team10

2022-12-05

```
library(ggplot2)
library(moderndive)
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
```

```
library(readr)
library(faraway)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v tibble  3.1.8     v stringr 1.4.1
## v tidyr   1.2.1     v forcats 0.5.2
## v purrr   0.3.5
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(olsrr)
```

```
##
## Attaching package: 'olsrr'
##
## The following object is masked from 'package:faraway':
##
##     hsb
##
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
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
```

Cleaning Data set:

```
lifeExpectancyData1 <- read_csv("Life Expectancy Data.csv")
```

```
## Rows: 2938 Columns: 22
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): Country, Status
## dbl (20): Year, Life expectancy, Adult Mortality, infant deaths, Alcohol, pe...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
lifeExpectancyData <- lifeExpectancyData1 %>%
  rename(Life.expectancy = "Life expectancy") %>%
  rename(Adult.Mortality = "Adult Mortality") %>%
  rename(infant.deaths = "infant deaths") %>%
  rename(percentage.expenditure = "percentage expenditure") %>%
  rename(Measles = "Measles") %>%
  rename(under.five.deaths = "under-five deaths" ) %>%
  rename(Polio = "Polio")%>%
  rename(Diphtheria = "Diphtheria" )%>%
  rename(HIV.AIDS = "HIV/AIDS")
```

```
lifeExpectancyDataUpdated <- lifeExpectancyData[which(grepl(pattern="2015", lifeExpectancyData$Year)),]
```

After looking at the dataset, only 8 of the variables had values for all the countries (some of the columns had NA entries):

```
lmod <- lm(Life.expectancy~Adult.Mortality+infant.deaths+percentage.expenditure+ Measles+under.five.dea
summary(lmod)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Adult.Mortality + infant.deaths +
##     percentage.expenditure + Measles + under.five.deaths + Polio +
##     Diphtheria + HIV.AIDS, data = lifeExpectancyDataUpdated)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.8445  -2.5420   0.3501   3.3574  10.8480
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             7.120e+01  1.809e+00  39.356  < 2e-16 ***
## Adult.Mortality        -4.881e-02  4.517e-03 -10.806  < 2e-16 ***
## infant.deaths           6.752e-02  4.644e-02   1.454  0.14780
## percentage.expenditure  2.410e-03  1.220e-02   0.198  0.84358
## Measles                 4.743e-06  8.175e-05   0.058  0.95380
## under.five.deaths      -5.958e-02  3.380e-02  -1.763  0.07972 .
## Polio                   4.352e-02  1.892e-02   2.300  0.02261 *
## Diphtheria              5.775e-02  2.091e-02   2.762  0.00636 **
## HIV.AIDS               -5.543e-01  3.381e-01  -1.639  0.10296
```

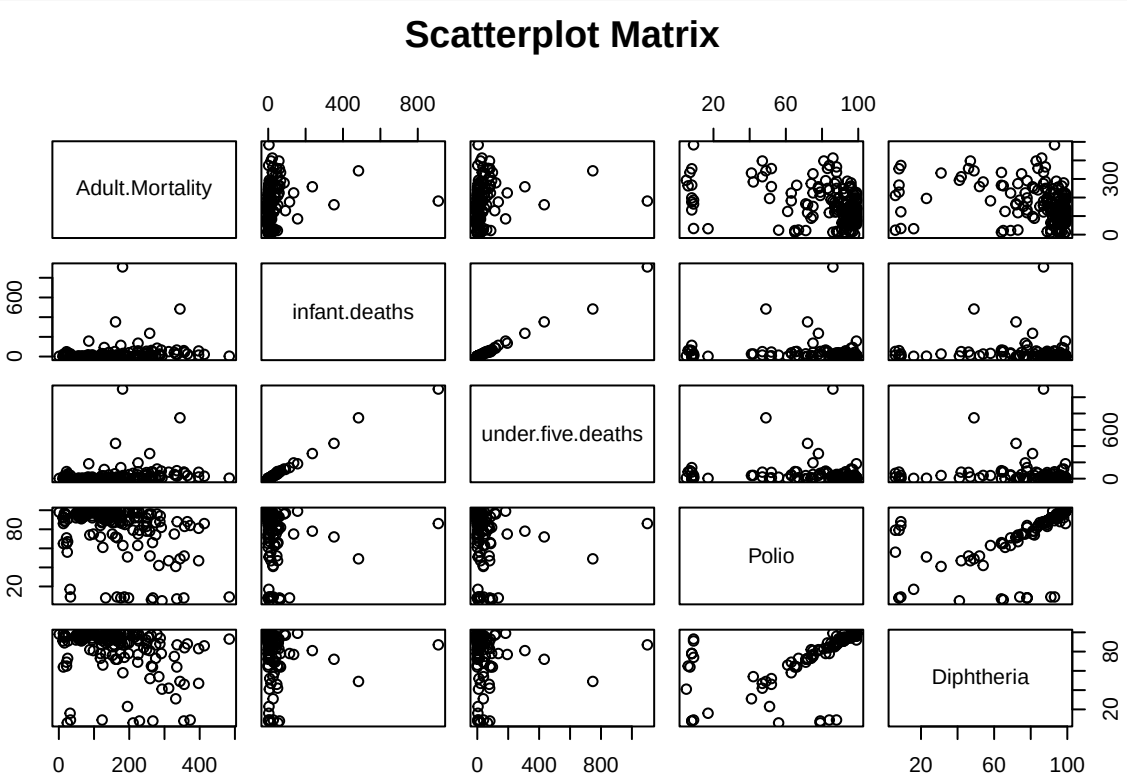
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.509 on 174 degrees of freedom
## Multiple R-squared:  0.7054, Adjusted R-squared:  0.6919
## F-statistic: 52.08 on 8 and 174 DF,  p-value: < 2.2e-16
```

```
anova(lmod)
```

```
## Analysis of Variance Table
##
## Response: Life.expectancy
##                          Df Sum Sq Mean Sq  F value    Pr(>F)
## Adult.Mortality           1 7291.0  7291.0 358.5390 < 2.2e-16 ***
## infant.deaths             1  111.6   111.6   5.4863  0.020297 *
## percentage.expenditure    1    1.2     1.2   0.0581  0.809784
## Measles                   1   63.7    63.7   3.1338  0.078436 .
## under.five.deaths         1  128.2   128.2   6.3030  0.012966 *
## Polio                     1  612.0   612.0  30.0950 1.432e-07 ***
## Diphtheria                1  210.4   210.4  10.3448  0.001548 **
## HIV.AIDS                  1   54.6    54.6   2.6873  0.102956
## Residuals               174 3538.3    20.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the ANOVA table, we see that Adult Mortality, infant deaths, under-five deaths, polio, diphtheria have significant p values

```
pairs(~ Adult.Mortality + infant.deaths + under.five.deaths + Polio + Diphtheria, data=lifeExpectancyDa
```




```
lifeExpectancyDataUpdated2 <- lifeExpectancyDataUpdated[c(-1,-2,-3,-7,-8,-9, -10, -11, -14, -16, -17, -
print(cor(lifeExpectancyDataUpdated2[,2:5]),digits=4)
```

```
##                   Adult.Mortality infant.deaths under.five.deaths   Polio
## Adult.Mortality            1.0000        0.1882            0.2120 -0.3724
## infant.deaths              0.1882        1.0000            0.9942 -0.1297
## under.five.deaths          0.2120        0.9942            1.0000 -0.1475
## Polio                     -0.3724       -0.1297           -0.1475  1.0000
```

Performing all possible, forward, and backward regressions

```
ourmodel <- lm(Life.expectancy~Adult.Mortality+infant.deaths+under.five.deaths+Polio+Diphtheria, data =
anova(ourmodel)
```

```
## Analysis of Variance Table
##
## Response: Life.expectancy
##                    Df Sum Sq Mean Sq  F value    Pr(>F)
## Adult.Mortality     1 7291.0  7291.0 359.0728 < 2.2e-16 ***
## infant.deaths       1  111.6   111.6   5.4945  0.020187 *
## under.five.deaths   1  188.5   188.5   9.2836  0.002666 **
## Polio               1  613.1   613.1  30.1957 1.344e-07 ***
## Diphtheria          1  212.8   212.8  10.4802  0.001441 **
## Residuals         177 3594.0    20.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(ourmodel)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Adult.Mortality + infant.deaths +
##     under.five.deaths + Polio + Diphtheria, data = lifeExpectancyDataUpdated2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3307  -2.5620   0.5472   3.1795  10.7421
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       70.656344   1.775388  39.798  < 2e-16 ***
## Adult.Mortality   -0.052790   0.003815 -13.837  < 2e-16 ***
## infant.deaths      0.076029   0.038280   1.986  0.04856 *
## under.five.deaths -0.065702   0.029900  -2.197  0.02929 *
## Polio              0.044868   0.018868   2.378  0.01847 *
## Diphtheria         0.065743   0.020308   3.237  0.00144 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.506 on 177 degrees of freedom
## Multiple R-squared:  0.7008, Adjusted R-squared:  0.6923
## F-statistic: 82.91 on 5 and 177 DF,  p-value: < 2.2e-16
```

```
k <- ols_step_all_possible(ourmodel)
k
```

```
##    Index N                                                  Predictors
## 1      1 1                                           Adult.Mortality
```

```
## 4      2 1                                                                  Polio
## 5      3 1                                                              Diphtheria
## 3      4 1                                                         under.five.deaths
## 2      5 1                                                             infant.deaths
## 9      6 2                                                 Adult.Mortality Diphtheria
## 8      7 2                                                      Adult.Mortality Polio
## 7      8 2                                          Adult.Mortality under.five.deaths
## 6      9 2                                              Adult.Mortality infant.deaths
## 15    10 2                                                        Polio Diphtheria
## 13    11 2                                                   under.five.deaths Polio
## 11    12 2                                                       infant.deaths Polio
## 14    13 2                                           under.five.deaths Diphtheria
## 12    14 2                                               infant.deaths Diphtheria
## 10    15 2                                         infant.deaths under.five.deaths
## 21    16 3                                       Adult.Mortality Polio Diphtheria
## 20    17 3                          Adult.Mortality under.five.deaths Diphtheria
## 18    18 3                              Adult.Mortality infant.deaths Diphtheria
## 19    19 3                               Adult.Mortality under.five.deaths Polio
## 17    20 3                                   Adult.Mortality infant.deaths Polio
## 16    21 3                        Adult.Mortality infant.deaths under.five.deaths
## 25    22 3                                     under.five.deaths Polio Diphtheria
## 22    23 3                                infant.deaths under.five.deaths Polio
## 24    24 3                                     infant.deaths Polio Diphtheria
## 23    25 3                           infant.deaths under.five.deaths Diphtheria
## 29    26 4                     Adult.Mortality under.five.deaths Polio Diphtheria
## 28    27 4                         Adult.Mortality infant.deaths Polio Diphtheria
## 27    28 4             Adult.Mortality infant.deaths under.five.deaths Diphtheria
## 26    29 4                  Adult.Mortality infant.deaths under.five.deaths Polio
## 30    30 4                       infant.deaths under.five.deaths Polio Diphtheria
## 31    31 5 Adult.Mortality infant.deaths under.five.deaths Polio Diphtheria
##      R-Square Adj. R-Square Mallow's Cp
## 1  0.60702715    0.60485603   53.453978
## 4  0.26584521    0.26178911  255.272250
## 5  0.25030006    0.24615807  264.467625
## 3  0.07394106    0.06882471  368.788705
## 2  0.05822064    0.05301744  378.087750
## 9  0.67617183    0.67257374   14.553047
## 8  0.66604570    0.66233510   20.542926
## 7  0.61895787    0.61472407   48.396640
## 6  0.61631578    0.61205263   49.959505
## 15 0.31090042    0.30324376  230.620886
## 13 0.30507170    0.29735027  234.068730
## 11 0.29678159    0.28896805  238.972552
## 14 0.29228605    0.28442256  241.631787
## 12 0.28398757    0.27603188  246.540558
## 10 0.14686529    0.13738602  327.652061
## 21 0.68671445    0.68146386   10.316804
## 20 0.68422466    0.67893234   11.789582
## 18 0.68262532    0.67730619   12.735638
## 19 0.67435180    0.66889401   17.629646
## 17 0.67255075    0.66706278   18.695013
## 16 0.63201008    0.62584265   42.675907
## 25 0.34579411    0.33482977  211.980329
## 22 0.34360456    0.33260352  213.275508
```

```
## 24 0.33874374    0.32766123  216.150815
## 23 0.32818330    0.31692380  222.397600
## 29 0.69410563    0.68723160    7.944727
## 28 0.69261147    0.68570387    8.828558
## 27 0.69121458    0.68427558    9.654857
## 26 0.68305720    0.67593489   14.480168
## 30 0.37707642    0.36307813  195.476008
## 31 0.70077435    0.69232164    6.000000
```

forward

```
f <- ols_step_forward_p(ourmodel, penter= 0.3, details = TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Adult.Mortality
## 2. infant.deaths
## 3. under.five.deaths
## 4. Polio
## 5. Diphtheria
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Adult.Mortality
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.779       RMSE                 5.107
## R-Squared               0.607       Coef. Var            7.130
## Adj. R-Squared          0.605       MSE                 26.077
## Pred R-Squared          0.596       MAE                  3.581
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------
## Regression     7291.014          1       7291.014    279.592    0.0000
## Residual       4720.004        181         26.077
## Total         12011.017        182
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------
##           model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
```

```
## ----------------------------------------------------------------------------------------------
##     (Intercept)    81.534         0.703                115.974    0.000    80.147    82.921
## Adult.Mortality    -0.065         0.004       -0.779    -16.721    0.000    -0.073    -0.057
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - Diphtheria
##
##                          Model Summary
## ------------------------------------------------------------
## R                        0.822       RMSE                4.648
## R-Squared                0.676       Coef. Var           6.491
## Adj. R-Squared           0.673       MSE                21.608
## Pred R-Squared           0.656       MAE                 3.473
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     8121.512          2       4060.756    187.925    0.0000
## Residual       3889.506        180         21.608
## Total         12011.017        182
## --------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------
##           model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
##     (Intercept)    71.774         1.699                42.236    0.000    68.421    75.128
## Adult.Mortality    -0.057         0.004       -0.689    -15.386    0.000    -0.065    -0.050
##      Diphtheria     0.102         0.016        0.278      6.200    0.000     0.069     0.134
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 3
##
## - Polio
##
##                          Model Summary
## ------------------------------------------------------------
## R                        0.829       RMSE                4.585
## R-Squared                0.687       Coef. Var           6.402
## Adj. R-Squared           0.681       MSE                21.022
## Pred R-Squared           0.664       MAE                 3.471
## ------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     8248.139          3       2749.380    130.788    0.0000
## Residual       3762.878        179         21.022
## Total         12011.017        182
## --------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##           model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
##     (Intercept)    70.175         1.798                 39.020    0.000    66.626    73.723
## Adult.Mortality    -0.055         0.004       -0.665    -14.654    0.000    -0.063    -0.048
##      Diphtheria     0.071         0.021        0.193      3.436    0.001     0.030     0.111
##           Polio     0.047         0.019        0.141      2.454    0.015     0.009     0.085
## ----------------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 4
##
## - under.five.deaths
##
##                         Model Summary
## ---------------------------------------------------------------
## R                       0.833       RMSE                4.543
## R-Squared               0.694       Coef. Var           6.344
## Adj. R-Squared          0.687       MSE                20.641
## Pred R-Squared          0.662       MAE                 3.412
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     8336.915          4       2084.229    100.975    0.0000
## Residual       3674.103        178         20.641
## Total         12011.017        182
## --------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##             model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
```

```
## ---------------------------------------------------------------------------------------------
##       (Intercept)    70.437         1.787                39.426    0.000    66.912    73.963
##   Adult.Mortality    -0.054         0.004       -0.649   -14.237    0.000    -0.062    -0.047
##        Diphtheria     0.069         0.020        0.189     3.390    0.001     0.029     0.109
##             Polio     0.046         0.019        0.136     2.398    0.018     0.008     0.083
## under.five.deaths    -0.007         0.003       -0.088    -2.074    0.040    -0.013     0.000
## ---------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 5
##
## - infant.deaths
##
##                         Model Summary
## ----------------------------------------------------------------
## R                        0.837       RMSE                4.506
## R-Squared                0.701       Coef. Var           6.292
## Adj. R-Squared           0.692       MSE                20.305
## Pred R-Squared           0.374       MAE                 3.384
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                  Sum of
##                 Squares         DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     8417.013          5       1683.403    82.905    0.0000
## Residual       3594.005        177         20.305
## Total         12011.017        182
## ---------------------------------------------------------------------
##
##                                    Parameter Estimates
## ---------------------------------------------------------------------------------------------
##             model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ---------------------------------------------------------------------------------------------
##       (Intercept)    70.656         1.775                39.798    0.000    67.153    74.160
##   Adult.Mortality    -0.053         0.004       -0.634   -13.837    0.000    -0.060    -0.045
##        Diphtheria     0.066         0.020        0.179     3.237    0.001     0.026     0.106
##             Polio     0.045         0.019        0.134     2.378    0.018     0.008     0.082
## under.five.deaths    -0.066         0.030       -0.872    -2.197    0.029    -0.125    -0.007
##     infant.deaths     0.076         0.038        0.783     1.986    0.049     0.000     0.152
## ---------------------------------------------------------------------------------------------
##
##
##
##
## Variables Entered:
##
## + Adult.Mortality
## + Diphtheria
```

```
## + Polio
## + under.five.deaths
## + infant.deaths
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.837       RMSE                 4.506
## R-Squared               0.701       Coef. Var            6.292
## Adj. R-Squared          0.692       MSE                 20.305
## Pred R-Squared          0.374       MAE                  3.384
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------------
## Regression     8417.013          5       1683.403    82.905    0.0000
## Residual       3594.005        177         20.305
## Total         12011.017        182
## --------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##             model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
##       (Intercept)    70.656         1.775                 39.798    0.000    67.153    74.160
##   Adult.Mortality    -0.053         0.004       -0.634    -13.837    0.000    -0.060    -0.045
##         Diphtheria     0.066         0.020        0.179      3.237    0.001     0.026     0.106
##             Polio      0.045         0.019        0.134      2.378    0.018     0.008     0.082
## under.five.deaths    -0.066         0.030       -0.872     -2.197    0.029    -0.125    -0.007
##     infant.deaths      0.076         0.038        0.783      1.986    0.049     0.000     0.152
## ----------------------------------------------------------------------------------------------------
```

backward

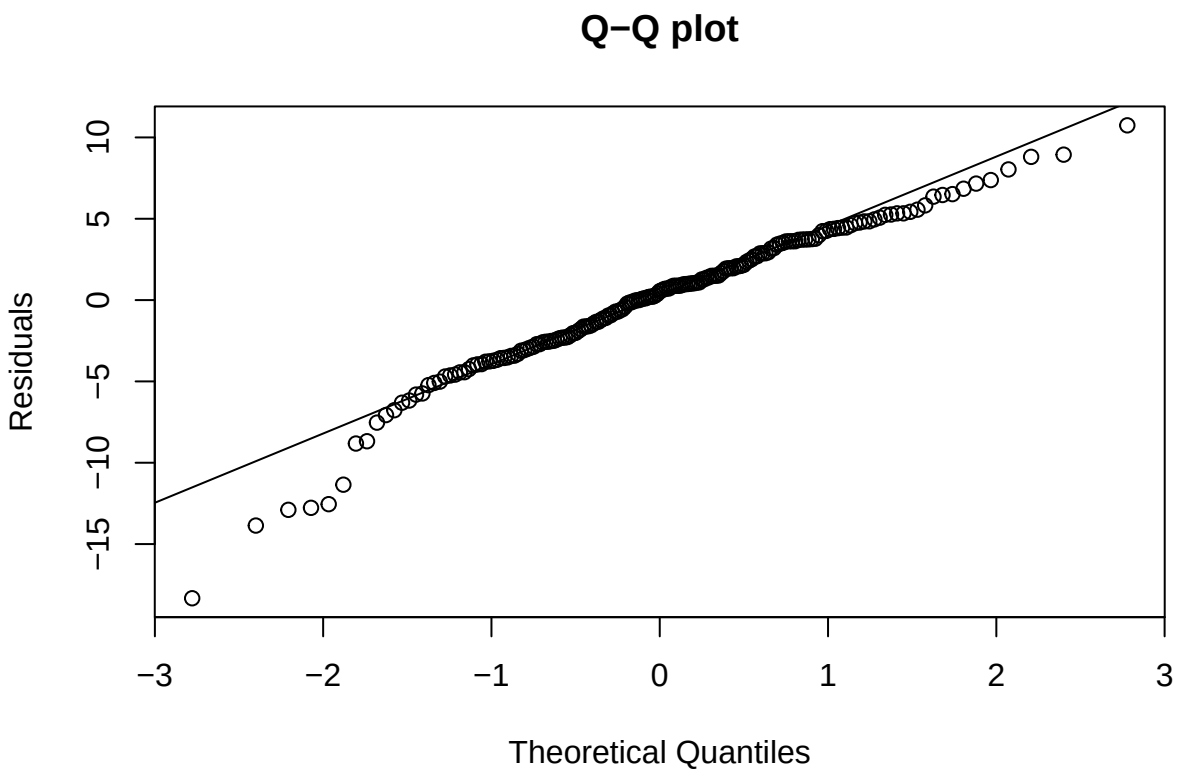
```
b <- ols_step_backward_p(ourmodel, penter = 0.3, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Adult.Mortality
## 2 . infant.deaths
## 3 . under.five.deaths
## 4 . Polio
## 5 . Diphtheria
```

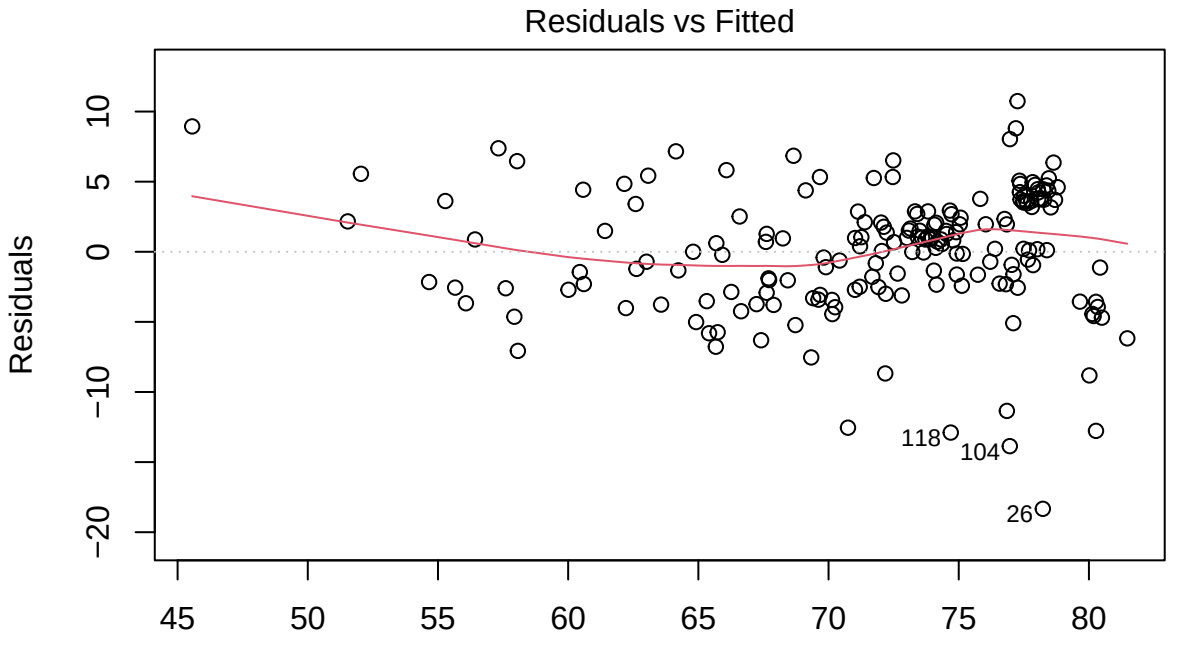
```
##
## We are eliminating variables based on p value...
##
##
## No more variables satisfy the condition of p value = 0.3
##
##
## Variables Removed:
##
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.837       RMSE                 4.506
## R-Squared               0.701       Coef. Var            6.292
## Adj. R-Squared          0.692       MSE                 20.305
## Pred R-Squared          0.374       MAE                  3.384
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     8417.013          5       1683.403    82.905    0.0000
## Residual       3594.005        177         20.305
## Total         12011.017        182
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------
##             model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
##       (Intercept)    70.656         1.775                 39.798    0.000    67.153    74.160
##   Adult.Mortality    -0.053         0.004       -0.634    -13.837    0.000    -0.060    -0.045
##     infant.deaths     0.076         0.038        0.783      1.986    0.049     0.000     0.152
## under.five.deaths    -0.066         0.030       -0.872     -2.197    0.029    -0.125    -0.007
##             Polio     0.045         0.019        0.134      2.378    0.018     0.008     0.082
##        Diphtheria     0.066         0.020        0.179      3.237    0.001     0.026     0.106
## ----------------------------------------------------------------------------------------------
```

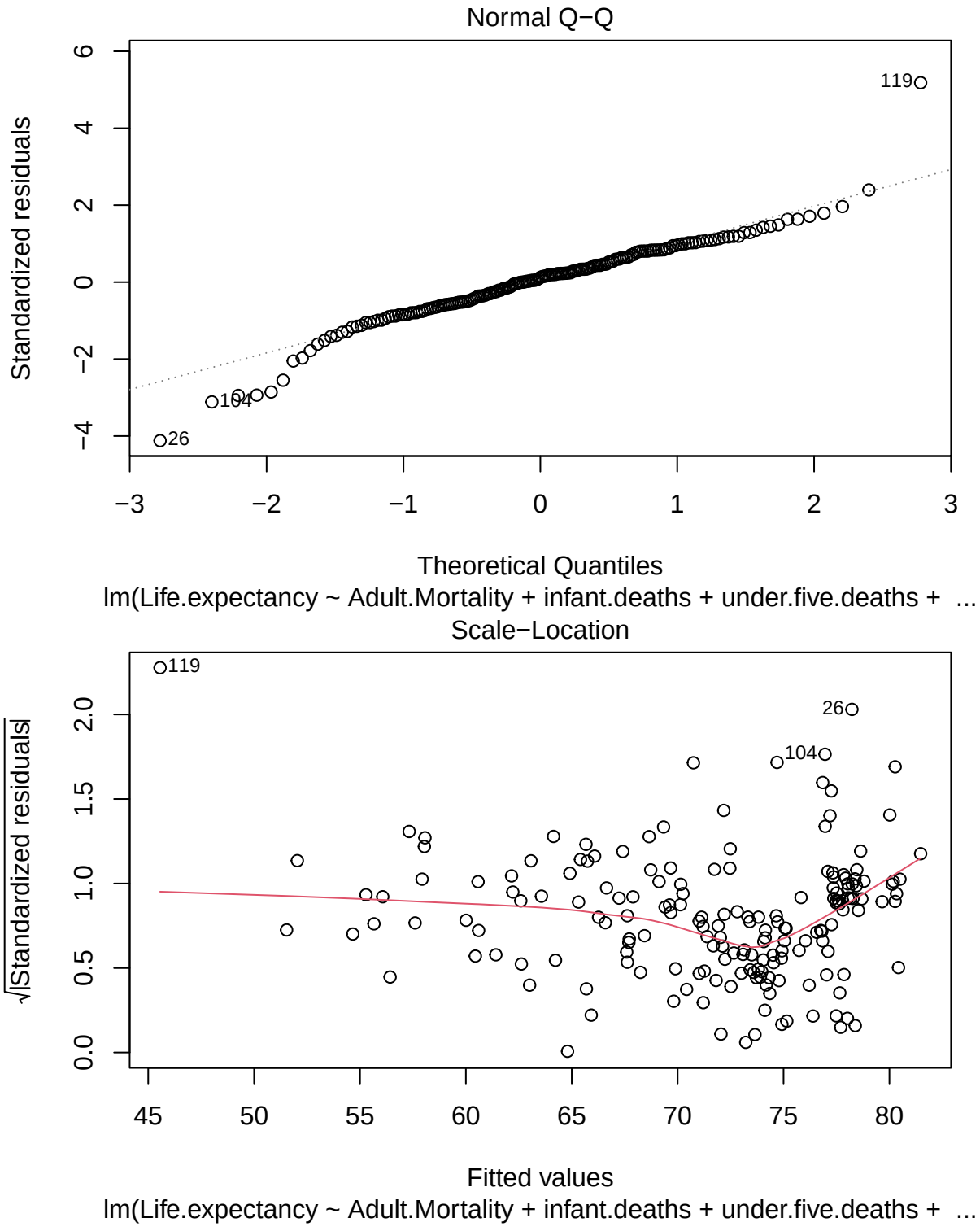
Check model assumptions

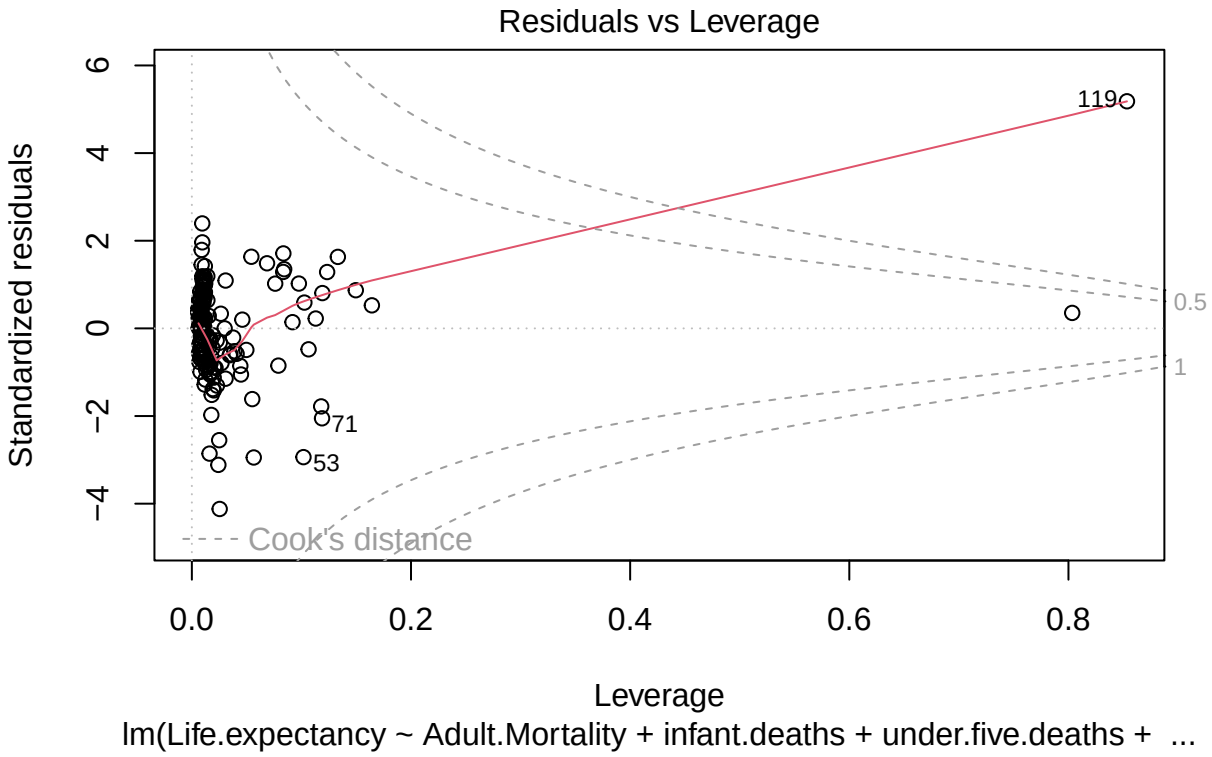
```
qqnorm(residuals(ourmodel),ylab="Residuals",main="Q-Q plot")
qqline(residuals(ourmodel))
```




```
plot(ourmodel)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Life.expectancy ~ Adult.Mortality + infant.deaths + under.five.deaths + ...
119
104
26
Scale−Location
√|Standardized residuals|
Fitted values
lm(Life.expectancy ~ Adult.Mortality + infant.deaths + under.five.deaths + ...



Check for multicollinearity

```
vif(ourmodel)
```

```
##   Adult.Mortality     infant.deaths under.five.deaths             Polio
##          1.241665         91.979835         93.084879          1.875786
##        Diphtheria
##          1.815719
```

Removed infant.deaths to remove multicollinearity? And to obtain final model, also this had higher r squared adj value and minimized Cp (error)

```
finalmod <- lm(Life.expectancy~Adult.Mortality+under.five.deaths+Polio+Diphtheria, data = lifeExpectancy
summary(finalmod)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Adult.Mortality + under.five.deaths +
##     Polio + Diphtheria, data = lifeExpectancyDataUpdated)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.1710  -2.6531   0.5757   3.2032  10.6572
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       70.437108   1.786550  39.426  < 2e-16 ***
## Adult.Mortality   -0.054026   0.003795 -14.237  < 2e-16 ***
## under.five.deaths -0.006654   0.003209  -2.074  0.03953 *
## Polio              0.045606   0.019020   2.398  0.01753 *
## Diphtheria         0.069168   0.020401   3.390  0.00086 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.543 on 178 degrees of freedom
## Multiple R-squared:  0.6941, Adjusted R-squared:  0.6872
## F-statistic:   101 on 4 and 178 DF,  p-value: < 2.2e-16
```

```
anova(finalmod)
```

```
## Analysis of Variance Table
##
## Response: Life.expectancy
##                     Df Sum Sq Mean Sq  F value    Pr(>F)
## Adult.Mortality      1 7291.0  7291.0 353.2292 < 2.2e-16 ***
## under.five.deaths    1  143.3   143.3   6.9425 0.0091594 **
## Polio                1  665.3   665.3  32.2337 5.467e-08 ***
## Diphtheria           1  237.3   237.3  11.4948 0.0008597 ***
## Residuals          178 3674.1    20.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

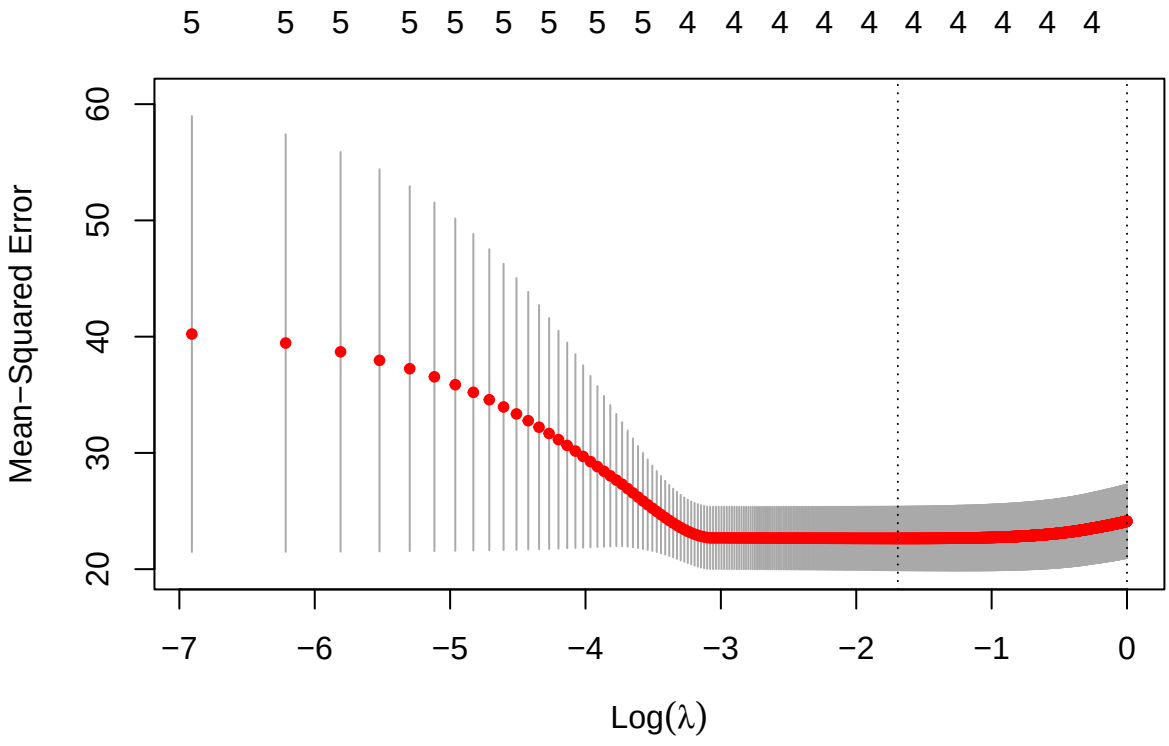
LASSO method

```
x<-(apply(lifeExpectancyDataUpdated2[,-1],2,as.numeric))
y<-(apply(lifeExpectancyDataUpdated2[,1],2,as.numeric))
set.seed(43215)


lasso_cv = cv.glmnet(x, y,family=c("gaussian"),standardize=TRUE,nfolds=10,lambda=seq(0,1,0.001,alpha = 
```

```
## Warning: In seq.default(0, 1, 0.001, alpha = 1) :
##  extra argument 'alpha' will be disregarded
```

```
plot(lasso_cv)
```



```r
(best_lambda = lasso_cv$lambda.min)
```

```
## [1] 0.184
```

```r
best_lasso = glmnet(x,y, standardize=TRUE,alpha = 1, lambda = best_lambda)
coef(best_lasso)
```

```
## 6 x 1 sparse Matrix of class "dgCMatrix"
##                              s0
## (Intercept)       70.804353219
## Adult.Mortality   -0.052988830
## infant.deaths      .
## under.five.deaths -0.005345686
## Polio              0.042803041
## Diphtheria         0.065223126
```

```r
best_lasso$dev.ratio
```

```
## [1] 0.6929939
```