality"
## [1] 0
## [1] "infant.deaths"
## [1] 0
## [1] "Alcohol"
## [1] 0
## [1] "percentage.expenditure"
## [1] 0
## [1] "Hepatitis.B"
## [1] 0
## [1] "Measles"
## [1] 0
## [1] "BMI"
## [1] 0
## [1] "under.five.deaths"
## [1] 0
## [1] "Polio"
## [1] 0
## [1] "Total.expenditure"
## [1] 0
## [1] "Diphtheria"
## [1] 0
```

```
## [1] "HIV.AIDS"
## [1] 0
## [1] "GDP"
## [1] 0
## [1] "Population"
## [1] 0
## [1] "thinness..1.19.years"
## [1] 0
## [1] "thinness.5.9.years"
## [1] 0
## [1] "Income.composition.of.resources"
## [1] 0
## [1] "Schooling"
## [1] 0
```

```
dim(my_data)
```

```
## [1] 2938 22
```

```
summary(my_data)
```

```
##      Country      Year      Status      Life.expectancy
## Length:2938      Min.   :2000      Length:2938      Min.   :36.30
## Class :character  1st Qu.:2004      Class :character  1st Qu.:63.10
## Mode  :character  Median :2008      Mode  :character  Median :72.10
##                               Mean   :2008      Mean   :69.22
##                               3rd Qu.:2012      3rd Qu.:75.70
##                               Max.    :2015      Max.    :89.00
##                               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
```

```
names(my_data)
```

```
## [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"
```

```
my_data1 <- my_data %>%
  na.omit() %>%
  mutate(Developing = as.integer(Status == "Developing")) # Change status to numeric
ncol(my_data1)
```

```
## [1] 23
```

```
my_data1<-my_data1[,-c(1, 2, 3)] # remove country, year, status
ncol(my_data1)
```

```
## [1] 20
```

```
summary(my_data1)
```

```
## Life.expectancy Adult.Mortality infant.deaths      Alcohol
## Min.   :44.0     Min.   : 1.0   Min.   : 0.00   Min.   : 0.010
## 1st Qu.:64.4     1st Qu.: 77.0   1st Qu.: 1.00   1st Qu.: 0.810
## Median :71.7     Median :148.0   Median : 3.00   Median : 3.790
## Mean   :69.3     Mean   :168.2   Mean   : 32.55   Mean   : 4.533
## 3rd Qu.:75.0     3rd Qu.:227.0   3rd Qu.: 22.00   3rd Qu.: 7.340
```

```

## Max. :89.0 Max. :723.0 Max. :1600.00 Max. :17.870
## percentage.expenditure Hepatitis.B Measles BMI
## Min. : 0.00 Min. : 2.00 Min. : 0 Min. : 2.00
## 1st Qu.: 37.44 1st Qu.:74.00 1st Qu.: 0 1st Qu.:19.50
## Median : 145.10 Median :89.00 Median : 15 Median :43.70
## Mean : 698.97 Mean :79.22 Mean : 2224 Mean :38.13
## 3rd Qu.: 509.39 3rd Qu.:96.00 3rd Qu.: 373 3rd Qu.:55.80
## Max. :18961.35 Max. :99.00 Max. :131441 Max. :77.10
## under.five.deaths Polio Total.expenditure Diphtheria
## Min. : 0.00 Min. : 3.00 Min. : 0.740 Min. : 2.00
## 1st Qu.: 1.00 1st Qu.:81.00 1st Qu.: 4.410 1st Qu.:82.00
## Median : 4.00 Median :93.00 Median : 5.840 Median :92.00
## Mean : 44.22 Mean :83.56 Mean : 5.956 Mean :84.16
## 3rd Qu.: 29.00 3rd Qu.:97.00 3rd Qu.: 7.470 3rd Qu.:97.00
## Max. :2100.00 Max. :99.00 Max. :14.390 Max. :99.00
## HIV.AIDS GDP Population thinness..1.19.years
## Min. : 0.100 Min. : 1.68 Min. :3.400e+01 Min. : 0.100
## 1st Qu.: 0.100 1st Qu.: 462.15 1st Qu.:1.919e+05 1st Qu.: 1.600
## Median : 0.100 Median : 1592.57 Median :1.420e+06 Median : 3.000
## Mean : 1.984 Mean : 5566.03 Mean :1.465e+07 Mean : 4.851
## 3rd Qu.: 0.700 3rd Qu.: 4718.51 3rd Qu.:7.659e+06 3rd Qu.: 7.100
## Max. :50.600 Max. :119172.74 Max. :1.294e+09 Max. :27.200
## thinness.5.9.years Income.composition.of.resources Schooling
## Min. : 0.100 Min. :0.0000 Min. : 4.20
## 1st Qu.: 1.700 1st Qu.:0.5090 1st Qu.:10.30
## Median : 3.200 Median :0.6730 Median :12.30
## Mean : 4.908 Mean :0.6316 Mean :12.12
## 3rd Qu.: 7.100 3rd Qu.:0.7510 3rd Qu.:14.00
## Max. :28.200 Max. :0.9360 Max. :20.70
## Developing
## Min. :0.0000
## 1st Qu.:1.0000
## Median :1.0000
## Mean :0.8532
## 3rd Qu.:1.0000
## Max. :1.0000

```

```
head(my_data1)
```

```

## Life.expectancy Adult.Mortality infant.deaths Alcohol percentage.expenditure
## 1 65.0 263 62 0.01 71.279624
## 2 59.9 271 64 0.01 73.523582
## 3 59.9 268 66 0.01 73.219243
## 4 59.5 272 69 0.01 78.184215
## 5 59.2 275 71 0.01 7.097109
## 6 58.8 279 74 0.01 79.679367
## Hepatitis.B Measles BMI under.five.deaths Polio Total.expenditure Diphtheria
## 1 65 1154 19.1 83 6 8.16 65
## 2 62 492 18.6 86 58 8.18 62
## 3 64 430 18.1 89 62 8.13 64
## 4 67 2787 17.6 93 67 8.52 67
## 5 68 3013 17.2 97 68 7.87 68
## 6 66 1989 16.7 102 66 9.20 66
## HIV.AIDS GDP Population thinness..1.19.years thinness.5.9.years

```

```
## 1      0.1 584.25921    33736494      17.2      17.3
## 2      0.1 612.69651     327582      17.5      17.5
## 3      0.1 631.74498    31731688      17.7      17.7
## 4      0.1 669.95900    3696958      17.9      18.0
## 5      0.1  63.53723    2978599      18.2      18.2
## 6      0.1 553.32894    2883167      18.4      18.4
##      Income.composition.of.resources  Schooling  Developing
## 1              0.479      10.1      1
## 2              0.476      10.0      1
## 3              0.470       9.9      1
## 4              0.463       9.8      1
## 5              0.454       9.5      1
## 6              0.448       9.2      1
```

```
M=cor(my_data1)
M
```

```
##      Life.expectancy  Adult.Mortality  infant.deaths
## Life.expectancy      1.000000000    -0.702523062    -0.169073804
## Adult.Mortality     -0.70252306     1.000000000     0.042450237
## infant.deaths       -0.16907380     0.042450237     1.000000000
## Alcohol              0.40271832    -0.175535086    -0.106216917
## percentage.expenditure 0.40963082    -0.237609890    -0.090764632
## Hepatitis.B          0.19993528    -0.105225443    -0.231768937
## Measles              -0.06888122    -0.003966685     0.532679832
## BMI                  0.54204159    -0.351542478    -0.234425154
## under.five.deaths     -0.19226530     0.060365026     0.996905622
## Polio                 0.32729440    -0.199853000    -0.156928805
## Total.expenditure     0.17471764    -0.085226535    -0.146951117
## Diphtheria            0.34133123    -0.191428759    -0.161871004
## HIV.AIDS              -0.59223629     0.550690745     0.007711547
## GDP                   0.44132181    -0.255034733    -0.098092020
## Population            -0.02230498    -0.015011838     0.671758310
## thinness..1.19.years  -0.45783819     0.272230044     0.463415256
## thinness.5.9.years    -0.45750829     0.286722882     0.461907925
## Income.composition.of.resources 0.72108259    -0.442203288    -0.134753863
## Schooling             0.72763003    -0.421170523    -0.214371900
## Developing           -0.44279758     0.278172847     0.108756919
##      Alcohol  percentage.expenditure  Hepatitis.B
## Life.expectancy      0.40271832      0.40963082  0.19993528
## Adult.Mortality     -0.17553509     -0.23760989  -0.10522544
## infant.deaths       -0.10621692     -0.09076463  -0.23176894
## Alcohol              1.00000000      0.41704736  0.10988939
## percentage.expenditure 0.41704736      1.00000000  0.01676017
## Hepatitis.B          0.10988939      0.01676017  1.00000000
## Measles              -0.05011023     -0.06307079  -0.12479999
## BMI                  0.35339621      0.24273824  0.14330179
## under.five.deaths     -0.10108216     -0.09215806  -0.24076603
## Polio                 0.24031453      0.12862605  0.46333080
## Total.expenditure     0.21488509      0.18387236  0.11332668
## Diphtheria            0.24295143      0.13481324  0.58898993
## HIV.AIDS              -0.02711264     -0.09508499  -0.09480197
## GDP                   0.44343279      0.95929886  0.04184950
## Population            -0.02888023     -0.01679214  -0.12972265
```



## thinness..1.19.years	-0.40375499	-0.25503460	-0.12940595
## thinness.5.9.years	-0.38620819	-0.25563544	-0.13325099
## Income.composition.of.resources	0.56107433	0.40216974	0.18492097
## Schooling	0.61697481	0.42208845	0.21518159
## Developing	-0.60778179	-0.46168832	-0.14035064
##	Measles	BMI	under.five.deaths
## Life.expectancy	-0.068881222	0.54204159	-0.19226530
## Adult.Mortality	-0.003966685	-0.35154248	0.06036503
## infant.deaths	0.532679832	-0.23442515	0.99690562
## Alcohol	-0.050110235	0.35339621	-0.10108216
## percentage.expenditure	-0.063070789	0.24273824	-0.09215806
## Hepatitis.B	-0.124799993	0.14330179	-0.24076603
## Measles	1.000000000	-0.15324546	0.51750556
## BMI	-0.153245464	1.000000000	-0.24213740
## under.five.deaths	0.517505563	-0.24213740	1.000000000
## Polio	-0.057850133	0.18626797	-0.17116419
## Total.expenditure	-0.113582738	0.18946896	-0.14580310
## Diphtheria	-0.058605907	0.17629450	-0.17844819
## HIV.AIDS	-0.003521854	-0.21089675	0.01947593
## GDP	-0.064767590	0.26611397	-0.10033126
## Population	0.321946377	-0.08141598	0.65867969
## thinness..1.19.years	0.180641506	-0.54701751	0.46478470
## thinness.5.9.years	0.174946217	-0.55409398	0.46228938
## Income.composition.of.resources	-0.058277256	0.51050483	-0.14809728
## Schooling	-0.115660481	0.55484390	-0.22601262
## Developing	0.071963263	-0.29837983	0.10984742
##	Polio	Total.expenditure	Diphtheria
## Life.expectancy	0.32729440	0.17471764	0.34133123
## Adult.Mortality	-0.19985300	-0.08522653	-0.19142876
## infant.deaths	-0.15692881	-0.14695112	-0.16187100
## Alcohol	0.24031453	0.21488509	0.24295143
## percentage.expenditure	0.12862605	0.18387236	0.13481324
## Hepatitis.B	0.46333080	0.11332668	0.58898993
## Measles	-0.05785013	-0.11358274	-0.05860591
## BMI	0.18626797	0.18946896	0.17629450
## under.five.deaths	-0.17116419	-0.14580310	-0.17844819
## Polio	1.000000000	0.11976798	0.60924547
## Total.expenditure	0.11976798	1.000000000	0.12991481
## Diphtheria	0.60924547	0.12991481	1.000000000
## HIV.AIDS	-0.10788547	0.04310066	-0.11760107
## GDP	0.15680869	0.18037347	0.15843774
## Population	-0.04538657	-0.07996224	-0.03989754
## thinness..1.19.years	-0.16406959	-0.20987232	-0.18724165
## thinness.5.9.years	-0.17448925	-0.21786479	-0.18095238
## Income.composition.of.resources	0.31468159	0.18365319	0.34326177
## Schooling	0.35014660	0.24378345	0.35039793
## Developing	-0.20191667	-0.19253779	-0.20165411
##	HIV.AIDS	GDP	Population
## Life.expectancy	-0.592236293	0.44132181	-0.022304978
## Adult.Mortality	0.550690745	-0.25503473	-0.015011838
## infant.deaths	0.007711547	-0.09809202	0.671758310
## Alcohol	-0.027112636	0.44343279	-0.028880232
## percentage.expenditure	-0.095084991	0.95929886	-0.016792141
## Hepatitis.B	-0.094801971	0.04184950	-0.129722655

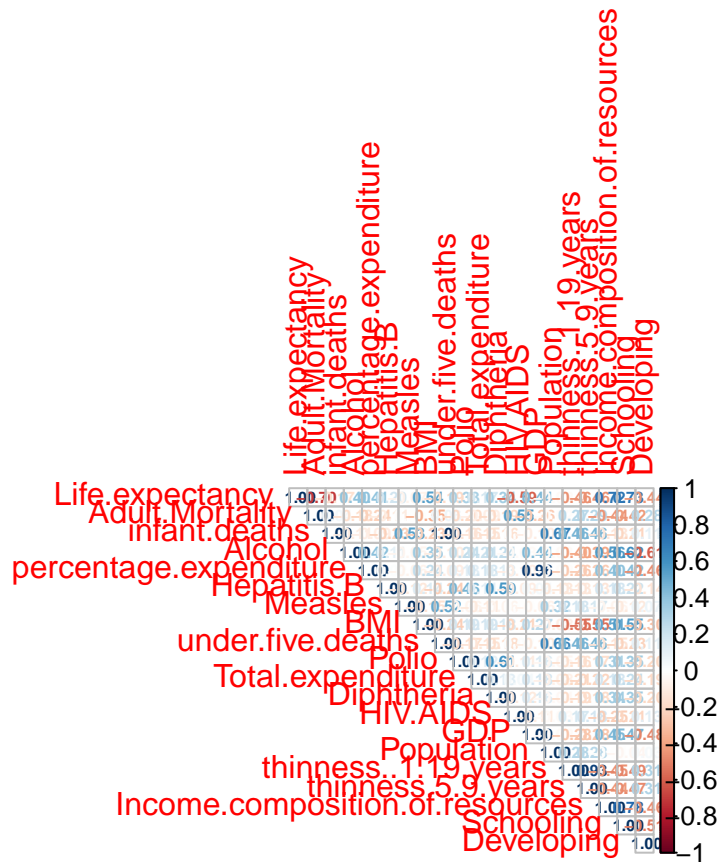
## Measles	-0.003521854	-0.06476759	0.321946377
## BMI	-0.210896746	0.26611397	-0.081415982
## under.five.deaths	0.019475927	-0.10033126	0.658679691
## Polio	-0.107885468	0.15680869	-0.045386572
## Total.expenditure	0.043100657	0.18037347	-0.079962237
## Diphtheria	-0.117601074	0.15843774	-0.039897537
## HIV.AIDS	1.000000000	-0.10808060	-0.027800562
## GDP	-0.108080600	1.000000000	-0.020368964
## Population	-0.027800562	-0.02036896	1.000000000
## thinness..1.19.years	0.172591767	-0.27749835	0.282529280
## thinness.5.9.years	0.183146727	-0.27795855	0.277913374
## Income.composition.of.resources	-0.248589855	0.44685551	-0.008132466
## Schooling	-0.211840201	0.46794697	-0.040312419
## Developing	0.129555493	-0.48480102	0.034790186
##	thinness..1.19.years	thinness.5.9.years	
## Life.expectancy	-0.4578382	-0.4575083	
## Adult.Mortality	0.2722300	0.2867229	
## infant.deaths	0.4634153	0.4619079	
## Alcohol	-0.4037550	-0.3862082	
## percentage.expenditure	-0.2550346	-0.2556354	
## Hepatitis.B	-0.1294060	-0.1332510	
## Measles	0.1806415	0.1749462	
## BMI	-0.5470175	-0.5540940	
## under.five.deaths	0.4647847	0.4622894	
## Polio	-0.1640696	-0.1744893	
## Total.expenditure	-0.2098723	-0.2178648	
## Diphtheria	-0.1872416	-0.1809524	
## HIV.AIDS	0.1725918	0.1831467	
## GDP	-0.2774983	-0.2779586	
## Population	0.2825293	0.2779134	
## thinness..1.19.years	1.0000000	0.9279134	
## thinness.5.9.years	0.9279134	1.0000000	
## Income.composition.of.resources	-0.4536789	-0.4384837	
## Schooling	-0.4911992	-0.4724820	
## Developing	0.3080053	0.3072786	
##	Income.composition.of.resources	Schooling	
## Life.expectancy	0.721082593	0.72763003	
## Adult.Mortality	-0.442203288	-0.42117052	
## infant.deaths	-0.134753863	-0.21437190	
## Alcohol	0.561074332	0.61697481	
## percentage.expenditure	0.402169736	0.42208845	
## Hepatitis.B	0.184920970	0.21518159	
## Measles	-0.058277256	-0.11566048	
## BMI	0.510504831	0.55484390	
## under.five.deaths	-0.148097276	-0.22601262	
## Polio	0.314681594	0.35014660	
## Total.expenditure	0.183653190	0.24378345	
## Diphtheria	0.343261772	0.35039793	
## HIV.AIDS	-0.248589855	-0.21184020	
## GDP	0.446855511	0.46794697	
## Population	-0.008132466	-0.04031242	
## thinness..1.19.years	-0.453678854	-0.49119921	
## thinness.5.9.years	-0.438483721	-0.47248203	
## Income.composition.of.resources	1.000000000	0.78474058	

## Schooling		0.784740581	1.00000000
## Developing		-0.463614945	-0.51254296
##	Developing		
## Life.expectancy	-0.44279758		
## Adult.Mortality	0.27817285		
## infant.deaths	0.10875692		
## Alcohol	-0.60778179		
## percentage.expenditure	-0.46168832		
## Hepatitis.B	-0.14035064		
## Measles	0.07196326		
## BMI	-0.29837983		
## under.five.deaths	0.10984742		
## Polio	-0.20191667		
## Total.expenditure	-0.19253779		
## Diphtheria	-0.20165411		
## HIV.AIDS	0.12955549		
## GDP	-0.48480102		
## Population	0.03479019		
## thinness..1.19.years	0.30800531		
## thinness.5.9.years	0.30727858		
## Income.composition.of.resources	-0.46361494		
## Schooling	-0.51254296		
## Developing	1.00000000		

```
which.min(abs(M))
```

```
## [1] 133
```

```
corrplot(M,method = "number",type="upper",number.cex = 0.5) # make correlation plot
```



Define Training and test dataset

```
set.seed(0)
tr_size <- nrow(my_data1) * 0.7 # training sample size
tr_ind <- sample(nrow(my_data1), tr_size)
data_tr <- my_data1[tr_ind, ] # training data
data_te <- my_data1[-tr_ind, ] # test data
ncol(my_data1)
```

```
## [1] 20
```

```
nrow(my_data1)
```

```
## [1] 1649
```

```
nrow(data_tr)
```

```
## [1] 1154
```

```
nrow(data_te)
```

```
## [1] 495
```

Train Model

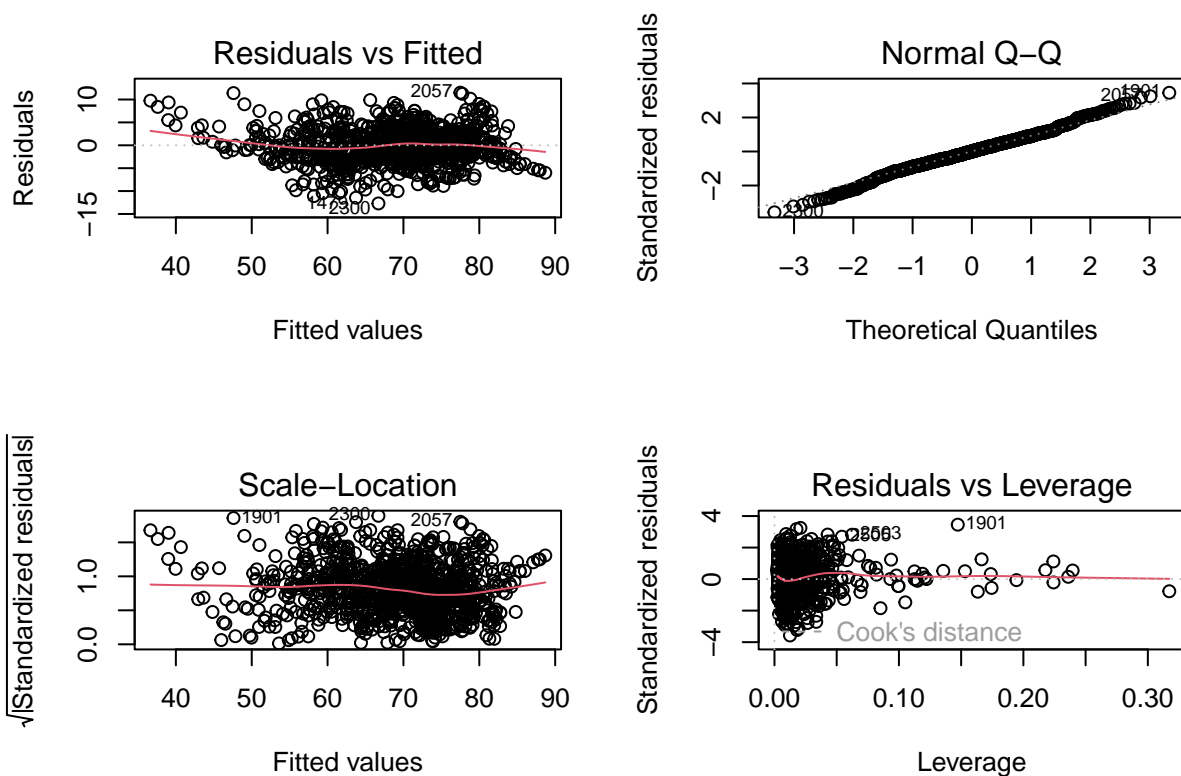
```
set.seed(0)
model <- lm(Life.expectancy~., data = data_tr)
summary(model)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ ., data = data_tr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.704  -2.164   0.010   2.225  11.494
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.461e+01  1.021e+00  53.507 < 2e-16 ***
## Adult.Mortality -1.594e-02  1.137e-03 -14.021 < 2e-16 ***
## infant.deaths    1.022e-01  1.487e-02   6.875 1.02e-11 ***
## Alcohol          -1.207e-01  3.887e-02  -3.105  0.00195 **
## percentage.expenditure 3.241e-04  2.169e-04   1.494  0.13533
## Hepatitis.B      -1.068e-02  5.205e-03  -2.052  0.04038 *
## Measles          -1.038e-05  1.317e-05  -0.788  0.43066
## BMI              3.401e-02  7.083e-03   4.802 1.78e-06 ***
## under.five.deaths -7.607e-02  1.071e-02  -7.102 2.16e-12 ***
## Polio            1.222e-02  6.156e-03   1.986  0.04731 *
## Total.expenditure 3.363e-02  4.801e-02   0.700  0.48377
## Diphtheria       1.590e-02  7.184e-03   2.214  0.02705 *
## HIV.AIDS         -4.386e-01  2.158e-02 -20.325 < 2e-16 ***
## GDP              9.858e-06  3.440e-05   0.287  0.77450
## Population       -1.724e-09  2.116e-09  -0.815  0.41542
## thinness..1.19.years -1.341e-02  5.635e-02  -0.238  0.81193
## thinness.5.9.years  -5.432e-02  5.579e-02  -0.974  0.33043
## Income.composition.of.resources 1.045e+01  1.015e+00  10.293 < 2e-16 ***
## Schooling        8.510e-01  6.999e-02  12.159 < 2e-16 ***
## Developing       -1.144e+00  4.059e-01  -2.818  0.00491 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.576 on 1134 degrees of freedom
## Multiple R-squared:  0.835, Adjusted R-squared:  0.8323
## F-statistic: 302.1 on 19 and 1134 DF, p-value: < 2.2e-16
```

```
alias(model)# Check that the predictors of the model do not contain a large number of repeated values
```

```
## Model :
## Life.expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
##     percentage.expenditure + Hepatitis.B + Measles + BMI + under.five.deaths +
##     Polio + Total.expenditure + Diphtheria + HIV.AIDS + GDP +
##     Population + thinness..1.19.years + thinness.5.9.years +
##     Income.composition.of.resources + Schooling + Developing
```

```
par (mfrow = c(2,2))
plot (model)
```



```
shapiro.test(model$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.99402, p-value = 0.0001406
```

```
dwtest(model)
```

```
##
##  Durbin-Watson test
##
## data:  model
## DW = 2.0436, p-value = 0.7729
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bptest(model)
```

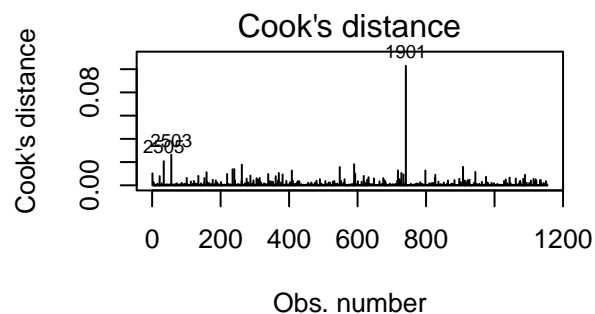
```
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 125.61, df = 19, p-value < 2.2e-16
```

```
AA<-rstudent(model) # Compute standardized residuals to check outlines.
p<-ncol(data_tr)
n<-nrow(data_tr)
which(abs(AA)>qt(1-0.05/(n*2),n-p-1))
```

```
## named integer(0)
```

```
plot(model,which=4)
vif(model)
```

##	Adult.Mortality	infant.deaths
##	1.819167	254.411298
##	Alcohol	percentage.expenditure
##	2.258261	14.421237
##	Hepatitis.B	Measles
##	1.660452	1.542678
##	BMI	under.five.deaths
##	1.772660	238.492903
##	Polio	Total.expenditure
##	1.712429	1.115458
##	Diphtheria	HIV.AIDS
##	2.049503	1.500308
##	GDP	Population
##	15.121553	2.264707
##	thinness..1.19.years	thinness.5.9.years
##	5.839108	5.826642
##	Income.composition.of.resources	Schooling
##	3.069910	3.606260
##	Developing	
##	1.867234	



AIC Selection

```
step(model)
```

```
## Start: AIC=2960.78
## Life expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
##   percentage.expenditure + Hepatitis.B + Measles + BMI + under.five.deaths +
##   Polio + Total.expenditure + Diphtheria + HIV.AIDS + GDP +
##   Population + thinness..1.19.years + thinness.5.9.years +
##   Income.composition.of.resources + Schooling + Developing
##
##
```

	Df	Sum of Sq	RSS	AIC
## - thinness..1.19.years	1	0.7	14502	2958.8
## - GDP	1	1.1	14502	2958.9
## - Total.expenditure	1	6.3	14508	2959.3
## - Measles	1	7.9	14509	2959.4
## - Population	1	8.5	14510	2959.5
## - thinness.5.9.years	1	12.1	14513	2959.7
## <none>			14501	2960.8
## - percentage.expenditure	1	28.6	14530	2961.0
## - Polio	1	50.4	14552	2962.8
## - Hepatitis.B	1	53.9	14555	2963.1
## - Diphtheria	1	62.7	14564	2963.8
## - Developing	1	101.6	14603	2966.8
## - Alcohol	1	123.3	14625	2968.5
## - BMI	1	294.8	14796	2982.0



```

## - infant.deaths          1      604.4 15106 3005.9
## - under.five.deaths      1      645.1 15146 3009.0
## - Income.composition.of.resources 1      1354.7 15856 3061.8
## - Schooling              1      1890.4 16392 3100.2
## - Adult.Mortality        1      2514.0 17015 3143.3
## - HIV.AIDS               1      5282.4 19784 3317.2
##
## Step: AIC=2958.84
## Life expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
##   percentage.expenditure + Hepatitis.B + Measles + BMI + under.five.deaths +
##   Polio + Total.expenditure + Diphtheria + HIV.AIDS + GDP +
##   Population + thinness.5.9.years + Income.composition.of.resources +
##   Schooling + Developing
##
##              Df Sum of Sq  RSS    AIC
## - GDP          1      1.0 14503 2956.9
## - Total.expenditure 1      6.3 14508 2957.3
## - Measles       1      8.0 14510 2957.5
## - Population    1      8.8 14511 2957.5
## <none>          14502 2958.8
## - percentage.expenditure 1     28.6 14531 2959.1
## - Polio         1     49.8 14552 2960.8
## - thinness.5.9.years 1     52.6 14555 2961.0
## - Hepatitis.B   1     54.3 14556 2961.1
## - Diphtheria    1     63.2 14565 2961.9
## - Developing    1    101.2 14603 2964.9
## - Alcohol       1    122.6 14625 2966.6
## - BMI           1    297.2 14799 2980.3
## - infant.deaths 1    606.4 15108 3004.1
## - under.five.deaths 1    647.7 15150 3007.3
## - Income.composition.of.resources 1   1361.3 15863 3060.4
## - Schooling     1   1906.4 16408 3099.4
## - Adult.Mortality 1   2514.6 17017 3141.4
## - HIV.AIDS      1   5293.1 19795 3315.9
##
## Step: AIC=2956.92
## Life expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
##   percentage.expenditure + Hepatitis.B + Measles + BMI + under.five.deaths +
##   Polio + Total.expenditure + Diphtheria + HIV.AIDS + Population +
##   thinness.5.9.years + Income.composition.of.resources + Schooling +
##   Developing
##
##              Df Sum of Sq  RSS    AIC
## - Total.expenditure 1      6.1 14509 2955.4
## - Measles           1      7.9 14511 2955.5
## - Population        1      9.0 14512 2955.6
## <none>              14503 2956.9
## - Polio             1     50.4 14554 2958.9
## - thinness.5.9.years 1     52.7 14556 2959.1
## - Hepatitis.B       1     53.7 14557 2959.2
## - Diphtheria        1     62.7 14566 2959.9
## - Developing        1    102.2 14605 2963.0
## - Alcohol           1    122.2 14625 2964.6
## - BMI               1    297.7 14801 2978.4

```

```

## - percentage.expenditure      1      403.1 14906 2986.6
## - infant.deaths               1      607.1 15110 3002.2
## - under.five.deaths           1      648.2 15151 3005.4
## - Income.composition.of.resources 1     1374.4 15878 3059.4
## - Schooling                   1     1921.7 16425 3098.5
## - Adult.Mortality             1     2514.1 17017 3139.4
## - HIV.AIDS                    1     5292.7 19796 3313.9
##
## Step: AIC=2955.4
## Life expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
##     percentage.expenditure + Hepatitis.B + Measles + BMI + under.five.deaths +
##     Polio + Diphtheria + HIV.AIDS + Population + thinness.5.9.years +
##     Income.composition.of.resources + Schooling + Developing
##
##              Df Sum of Sq  RSS    AIC
## - Measles      1      8.4 14518 2954.1
## - Population    1      8.9 14518 2954.1
## <none>          1     14509 2955.4
## - Polio         1     51.1 14560 2957.5
## - Hepatitis.B   1     52.1 14561 2957.5
## - thinness.5.9.years 1     55.8 14565 2957.8
## - Diphtheria    1     63.6 14573 2958.5
## - Developing    1    103.8 14613 2961.6
## - Alcohol       1    121.7 14631 2963.0
## - BMI           1    303.0 14812 2977.3
## - percentage.expenditure 1    411.2 14920 2985.7
## - infant.deaths 1    603.6 15113 3000.4
## - under.five.deaths 1    644.7 15154 3003.6
## - Income.composition.of.resources 1   1371.1 15880 3057.6
## - Schooling     1   1938.1 16447 3098.1
## - Adult.Mortality 1   2521.7 17031 3138.3
## - HIV.AIDS      1   5322.5 19832 3314.0
##
## Step: AIC=2954.07
## Life expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
##     percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +
##     Polio + Diphtheria + HIV.AIDS + Population + thinness.5.9.years +
##     Income.composition.of.resources + Schooling + Developing
##
##              Df Sum of Sq  RSS    AIC
## - Population    1      6.9 14524 2952.6
## <none>          1     14518 2954.1
## - Polio         1     50.6 14568 2956.1
## - Hepatitis.B   1     51.5 14569 2956.2
## - thinness.5.9.years 1     51.7 14569 2956.2
## - Diphtheria    1     64.1 14582 2957.2
## - Developing    1    104.8 14622 2960.4
## - Alcohol       1    125.2 14643 2962.0
## - BMI           1    314.6 14832 2976.8
## - percentage.expenditure 1    413.3 14931 2984.5
## - infant.deaths 1    644.6 15162 3002.2
## - under.five.deaths 1    675.0 15192 3004.5
## - Income.composition.of.resources 1   1375.5 15893 3056.5
## - Schooling     1   1938.8 16456 3096.7

```

```

## - Adult.Mortality          1    2536.2 17054 3137.9
## - HIV.AIDS                 1    5328.9 19846 3312.9
##
## Step:  AIC=2952.62
## Life.expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
##     percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +
##     Polio + Diphtheria + HIV.AIDS + thinness.5.9.years + Income.composition.of.resources +
##     Schooling + Developing
##
##                                Df Sum of Sq  RSS    AIC
## <none>                                14524 2952.6
## - Polio                             1      50.0 14574 2954.6
## - Hepatitis.B                       1      50.9 14575 2954.7
## - thinness.5.9.years                1      53.7 14578 2954.9
## - Diphtheria                       1      62.7 14587 2955.6
## - Developing                       1     105.1 14630 2958.9
## - Alcohol                          1     125.3 14650 2960.5
## - BMI                             1     310.1 14834 2975.0
## - percentage.expenditure            1     411.4 14936 2982.8
## - infant.deaths                    1     660.9 15185 3002.0
## - under.five.deaths                 1     677.4 15202 3003.2
## - Income.composition.of.resources  1    1382.0 15906 3055.5
## - Schooling                        1    1933.8 16458 3094.9
## - Adult.Mortality                   1    2554.6 17079 3137.6
## - HIV.AIDS                         1    5324.4 19849 3311.0
##
##
## Call:
## lm(formula = Life.expectancy ~ Adult.Mortality + infant.deaths +
##     Alcohol + percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +
##     Polio + Diphtheria + HIV.AIDS + thinness.5.9.years + Income.composition.of.resources +
##     Schooling + Developing, data = data_tr)
##
## Coefficients:
##                (Intercept)                Adult.Mortality
##                54.7468812                -0.0160171
##                infant.deaths                Alcohol
##                0.0945725                -0.1212779
##                percentage.expenditure        Hepatitis.B
##                0.0003858                -0.0103587
##                BMI                under.five.deaths
##                0.0345796                -0.0713724
##                Polio                Diphtheria
##                0.0121297                0.0158760
##                HIV.AIDS                thinness.5.9.years
##                -0.4370189                -0.0651349
## Income.composition.of.resources                Schooling
##                10.4991049                0.8516676
##                Developing
##                -1.1615165

```

```

# Find model with lowest AIC

```

```

lmod_AIC_B<-lm(Life.expectancy ~ Adult.Mortality + infant.deaths + Alcohol +
    percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +

```

```

Polio + Diphtheria + HIV.AIDS + thinness.5.9.years +
Income.composition.of.resources + Schooling + Developing,
data = data_tr) # AIC selected model
sum_AIC_B<-summary(lmod_AIC_B)
sum_AIC_B

##
## Call:
## lm(formula = Life.expectancy ~ Adult.Mortality + infant.deaths +
##     Alcohol + percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +
##     Polio + Diphtheria + HIV.AIDS + thinness.5.9.years + Income.composition.of.resources +
##     Schooling + Developing, data = data_tr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4930  -2.1727   0.0338   2.2373  11.6085
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.475e+01  9.748e-01  56.161 < 2e-16 ***
## Adult.Mortality  -1.602e-02  1.132e-03 -14.154 < 2e-16 ***
## infant.deaths     9.457e-02  1.314e-02   7.199 1.10e-12 ***
## Alcohol          -1.213e-01  3.868e-02  -3.135  0.00176 **
## percentage.expenditure  3.858e-04  6.793e-05   5.680 1.71e-08 ***
## Hepatitis.B       -1.036e-02  5.182e-03  -1.999  0.04587 *
## BMI               3.458e-02  7.012e-03   4.932 9.37e-07 ***
## under.five.deaths -7.137e-02  9.792e-03  -7.289 5.84e-13 ***
## Polio             1.213e-02  6.127e-03   1.980  0.04797 *
## Diphtheria        1.588e-02  7.160e-03   2.217  0.02680 *
## HIV.AIDS          -4.370e-01  2.139e-02 -20.434 < 2e-16 ***
## thinness.5.9.years -6.513e-02  3.175e-02  -2.052  0.04043 *
## Income.composition.of.resources  1.050e+01  1.009e+00  10.411 < 2e-16 ***
## Schooling         8.517e-01  6.916e-02  12.314 < 2e-16 ***
## Developing       -1.162e+00  4.045e-01  -2.871  0.00416 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.571 on 1139 degrees of freedom
## Multiple R-squared:  0.8348, Adjusted R-squared:  0.8328
## F-statistic: 411.1 on 14 and 1139 DF, p-value: < 2.2e-16

vif(lmod_AIC_B)

##              Adult.Mortality              infant.deaths
##              1.808486              199.139250
##              Alcohol              percentage.expenditure
##              2.242487              1.418817
##              Hepatitis.B              BMI
##              1.651012              1.742053
##              under.five.deaths              Polio
##              199.915027              1.701046
##              Diphtheria              HIV.AIDS
##              2.041567              1.477832

```

```
##          thinness.5.9.years Income.composition.of.resources
##          1.892364          3.040356
##          Schooling          Developing
##          3.531003          1.860123
```

BIC Selection

```
set.seed(0)
fit_null<-lm(Life.expectancy~1,data_tr)
step(fit_null, scope = list(lower = fit_null, upper = model), direction = "both",
criterion = "BIC", k = log(n))
```

```
## Start:  AIC=5007.43
## Life.expectancy ~ 1
##
##          Df Sum of Sq  RSS   AIC
## + Schooling          1    46348 41563 4150.0
## + Income.composition.of.resources  1    46324 41586 4150.7
## + Adult.Mortality      1    42969 44941 4240.2
## + HIV.AIDS             1    30508 57403 4522.6
## + BMI                  1    24795 63116 4632.1
## + thinness.5.9.years    1    20496 67415 4708.1
## + thinness..1.19.years  1    20454 67456 4708.9
## + Developing           1    17741 70169 4754.4
## + GDP                  1    16988 70923 4766.7
## + percentage.expenditure 1    14667 73244 4803.8
## + Alcohol              1    14322 73589 4809.3
## + Polio                1     9706 78205 4879.5
## + Diphtheria           1     9526 78385 4882.1
## + under.five.deaths     1     3146 84765 4972.4
## + Hepatitis.B           1     2772 85139 4977.5
## + infant.deaths         1     2460 85450 4981.7
## + Total.expenditure     1     1948 85962 4988.6
## <none>                  87911 5007.4
## + Measles              1       276 87635 5010.9
## + Population            1        38 87873 5014.0
##
## Step:  AIC=4150.01
## Life.expectancy ~ Schooling
##
##          Df Sum of Sq  RSS   AIC
## + HIV.AIDS          1    17966 23597 3503.8
## + Adult.Mortality    1    16736 24827 3562.4
## + Income.composition.of.resources  1     5400 36164 3996.5
## + BMI                1     2063 39500 4098.3
## + thinness.5.9.years  1     1844 39720 4104.7
## + thinness..1.19.years 1     1437 40126 4116.5
## + GDP                1     1100 40463 4126.1
## + percentage.expenditure 1     1046 40517 4127.7
## + Polio              1       776 40787 4135.3
## + Diphtheria         1       644 40919 4139.1
## + Developing          1       555 41008 4141.5
## + Alcohol             1       355 41208 4147.2
```

```

## <none> 41563 4150.0
## + Hepatitis.B 1 75 41488 4155.0
## + Measles 1 38 41525 4156.0
## + under.five.deaths 1 33 41530 4156.1
## + Total.expenditure 1 25 41538 4156.4
## + Population 1 8 41555 4156.8
## + infant.deaths 1 2 41561 4157.0
## - Schooling 1 46348 87911 5007.4
##
## Step: AIC=3503.81
## Life expectancy ~ Schooling + HIV.AIDS
##
## Df Sum of Sq RSS AIC
## + Adult.Mortality 1 4770 18827 3250.3
## + Income.composition.of.resources 1 3110 20487 3347.7
## + GDP 1 1017 22581 3460.0
## + percentage.expenditure 1 992 22605 3461.3
## + BMI 1 984 22613 3461.7
## + thinness.5.9.years 1 641 22956 3479.1
## + thinness..1.19.years 1 500 23097 3486.1
## + Diphtheria 1 486 23111 3486.8
## + Polio 1 452 23146 3488.6
## + Developing 1 395 23202 3491.4
## <none> 23597 3503.8
## + under.five.deaths 1 98 23499 3506.0
## + Total.expenditure 1 90 23507 3506.4
## + infant.deaths 1 47 23550 3508.5
## + Hepatitis.B 1 21 23577 3509.9
## + Measles 1 8 23590 3510.5
## + Population 1 5 23592 3510.6
## + Alcohol 1 1 23596 3510.8
## - HIV.AIDS 1 17966 41563 4150.0
## - Schooling 1 33805 57403 4522.6
##
## Step: AIC=3250.26
## Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality
##
## Df Sum of Sq RSS AIC
## + Income.composition.of.resources 1 2088.5 16739 3121.6
## + GDP 1 725.8 18102 3211.9
## + percentage.expenditure 1 704.4 18123 3213.3
## + BMI 1 623.1 18204 3218.5
## + thinness.5.9.years 1 428.3 18399 3230.8
## + thinness..1.19.years 1 388.2 18439 3233.3
## + Diphtheria 1 357.2 18470 3235.2
## + Polio 1 285.0 18542 3239.7
## + Developing 1 231.8 18596 3243.0
## <none> 18827 3250.3
## + under.five.deaths 1 105.0 18722 3250.9
## + infant.deaths 1 63.3 18764 3253.4
## + Total.expenditure 1 49.4 18778 3254.3
## + Hepatitis.B 1 17.6 18810 3256.2
## + Alcohol 1 15.4 18812 3256.4
## + Population 1 5.1 18822 3257.0

```

```

## + Measles                1      2.9 18824 3257.1
## - Adult.Mortality        1    4769.9 23597 3503.8
## - HIV.AIDS               1    5999.7 24827 3562.4
## - Schooling              1   20937.3 39765 4106.0
##
## Step: AIC=3121.63
## Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality + Income.composition.of.resources
##
##
## Df Sum of Sq  RSS    AIC
## + percentage.expenditure  1    459.1 16280 3096.6
## + GDP                     1    437.7 16301 3098.1
## + BMI                     1    425.0 16314 3099.0
## + thinness.5.9.years      1    249.5 16489 3111.4
## + thinness..1.19.years    1    206.5 16532 3114.4
## + Diphtheria              1    194.5 16544 3115.2
## + Polio                   1    169.1 16570 3117.0
## + under.five.deaths       1    157.3 16582 3117.8
## + Developing              1    124.6 16614 3120.1
## + infant.deaths           1    109.6 16629 3121.1
## <none>                     1    16739 3121.6
## + Total.expenditure       1     44.7 16694 3125.6
## + Alcohol                  1     29.1 16710 3126.7
## + Population              1     14.7 16724 3127.7
## + Hepatitis.B             1      7.9 16731 3128.1
## + Measles                 1      0.1 16739 3128.7
## - Income.composition.of.resources 1   2088.5 18827 3250.3
## - Adult.Mortality         1   3748.0 20487 3347.7
## - Schooling               1   3916.8 20656 3357.2
## - HIV.AIDS                1   5740.0 22479 3454.8
##
## Step: AIC=3096.59
## Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality + Income.composition.of.resources +
##   percentage.expenditure
##
##
## Df Sum of Sq  RSS    AIC
## + BMI                     1    435.4 15844 3072.4
## + thinness.5.9.years      1    217.3 16062 3088.1
## + Diphtheria              1    212.4 16067 3088.5
## + Polio                   1    193.5 16086 3089.8
## + thinness..1.19.years    1    185.3 16094 3090.4
## + under.five.deaths       1    158.8 16121 3092.3
## + infant.deaths           1    108.4 16171 3095.9
## <none>                     1    16280 3096.6
## + Alcohol                  1     96.8 16183 3096.8
## + Total.expenditure       1     24.8 16255 3101.9
## + Hepatitis.B             1     22.7 16257 3102.0
## + Developing              1     20.2 16260 3102.2
## + Population              1     14.2 16266 3102.6
## + GDP                     1      2.4 16277 3103.5
## + Measles                 1      0.1 16280 3103.6
## - percentage.expenditure  1    459.1 16739 3121.6
## - Income.composition.of.resources 1   1843.1 18123 3213.3
## - Schooling               1   3328.1 19608 3304.2
## - Adult.Mortality         1   3594.7 19875 3319.8

```

```

## - HIV.AIDS          1    5854.4 22134 3444.0
##
## Step: AIC=3072.35
## Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality + Income.composition.of.resources +
##     percentage.expenditure + BMI
##
##              Df Sum of Sq  RSS    AIC
## + Diphtheria    1    239.9 15604 3061.8
## + Polio          1    207.0 15637 3064.2
## + Alcohol        1    118.1 15726 3070.8
## <none>                      15844 3072.4
## + under.five.deaths    1     92.2 15752 3072.7
## + thinness.5.9.years   1     56.2 15788 3075.3
## + infant.deaths       1     53.9 15790 3075.5
## + thinness..1.19.years 1     46.1 15798 3076.0
## + Hepatitis.B         1     20.5 15824 3077.9
## + Developing          1     18.5 15826 3078.1
## + Measles             1      8.8 15836 3078.8
## + Total.expenditure    1      8.8 15836 3078.8
## + Population          1      4.1 15840 3079.1
## + GDP                 1      1.7 15843 3079.3
## - BMI                1    435.4 16280 3096.6
## - percentage.expenditure 1    469.6 16314 3099.0
## - Income.composition.of.resources 1 1655.0 17499 3179.9
## - Schooling          1 2429.6 18274 3229.9
## - Adult.Mortality     1 3376.7 19221 3288.2
## - HIV.AIDS           1 5727.9 21572 3421.4
##
## Step: AIC=3061.8
## Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality + Income.composition.of.resources +
##     percentage.expenditure + BMI + Diphtheria
##
##              Df Sum of Sq  RSS    AIC
## + Alcohol        1    132.3 15472 3059.0
## <none>                      15604 3061.8
## + under.five.deaths    1     56.1 15548 3064.7
## + Polio              1     50.5 15554 3065.1
## + thinness.5.9.years   1     46.3 15558 3065.4
## + thinness..1.19.years 1     38.8 15566 3066.0
## + infant.deaths       1     29.2 15575 3066.7
## + Hepatitis.B         1     24.9 15580 3067.0
## + Developing          1     15.6 15589 3067.7
## + Measles             1     11.4 15593 3068.0
## + Total.expenditure    1      3.2 15601 3068.6
## + Population          1      2.3 15602 3068.7
## + GDP                 1      1.3 15603 3068.8
## - Diphtheria         1    239.9 15844 3072.4
## - BMI                1    462.9 16067 3088.5
## - percentage.expenditure 1    489.1 16094 3090.4
## - Income.composition.of.resources 1 1490.0 17095 3160.0
## - Schooling          1 2186.1 17791 3206.1
## - Adult.Mortality     1 3312.8 18917 3276.9
## - HIV.AIDS           1 5743.2 21348 3416.4
##

```



```

## Step: AIC=3059.02
## Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality + Income.composition.of.resources +
##     percentage.expenditure + BMI + Diphtheria + Alcohol
##
##
##           Df Sum of Sq  RSS   AIC
## <none>                15472 3059.0
## + Developing           1    83.5 15389 3059.8
## + thinness.5.9.years    1    66.5 15406 3061.1
## + thinness..1.19.years  1    58.4 15414 3061.7
## - Alcohol              1   132.3 15604 3061.8
## + Polio                1    52.8 15419 3062.1
## + under.five.deaths     1    49.6 15423 3062.4
## + Hepatitis.B           1    27.6 15445 3064.0
## + infant.deaths         1    26.0 15446 3064.1
## + Measles               1    12.8 15459 3065.1
## + Total.expenditure     1     4.9 15467 3065.7
## + Population            1     2.7 15470 3065.9
## + GDP                   1     2.4 15470 3065.9
## - Diphtheria            1   254.1 15726 3070.8
## - BMI                   1   487.0 15959 3087.7
## - percentage.expenditure 1   574.6 16047 3094.1
## - Income.composition.of.resources 1 1600.0 17072 3165.5
## - Schooling             1 2309.9 17782 3212.5
## - Adult.Mortality       1 3102.7 18575 3262.9
## - HIV.AIDS              1 5571.4 21044 3406.9

##
## Call:
## lm(formula = Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality +
##     Income.composition.of.resources + percentage.expenditure +
##     BMI + Diphtheria + Alcohol, data = data_tr)
##
## Coefficients:
##              (Intercept)                Schooling
##              51.6253207                0.9161955
##              HIV.AIDS                Adult.Mortality
##              -0.4441776                -0.0174209
## Income.composition.of.resources    percentage.expenditure
##              11.1794749                0.0004363
##              BMI                Diphtheria
##              0.0401836                0.0240417
##              Alcohol
##              -0.1124538

lmod_BIC_B0<-lm(Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality +
                Income.composition.of.resources +
                percentage.expenditure + BMI + Diphtheria + Alcohol,data_tr) # BIC selected model
sum_BIC_B0<-summary(lmod_BIC_B0)
sum_BIC_B0

##
## Call:
## lm(formula = Life expectancy ~ Schooling + HIV.AIDS + Adult.Mortality +

```

```
##      Income.composition.of.resources + percentage.expenditure +
##      BMI + Diphtheria + Alcohol, data = data_tr)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -12.7365  -2.1572   0.0879   2.3716  12.3880
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      51.6253207   0.7516596   68.682 < 2e-16 ***
## Schooling         0.9161955   0.0700758   13.074 < 2e-16 ***
## HIV.AIDS        -0.4441776   0.0218750  -20.305 < 2e-16 ***
## Adult.Mortality  -0.0174209   0.0011497  -15.153 < 2e-16 ***
## Income.composition.of.resources 11.1794749   1.0273773   10.882 < 2e-16 ***
## percentage.expenditure 0.0004363   0.0000669    6.521 1.05e-10 ***
## BMI              0.0401836   0.0066937    6.003 2.59e-09 ***
## Diphtheria       0.0240417   0.0055439    4.337 1.57e-05 ***
## Alcohol          -0.1124538   0.0359347   -3.129 0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.676 on 1145 degrees of freedom
## Multiple R-squared:  0.824, Adjusted R-squared:  0.8228
## F-statistic: 670.1 on 8 and 1145 DF, p-value: < 2.2e-16
```

```
vif(lmod_BIC_B0)
```

```
##              Schooling              HIV.AIDS
##              3.420966              1.458970
##      Adult.Mortality Income.composition.of.resources
##              1.761482              2.977498
##      percentage.expenditure              BMI
##              1.298893              1.498168
##              Diphtheria              Alcohol
##              1.155086              1.826288
```

```
#lasso
```

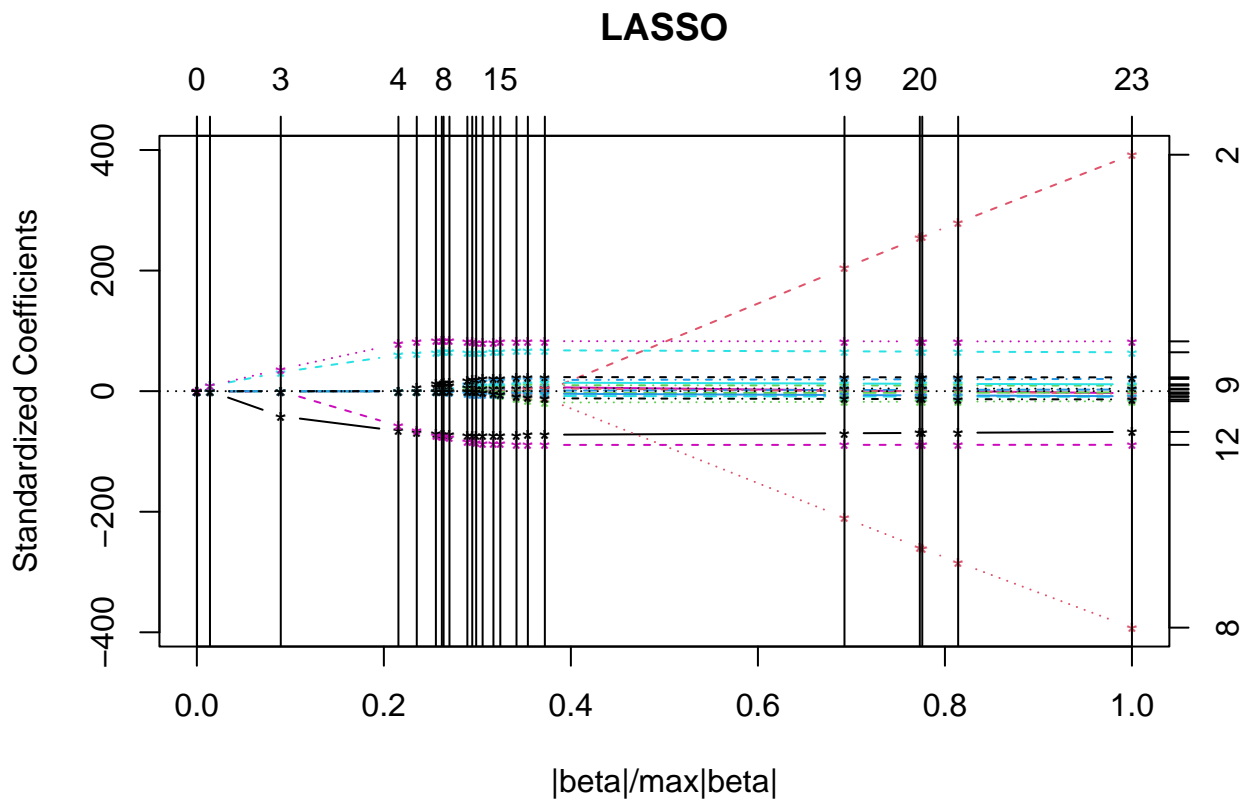
```
set.seed(0)
x_tr = as.matrix(data_tr[,c(2:ncol(data_tr))])
y_tr = as.matrix(data_tr[,1])

lar1 <- lars(x_tr, y_tr, type = "lasso")
lar1
```

```
##
## Call:
## lars(x = x_tr, y = y_tr, type = "lasso")
## R-squared: 0.835
## Sequence of LASSO moves:
##      Schooling Income.composition.of.resources Adult.Mortality HIV.AIDS BMI GDP
## Var      18              17              1      12  7  13
## Step      1              2              3      4  5  6
```

```
##      thinness.5.9.years percentage.expenditure Diphtheria Polio Developing
## Var          16              4          11    9          19
## Step          7              8          9    10          11
##      thinness..1.19.years under.five.deaths Alcohol Measles Hepatitis.B
## Var          15              8    3    6          5
## Step         12             13    14    15          16
##      Population Total.expenditure infant.deaths Population Population Measles
## Var          14              10    2   -14         14    -6
## Step         17             18    19    20         21    22
##      Measles
## Var          6
## Step         23
```

```
plot(lar1)
```



```
sum1<-summary(lar1)
sum1
```

```
## LARS/LASSO
## Call: lars(x = x_tr, y = y_tr, type = "lasso")
##      Df    Rss      Cp
## 0     1 87911 5722.608
## 1     2 87800 5715.949
## 2     3 80822 5172.240
## 3     4 49795 2747.932
```

```
## 4 5 21174 511.822
## 5 6 19075 349.630
## 6 7 17470 226.175
## 7 8 17084 197.984
## 8 9 16979 191.740
## 9 10 16657 168.563
## 10 11 15922 113.129
## 11 12 15785 104.360
## 12 13 15694 99.237
## 13 14 15579 92.247
## 14 15 15439 83.307
## 15 16 15368 79.803
## 16 17 15242 71.888
## 17 18 15179 69.022
## 18 19 15122 66.523
## 19 20 14644 31.127
## 20 19 14577 23.924
## 21 20 14575 25.786
## 22 19 14552 21.952
## 23 20 14501 20.000
```

```
lar1$Cp[which.min(lar1$Cp)]
```

```
## 23
## 20
```

```
lar1$beta
```

##	Adult.Mortality	infant.deaths	Alcohol	percentage.expenditure
## 0	0.00000000	0.00000000	0.00000000	0.000000e+00
## 1	0.00000000	0.00000000	0.00000000	0.000000e+00
## 2	0.00000000	0.00000000	0.00000000	0.000000e+00
## 3	-0.00977413	0.00000000	0.00000000	0.000000e+00
## 4	-0.01555856	0.00000000	0.00000000	0.000000e+00
## 5	-0.01625819	0.00000000	0.00000000	0.000000e+00
## 6	-0.01678625	0.00000000	0.00000000	0.000000e+00
## 7	-0.01693850	0.00000000	0.00000000	0.000000e+00
## 8	-0.01698179	0.00000000	0.00000000	1.263759e-05
## 9	-0.01708372	0.00000000	0.00000000	5.010667e-05
## 10	-0.01734371	0.00000000	0.00000000	1.683132e-04
## 11	-0.01740584	0.00000000	0.00000000	1.979189e-04
## 12	-0.01745870	0.00000000	0.00000000	2.213531e-04
## 13	-0.01754419	0.00000000	0.00000000	2.547775e-04
## 14	-0.01742236	0.00000000	-0.03848159	2.752585e-04
## 15	-0.01735656	0.00000000	-0.05763878	2.868010e-04
## 16	-0.01721769	0.00000000	-0.09127957	2.946223e-04
## 17	-0.01714550	0.00000000	-0.10985228	2.966879e-04
## 18	-0.01703610	0.00000000	-0.13670145	2.967200e-04
## 19	-0.01646289	0.05327328	-0.12836635	3.109836e-04
## 20	-0.01632439	0.06640825	-0.12625739	3.130763e-04
## 21	-0.01631982	0.06683012	-0.12619202	3.132060e-04
## 22	-0.01625072	0.07292281	-0.12541442	3.156637e-04
## 23	-0.01593616	0.10222767	-0.12070699	3.240908e-04

##	Hepatitis.B	Measles	BMI	under.five.deaths	Polio
## 0	0.000000000	0.000000e+00	0.000000000	0.0000000000	0.0000000000
## 1	0.000000000	0.000000e+00	0.000000000	0.0000000000	0.0000000000
## 2	0.000000000	0.000000e+00	0.000000000	0.0000000000	0.0000000000
## 3	0.000000000	0.000000e+00	0.000000000	0.0000000000	0.0000000000
## 4	0.000000000	0.000000e+00	0.000000000	0.0000000000	0.0000000000
## 5	0.000000000	0.000000e+00	0.009226295	0.0000000000	0.0000000000
## 6	0.000000000	0.000000e+00	0.017592735	0.0000000000	0.0000000000
## 7	0.000000000	0.000000e+00	0.019409866	0.0000000000	0.0000000000
## 8	0.000000000	0.000000e+00	0.019949278	0.0000000000	0.0000000000
## 9	0.000000000	0.000000e+00	0.021655012	0.0000000000	0.0000000000
## 10	0.000000000	0.000000e+00	0.026638834	0.0000000000	0.004705920
## 11	0.000000000	0.000000e+00	0.027970276	0.0000000000	0.005949128
## 12	0.000000000	0.000000e+00	0.029006007	0.0000000000	0.006964178
## 13	0.000000000	0.000000e+00	0.030508272	-0.0003299328	0.008273306
## 14	0.000000000	0.000000e+00	0.031388973	-0.0004451650	0.009131204
## 15	0.000000000	3.892713e-06	0.032029851	-0.0006322507	0.009512640
## 16	-0.003821677	1.020530e-05	0.033323547	-0.0010250180	0.011008049
## 17	-0.005932449	1.387622e-05	0.033892430	-0.0015209059	0.011790546
## 18	-0.009032651	1.930010e-05	0.034618761	-0.0022343171	0.012887703
## 19	-0.009891428	3.833197e-06	0.034301700	-0.0407117606	0.012541709
## 20	-0.010098465	1.237070e-07	0.034132101	-0.0503661502	0.012427408
## 21	-0.010105321	0.000000e+00	0.034130667	-0.0506688800	0.012425010
## 22	-0.010208786	0.000000e+00	0.034202682	-0.0551016832	0.012372422
## 23	-0.010680583	-1.037980e-05	0.034010342	-0.0760698238	0.012223766
##	Total.expenditure	Diphtheria	HIV.AIDS	GDP	Population
## 0	0.000000000	0.000000000	0.00000000	0.000000e+00	0.000000e+00
## 1	0.000000000	0.000000000	0.00000000	0.000000e+00	0.000000e+00
## 2	0.000000000	0.000000000	0.00000000	0.000000e+00	0.000000e+00
## 3	0.000000000	0.000000000	0.00000000	0.000000e+00	0.000000e+00
## 4	0.000000000	0.000000000	-0.2844093	0.000000e+00	0.000000e+00
## 5	0.000000000	0.000000000	-0.3237722	0.000000e+00	0.000000e+00
## 6	0.000000000	0.000000000	-0.3603766	1.753426e-05	0.000000e+00
## 7	0.000000000	0.000000000	-0.3709481	2.262704e-05	0.000000e+00
## 8	0.000000000	0.000000000	-0.3740586	2.219802e-05	0.000000e+00
## 9	0.000000000	0.002753284	-0.3828233	2.079696e-05	0.000000e+00
## 10	0.000000000	0.008042292	-0.4082185	1.547943e-05	0.000000e+00
## 11	0.000000000	0.009433611	-0.4149946	1.372975e-05	0.000000e+00
## 12	0.000000000	0.010526295	-0.4203554	1.236560e-05	0.000000e+00
## 13	0.000000000	0.011913307	-0.4287123	1.053257e-05	0.000000e+00
## 14	0.000000000	0.012860711	-0.4313347	1.002596e-05	0.000000e+00
## 15	0.000000000	0.013283016	-0.4326789	9.641978e-06	0.000000e+00
## 16	0.000000000	0.016199077	-0.4354363	1.007761e-05	0.000000e+00
## 17	0.000000000	0.017649942	-0.4368068	1.068202e-05	7.677794e-10
## 18	0.01188347	0.019705278	-0.4393990	1.185902e-05	1.876286e-09
## 19	0.02321738	0.017723575	-0.4389777	1.081630e-05	0.000000e+00
## 20	0.02623053	0.017133681	-0.4387891	1.078337e-05	0.000000e+00
## 21	0.02631770	0.017119181	-0.4387868	1.077247e-05	-2.030679e-11
## 22	0.02799718	0.016890803	-0.4388035	1.056532e-05	-3.110455e-10
## 23	0.03363246	0.015902530	-0.4385905	9.858124e-06	-1.724175e-09
##	thinness..1.19.years	thinness.5.9.years	Income.composition.of.resources		
## 0	0.000000000	0.000000000	0.000000		
## 1	0.000000000	0.000000000	0.000000		
## 2	0.000000000	0.000000000	1.361789		

```
## 3      0.000000000      0.000000000      5.065665
## 4      0.000000000      0.000000000      9.684905
## 5      0.000000000      0.000000000     10.208501
## 6      0.000000000      0.000000000     10.447586
## 7      0.000000000     -0.007647153     10.500950
## 8      0.000000000     -0.009862181     10.519472
## 9      0.000000000     -0.015354612     10.514569
## 10     0.000000000     -0.031216859     10.477604
## 11     0.000000000     -0.035232076     10.460145
## 12     -0.003686124     -0.035327440     10.442942
## 13     -0.006269579     -0.032955248     10.467012
## 14     -0.011293011     -0.032821529     10.625859
## 15     -0.013575342     -0.032039669     10.699948
## 16     -0.016194088     -0.030454393     10.810363
## 17     -0.018745256     -0.029476938     10.872040
## 18     -0.022459133     -0.027415357     10.966279
## 19     -0.017743114     -0.041434424     10.694726
## 20     -0.017242693     -0.044819941     10.628122
## 21     -0.017197546     -0.044931794     10.625967
## 22     -0.016470887     -0.046195684     10.593423
## 23     -0.013409425     -0.054316961     10.445188
##      Schooling  Developing
## 0  0.000000000  0.000000000
## 1  0.002652275  0.000000000
## 2  0.089308465  0.000000000
## 3  0.375602843  0.000000000
## 4  0.811537097  0.000000000
## 5  0.847559456  0.000000000
## 6  0.858130544  0.000000000
## 7  0.859416124  0.000000000
## 8  0.859926272  0.000000000
## 9  0.857069435  0.000000000
## 10 0.845149516  0.000000000
## 11 0.839532849 -0.06585915
## 12 0.834635032 -0.11855471
## 13 0.825369173 -0.20029570
## 14 0.835606260 -0.40270701
## 15 0.840750630 -0.50663931
## 16 0.849590748 -0.70995595
## 17 0.851879851 -0.82708722
## 18 0.854237023 -0.99228196
## 19 0.852547597 -1.07124972
## 20 0.850537862 -1.09326016
## 21 0.850543248 -1.09385558
## 22 0.850644785 -1.10466522
## 23 0.850995134 -1.14381554
## attr(,"scaled:scale")
## [1] 4.243627e+03 3.835913e+03 1.382431e+02 6.262055e+04 8.853754e+02
## [6] 3.373326e+05 6.721794e+02 5.156200e+03 7.601753e+02 7.866082e+01
## [11] 7.126279e+02 2.029775e+02 4.042208e+05 2.542815e+09 1.533606e+02
## [16] 1.547321e+02 6.174032e+00 9.702399e+01 1.203980e+01
```

```
lmod_la<-glm(Life.expectancy ~.,family='gaussian',data_tr)
sum_la<-summary(lmod_la)
```

```
sum_la
```

```
##
## Call:
## glm(formula = Life.expectancy ~ ., family = "gaussian", data = data_tr)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -12.704   -2.164    0.010    2.225   11.494
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.461e+01  1.021e+00  53.507 < 2e-16 ***
## Adult.Mortality  -1.594e-02  1.137e-03 -14.021 < 2e-16 ***
## infant.deaths     1.022e-01  1.487e-02   6.875 1.02e-11 ***
## Alcohol          -1.207e-01  3.887e-02  -3.105  0.00195 **
## percentage.expenditure  3.241e-04  2.169e-04   1.494  0.13533
## Hepatitis.B       -1.068e-02  5.205e-03  -2.052  0.04038 *
## Measles           -1.038e-05  1.317e-05  -0.788  0.43066
## BMI               3.401e-02  7.083e-03   4.802 1.78e-06 ***
## under.five.deaths -7.607e-02  1.071e-02  -7.102 2.16e-12 ***
## Polio             1.222e-02  6.156e-03   1.986  0.04731 *
## Total.expenditure  3.363e-02  4.801e-02   0.700  0.48377
## Diphtheria        1.590e-02  7.184e-03   2.214  0.02705 *
## HIV.AIDS          -4.386e-01  2.158e-02 -20.325 < 2e-16 ***
## GDP               9.858e-06  3.440e-05   0.287  0.77450
## Population        -1.724e-09  2.116e-09  -0.815  0.41542
## thinness..1.19.years -1.341e-02  5.635e-02  -0.238  0.81193
## thinness.5.9.years  -5.432e-02  5.579e-02  -0.974  0.33043
## Income.composition.of.resources 1.045e+01  1.015e+00  10.293 < 2e-16 ***
## Schooling         8.510e-01  6.999e-02  12.159 < 2e-16 ***
## Developing        -1.144e+00  4.059e-01  -2.818  0.00491 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 12.78774)
##
##      Null deviance: 87911  on 1153  degrees of freedom
## Residual deviance: 14501  on 1134  degrees of freedom
## AIC: 6237.7
##
## Number of Fisher Scoring iterations: 2

shapiro.test(lmod_la$residuals)

##
## Shapiro-Wilk normality test
##
## data:  lmod_la$residuals
## W = 0.99402, p-value = 0.0001406
```

```
dwtest(lmod_la)
```

```
##
## Durbin-Watson test
##
## data: lmod_la
## DW = 2.0436, p-value = 0.7729
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bptest(lmod_la)
```

```
##
## studentized Breusch-Pagan test
##
## data: lmod_la
## BP = 125.61, df = 19, p-value < 2.2e-16
```

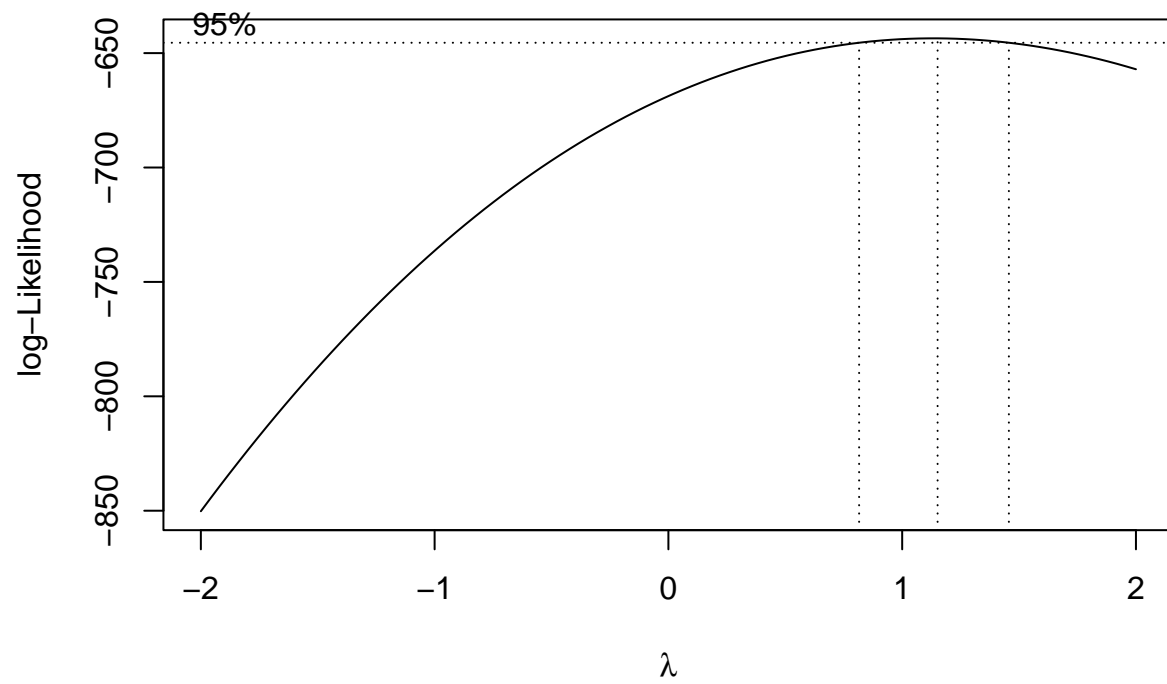
```
vif(lmod_la)
```

```
##           Adult.Mortality           infant.deaths
##           1.819167           254.411298
##           Alcohol           percentage.expenditure
##           2.258261           14.421237
##           Hepatitis.B           Measles
##           1.660452           1.542678
##           BMI           under.five.deaths
##           1.772660           238.492903
##           Polio           Total.expenditure
##           1.712429           1.115458
##           Diphtheria           HIV.AIDS
##           2.049503           1.500308
##           GDP           Population
##           15.121553           2.264707
##           thinness..1.19.years           thinness.5.9.years
##           5.839108           5.826642
## Income.composition.of.resources           Schooling
##           3.069910           3.606260
##           Developing
##           1.867234
```

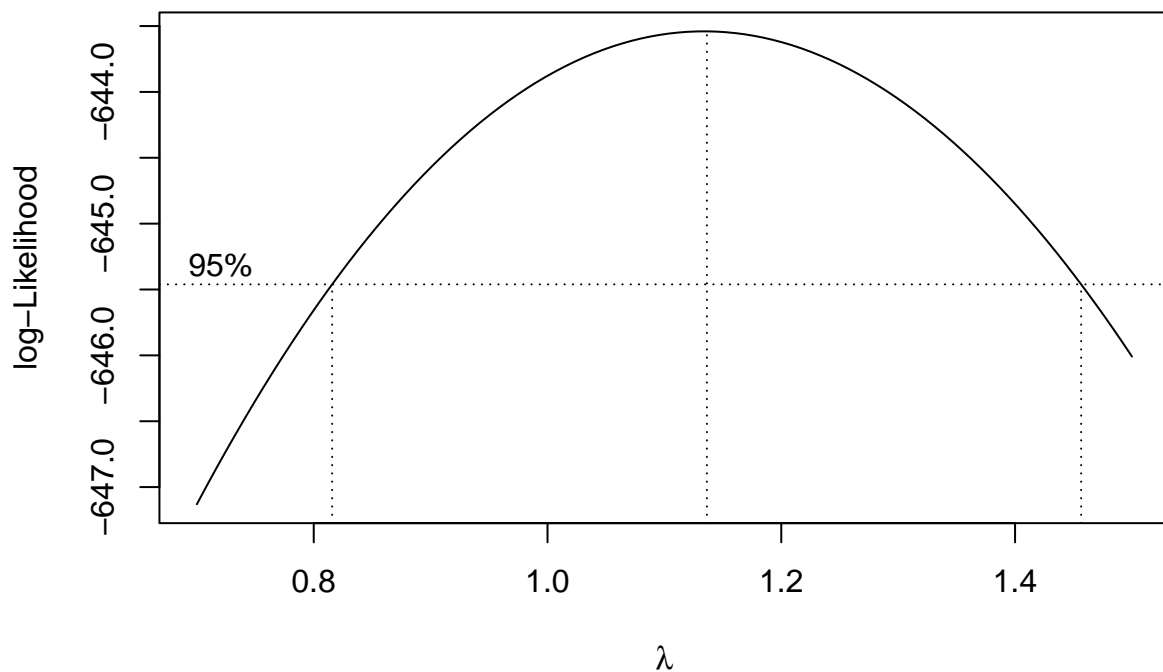
Transformation

```
boxcox(model, plotit=T)
```





```
b<-boxcox(model, plotit=T, lambda=seq(0.7, 1.5,by=0.05))
```



b

```
## $x
## [1] 0.7000000 0.7080808 0.7161616 0.7242424 0.7323232 0.7404040 0.7484848
## [8] 0.7565657 0.7646465 0.7727273 0.7808081 0.7888889 0.7969697 0.8050505
## [15] 0.8131313 0.8212121 0.8292929 0.8373737 0.8454545 0.8535354 0.8616162
## [22] 0.8696970 0.8777778 0.8858586 0.8939394 0.9020202 0.9101010 0.9181818
## [29] 0.9262626 0.9343434 0.9424242 0.9505051 0.9585859 0.9666667 0.9747475
## [36] 0.9828283 0.9909091 0.9989899 1.0070707 1.0151515 1.0232323 1.0313131
## [43] 1.0393939 1.0474747 1.0555556 1.0636364 1.0717172 1.0797980 1.0878788
## [50] 1.0959596 1.1040404 1.1121212 1.1202020 1.1282828 1.1363636 1.1444444
## [57] 1.1525253 1.1606061 1.1686869 1.1767677 1.1848485 1.1929293 1.2010101
## [64] 1.2090909 1.2171717 1.2252525 1.2333333 1.2414141 1.2494949 1.2575758
## [71] 1.2656566 1.2737374 1.2818182 1.2898990 1.2979798 1.3060606 1.3141414
## [78] 1.3222222 1.3303030 1.3383838 1.3464646 1.3545455 1.3626263 1.3707071
## [85] 1.3787879 1.3868687 1.3949495 1.4030303 1.4111111 1.4191919 1.4272727
## [92] 1.4353535 1.4434343 1.4515152 1.4595960 1.4676768 1.4757576 1.4838384
## [99] 1.4919192 1.5000000
##
## $y
## [1] -647.1302 -646.9966 -646.8655 -646.7370 -646.6111 -646.4877 -646.3669
## [8] -646.2487 -646.1330 -646.0199 -645.9094 -645.8014 -645.6960 -645.5931
## [15] -645.4928 -645.3950 -645.2998 -645.2071 -645.1169 -645.0293 -644.9442
## [22] -644.8617 -644.7816 -644.7041 -644.6291 -644.5567 -644.4867 -644.4193
## [29] -644.3544 -644.2920 -644.2321 -644.1747 -644.1198 -644.0674 -644.0175
## [36] -643.9701 -643.9252 -643.8828 -643.8429 -643.8054 -643.7705 -643.7380
```

```
## [43] -643.7080 -643.6805 -643.6554 -643.6328 -643.6127 -643.5950 -643.5798
## [50] -643.5671 -643.5568 -643.5490 -643.5436 -643.5407 -643.5402 -643.5422
## [57] -643.5466 -643.5534 -643.5627 -643.5744 -643.5885 -643.6051 -643.6241
## [64] -643.6455 -643.6693 -643.6956 -643.7243 -643.7553 -643.7888 -643.8247
## [71] -643.8630 -643.9037 -643.9468 -643.9923 -644.0402 -644.0905 -644.1432
## [78] -644.1983 -644.2557 -644.3156 -644.3778 -644.4424 -644.5093 -644.5787
## [85] -644.6504 -644.7245 -644.8009 -644.8797 -644.9608 -645.0443 -645.1302
## [92] -645.2184 -645.3090 -645.4019 -645.4972 -645.5948 -645.6947 -645.7970
## [99] -645.9016 -646.0085
```

```
I=which(b$y==max(b$y))
b$x[I]
```

```
## [1] 1.136364
```

```
lmod_trans<-lm(Life.expectancy ~(1.14) ~ Adult.Mortality + infant.deaths + Alcohol +
percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +
Polio + Diphtheria + HIV.AIDS + thinness.5.9.years +
Income.composition.of.resources + Schooling + Developing,
data = data_tr)
summary(lmod_trans)
```

```
##
## Call:
## lm(formula = Life.expectancy^(1.14) ~ Adult.Mortality + infant.deaths +
##     Alcohol + percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +
##     Polio + Diphtheria + HIV.AIDS + thinness.5.9.years + Income.composition.of.resources +
##     Schooling + Developing, data = data_tr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.3506  -4.5555   0.0295   4.6070  24.5090
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      95.7448112   2.0095217  47.646 < 2e-16 ***
## Adult.Mortality  -0.0328397   0.0023328 -14.077 < 2e-16 ***
## infant.deaths     0.1908090   0.0270811   7.046 3.19e-12 ***
## Alcohol          -0.2502517   0.0797404  -3.138 0.00174 **
## percentage.expenditure  0.0008192   0.0001400   5.850 6.41e-09 ***
## Hepatitis.B      -0.0216771   0.0106833  -2.029 0.04268 *
## BMI              0.0705189   0.0144545   4.879 1.22e-06 ***
## under.five.deaths -0.1439850   0.0201860  -7.133 1.74e-12 ***
## Polio            0.0247558   0.0126299   1.960 0.05023 .
## Diphtheria       0.0327268   0.0147596   2.217 0.02680 *
## HIV.AIDS         -0.8758899   0.0440881 -19.867 < 2e-16 ***
## thinness.5.9.years -0.1393322   0.0654453  -2.129 0.03347 *
## Income.composition.of.resources 21.5970832   2.0789783  10.388 < 2e-16 ***
## Schooling        1.7550369   0.1425695  12.310 < 2e-16 ***
## Developing      -2.5148220   0.8338894  -3.016 0.00262 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 7.361 on 1139 degrees of freedom
## Multiple R-squared:  0.8332, Adjusted R-squared:  0.8311
## F-statistic: 406.4 on 14 and 1139 DF,  p-value: < 2.2e-16
```

```
shapiro.test(lmod_trans$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lmod_trans$residuals
## W = 0.99451, p-value = 0.0003153
```

```
bptest(lmod_trans)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lmod_trans
## BP = 113.79, df = 14, p-value < 2.2e-16
```

```
confint(lmod_trans)
```

	2.5 %	97.5 %
## (Intercept)	9.180203e+01	99.6875911695
## Adult.Mortality	-3.741673e-02	-0.0282625998
## infant.deaths	1.376746e-01	0.2439434815
## Alcohol	-4.067062e-01	-0.0937971872
## percentage.expenditure	5.444340e-04	0.0010939025
## Hepatitis.B	-4.263817e-02	-0.0007159645
## BMI	4.215856e-02	0.0988793185
## under.five.deaths	-1.835909e-01	-0.1043791674
## Polio	-2.471807e-05	0.0495363565
## Diphtheria	3.767671e-03	0.0616859297
## HIV.AIDS	-9.623929e-01	-0.7893868858
## thinness.5.9.years	-2.677391e-01	-0.0109252171
## Income.composition.of.resources	1.751803e+01	25.6761404526
## Schooling	1.475309e+00	2.0347652050
## Developing	-4.150954e+00	-0.8786902295

```
vif(lmod_trans)
```

##	Adult.Mortality	infant.deaths
##	1.808486	199.139250
##	Alcohol	percentage.expenditure
##	2.242487	1.418817
##	Hepatitis.B	BMI
##	1.651012	1.742053
##	under.five.deaths	Polio
##	199.915027	1.701046
##	Diphtheria	HIV.AIDS
##	2.041567	1.477832

```
##          thinness.5.9.years Income.composition.of.resources
##          1.892364          3.040356
##          Schooling          Developing
##          3.531003          1.860123
```

Prediction

```
# AIC model contain the highest R ^ Square.
result<-predict(lmod_AIC_B,newdata = data_te,interval='prediction') # Prediction
error <- sum((data_te$Life.expectancy - result)^2)
error
```

```
## [1] 68964.6
```

```
result2<-predict(lmod_BIC_B0,newdata = data_te,interval='prediction') # Prediction
error2 <- sum((data_te$Life.expectancy - result2)^2)
error2
```

```
## [1] 72927.3
```

```
result3<-predict(lmod_trans,newdata = data_te,interval='prediction') # Prediction
error3 <- sum((data_te$Life.expectancy - result3)^2)
error3
```

```
## [1] 4971339
```

```
result4<-predict(lmod_la,newdata = data_te,interval='prediction') # Prediction
error4 <- sum((data_te$Life.expectancy - result4)^2)
error4
```

```
## [1] 6552.674
```

```
which.min(c(error, error2, error3, error4)) # Check which method provide lowest errors.
```

```
## [1] 4
```

```
x <- model.matrix(lmod_la)
x0 <- apply(x,2,median) # get median characteristics
x0
```

```
##          (Intercept)          Adult.Mortality
##          1.0000          147.0000
##          infant.deaths          Alcohol
##          3.0000          3.9550
##          percentage.expenditure          Hepatitis.B
##          155.4500          91.0000
##          Measles          BMI
##          14.0000          43.5000
##          under.five.deaths          Polio
##          4.0000          93.0000
```

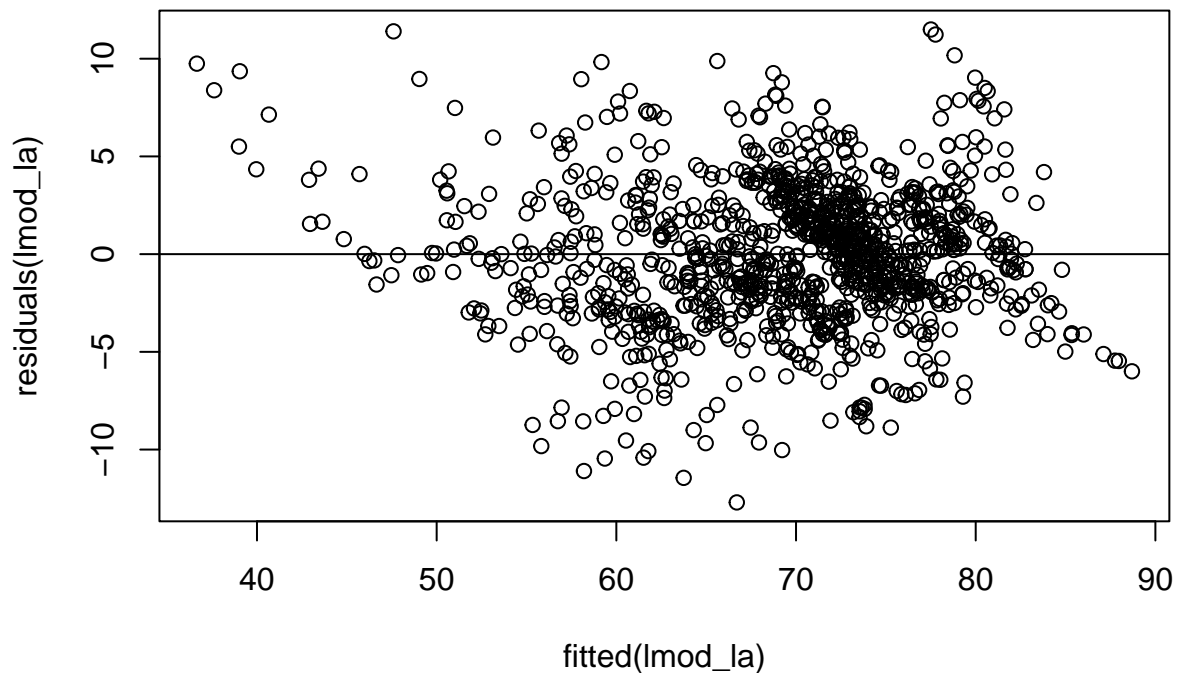
```
##          Total.expenditure          Diphtheria
##              5.8250              93.0000
##          HIV.AIDS              GDP
##              0.1000              1677.7636
##          Population      thinness..1.19.years
##          1420384.0000              2.9000
##          thinness.5.9.years Income.composition.of.resources
##              3.1000              0.6755
##          Schooling      Developing
##          12.3500              1.0000
```

```
predict(lmod_la, newdata = data.frame(t(x0)), interval = "prediction")
```

```
##          1
## 71.34512
```

```
#sqrt(pred1)
```

```
plot(residuals(lmod_la)~fitted(lmod_la))
abline(h = 0)
```



```
qqnorm(residuals(lmod_la))
qqline(residuals(lmod_la))
```

Normal Q-Q Plot

