Untitled

Avinandan

25/11/2021

Modelling

```r
#Set the path
setwd("C:/Users/patel/OneDrive/Desktop/0BrandNew0/Modelling")
getwd()
```

```
## [1] "C:/Users/patel/OneDrive/Desktop/0BrandNew0/Modelling"
```

```r
#Import the life expectancy dataset
data <- read.csv("LED.csv")
#Dimension of dataset
dim(data)
```

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

```r
#Variables of dataset
names(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"
```

NA Values Handling

```r
#Replace the NA values by median
data2 = data
sum(is.na(data2))
```

```
## [1] 2563
```

```r
for(i in 1:ncol(data)){
  data[is.na(data[,i]),i] <- median(data[,i] , na.rm = T)
}
sum(is.na(data))
```

```
## [1] 0
```

There are 2563 number of Missing values in our model so removing of these values from our dataset means loosing of 50% of data from our model. So Instead of dropping we replace it by median values on the concern of outliers.

Baseline Model:

Firstly we are making model including all the variables.

```
data3 = data[,-c(1:3)]
attach(data3)
model1 = lm(Life.expectancy ~ .,data3)
summary(model1)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ ., data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.4719  -2.1696  -0.0549   2.3290  16.3588
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     5.466e+01  5.793e-01  94.357  < 2e-16 ***
## Adult.Mortality                -2.039e-02  7.940e-04 -25.676  < 2e-16 ***
## infant.deaths                   9.946e-02  8.484e-03  11.722  < 2e-16 ***
## Alcohol                         1.190e-01  2.416e-02   4.925 8.92e-07 ***
## percentage.expenditure          7.994e-05  9.057e-05   0.883 0.377483
## Hepatitis.B                    -1.526e-02  3.735e-03  -4.086 4.50e-05 ***
## Measles                        -2.050e-05  7.694e-06  -2.665 0.007749 **
## BMI                             4.394e-02  4.944e-03   8.888  < 2e-16 ***
## under.five.deaths              -7.433e-02  6.217e-03 -11.956  < 2e-16 ***
## Polio                           2.872e-02  4.478e-03   6.415 1.64e-10 ***
## Total.expenditure               9.361e-02  3.410e-02   2.745 0.006081 **
## Diphtheria                      4.025e-02  4.665e-03   8.628  < 2e-16 ***
## HIV.AIDS                       -4.718e-01  1.765e-02 -26.723  < 2e-16 ***
## GDP                             4.681e-05  1.381e-05   3.391 0.000707 ***
## Population                     -7.820e-11  1.698e-09  -0.046 0.963277
## thinness..1.19.years           -8.458e-02  5.063e-02  -1.671 0.094903 .
## thinness.5.9.years              3.579e-03  4.990e-02   0.072 0.942827
## Income.composition.of.resources  5.805e+00  6.331e-01   9.168  < 2e-16 ***
## Schooling                       6.720e-01  4.179e-02  16.079  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.074 on 2919 degrees of freedom
## Multiple R-squared:  0.8176, Adjusted R-squared:  0.8165
## F-statistic:   727 on 18 and 2919 DF,  p-value: < 2.2e-16
```

Diagnostic Check of model1

```
library(ggplot2)
library(ggfortify)
autoplot(model1)
```

Clearly from the above summary of model1 we can conclude that the following parameters such as percentage expenditure, population, thinness(1-19 years), thinness(5-9 years), income composition of resources are insignificant(p-value of t-test is more than 0.05) to estimate our predictor variable.So the above variables are dropped from our model.

MultiCollinearity Check

```
library(carData)
library(car)
vif(model1)
```

```
##                  Adult.Mortality                    infant.deaths
##                         1.718038                       177.181727
##                          Alcohol          percentage.expenditure
##                         1.589377                         5.737567
##                      Hepatitis.B                         Measles
##                         1.305980                         1.377767
##                              BMI                under.five.deaths
##                         1.719067                       176.113622
##                            Polio                Total.expenditure
##                         1.937826                         1.186037
##                       Diphtheria                         HIV.AIDS
##                         2.155516                         1.422386
##                              GDP                       Population
##                         5.965483                         1.489929
##             thinness..1.19.years               thinness.5.9.years
```

```
##                        8.773122                        8.869434
## Income.composition.of.resources                       Schooling
##                        2.985808                        3.296186
```

Clearly variables like infant deaths, under five deaths have high vif values(> 100) that means it's highly correlated with other variables.So it violates the independency of explanatory variables(X) so it will be a wise decision to drop these variables.
So we are going to drop 7 variables percentage expenditure, population, thinness(1-19 years), thinness(5-9 years), income composition of resources, infant deaths, under five deaths as these are coming under the low significance case and very high multicollinearity case.

```
model2 = lm(Life.expectancy~.-percentage.expenditure-thinness..1.19.years-thinness.5.9.years-Income.comp
summary(model2)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ . - percentage.expenditure - thinness..1.19.years -
##     thinness.5.9.years - Income.composition.of.resources - Population -
##     infant.deaths - under.five.deaths, data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.3723  -2.3559   0.0234   2.5060  16.6845
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.334e+01  5.360e-01  99.514  < 2e-16 ***
## Adult.Mortality   -2.168e-02  8.233e-04 -26.338  < 2e-16 ***
## Alcohol            1.078e-01  2.398e-02   4.495 7.22e-06 ***
## Hepatitis.B       -2.006e-02  3.841e-03  -5.223 1.88e-07 ***
## Measles           -3.657e-05  7.022e-06  -5.209 2.03e-07 ***
## BMI                5.492e-02  4.778e-03  11.496  < 2e-16 ***
## Polio              3.226e-02  4.659e-03   6.924 5.37e-12 ***
## Total.expenditure  9.212e-02  3.482e-02   2.645   0.0082 **
## Diphtheria         5.101e-02  4.807e-03  10.611  < 2e-16 ***
## HIV.AIDS          -4.942e-01  1.828e-02 -27.036  < 2e-16 ***
## GDP                6.365e-05  6.652e-06   9.570  < 2e-16 ***
## Schooling          9.697e-01  3.392e-02  28.589  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.251 on 2926 degrees of freedom
## Multiple R-squared:  0.8009, Adjusted R-squared:  0.8002
## F-statistic:  1070 on 11 and 2926 DF,  p-value: < 2.2e-16
```

After removing the 7 insignificant variables model2 is been made and R^2 value of model2 is 80.02 that means it decreases by only 0.01.

Diagnostic Check For model2

```
library(ggplot2)
library(ggfortify)
autoplot(model2)
```

Interaction terms

Now we can check whether there is correlation between any two variables or not by the help of correlation plot and correlation matrix.

```
library(corrplot)
```

```
## corrplot 0.90 loaded
```

```
corrplot(cor(data3),type="upper",method="circle",title="Correlation plot between variables",
         mar=c(0.7,0.7,0.7,0.7),tl.cex = 0.6)
```

```
cor(Adult.Mortality,HIV.AIDS)
```

```
## [1] 0.5238865
```

```
cor(BMI, thinness..1.19.years)
```

```
## [1] -0.5324869
```

```
cor(BMI, thinness.5.9.years)
```

```
## [1] -0.5393652
```

```
cor(BMI, Schooling)
```

```
## [1] 0.4998062
```

```
cor(Polio, Diphtheria)
```

```
## [1] 0.6739768
```

```
cor(Schooling, Income.composition.of.resources)
```

```
## [1] 0.7953833
```

From the above correlation plot and correlation value of each variables it is found that a decent amount of correlation exsit between Adult-mortality & HIV/AIDS, BMI & thinness(1-19years), BMI & thinness(5-9years), BMI & Schooling, Polio & Diphtheria, Schooling & Income-composition of resources.So their interaction terms is been added in our model.

```r
test1 = lm(Life.expectancy~.-percentage.expenditure-thinness..1.19.years-thinness.5.9.years-Income.compo
summary(test1)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ . - percentage.expenditure - thinness..1.19.years -
##     thinness.5.9.years - Income.composition.of.resources - Population -
##     infant.deaths - under.five.deaths + Adult.Mortality:HIV.AIDS +
##     Schooling:Income.composition.of.resources + BMI:Schooling +
##     Polio:Diphtheria + BMI:thinness..1.19.years + BMI:thinness.5.9.years,
##     data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.2899  -2.0040  -0.0193   2.1680  18.1468
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                              5.405e+01  8.729e-01  61.918
## Adult.Mortality                         -2.686e-02  8.690e-04 -30.915
## Alcohol                                  8.678e-02  2.315e-02   3.748
## Hepatitis.B                             -1.472e-02  3.583e-03  -4.107
## Measles                                 -3.266e-05  6.473e-06  -5.047
## BMI                                      2.278e-01  1.628e-02  13.988
## Polio                                    1.293e-02  8.336e-03   1.551
## Total.expenditure                        4.846e-02  3.259e-02   1.487
## Diphtheria                               2.864e-02  8.290e-03   3.455
## HIV.AIDS                                -8.648e-01  3.061e-02 -28.255
## GDP                                      4.650e-05  6.453e-06   7.206
## Schooling                                8.897e-01  6.786e-02  13.110
## Adult.Mortality:HIV.AIDS                 1.040e-03  6.653e-05  15.626
## Income.composition.of.resources:Schooling  5.318e-01  4.755e-02  11.184
## BMI:Schooling                           -1.461e-02  1.248e-03 -11.702
## Polio:Diphtheria                         1.851e-04  1.096e-04   1.688
## BMI:thinness..1.19.years                -2.013e-03  2.653e-03  -0.759
## BMI:thinness.5.9.years                  -1.581e-03  2.621e-03  -0.603
##                                         Pr(>|t|)
## (Intercept)                              < 2e-16 ***
## Adult.Mortality                          < 2e-16 ***
## Alcohol                                 0.000182 ***
## Hepatitis.B                             4.12e-05 ***
## Measles                                 4.77e-07 ***
## BMI                                      < 2e-16 ***
## Polio                                   0.121051
## Total.expenditure                       0.137164
## Diphtheria                              0.000558 ***
## HIV.AIDS                                 < 2e-16 ***
## GDP                                     7.33e-13 ***
## Schooling                                < 2e-16 ***
## Adult.Mortality:HIV.AIDS                 < 2e-16 ***
## Income.composition.of.resources:Schooling  < 2e-16 ***
## BMI:Schooling                            < 2e-16 ***
## Polio:Diphtheria                        0.091475 .
## BMI:thinness..1.19.years                0.447948
```

```
## BMI:thinness.5.9.years                    0.546402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.899 on 2920 degrees of freedom
## Multiple R-squared:  0.8329, Adjusted R-squared:  0.8319
## F-statistic: 855.8 on 17 and 2920 DF,  p-value: < 2.2e-16
```

From the above testing model we can see that Polio & Diphtheria,BMI & thinness..1.19.years ,BMI & thinness.5.9.years are not significant as their p-values are greater than 0.05.So we can drop this interaction terms.

```
model3 = lm(Life.expectancy~.-percentage.expenditure-thinness..1.19.years-thinness.5.9.years-Income.comp
summary(model3)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ . - percentage.expenditure - thinness..1.19.years -
##     thinness.5.9.years - Income.composition.of.resources - Population -
##     infant.deaths - under.five.deaths + Adult.Mortality:HIV.AIDS +
##     Schooling:Income.composition.of.resources + BMI:Schooling,
##     data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7886  -2.0716  -0.0187   2.2203  18.3688
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             5.299e+01  7.432e-01  71.308
## Adult.Mortality                        -2.706e-02  8.690e-04 -31.137
## Alcohol                                 1.145e-01  2.240e-02   5.110
## Hepatitis.B                            -1.431e-02  3.548e-03  -4.034
## Measles                                -3.508e-05  6.469e-06  -5.424
## BMI                                     2.175e-01  1.615e-02  13.467
## Polio                                   2.441e-02  4.302e-03   5.675
## Total.expenditure                       7.560e-02  3.208e-02   2.357
## Diphtheria                              4.022e-02  4.450e-03   9.038
## HIV.AIDS                               -8.766e-01  3.058e-02 -28.668
## GDP                                     4.777e-05  6.464e-06   7.390
## Schooling                               8.706e-01  6.784e-02  12.833
## Adult.Mortality:HIV.AIDS                1.042e-03  6.667e-05  15.626
## Income.composition.of.resources:Schooling  5.497e-01  4.755e-02  11.562
## BMI:Schooling                          -1.422e-02  1.247e-03 -11.397
##                                        Pr(>|t|)
## (Intercept)                             < 2e-16 ***
## Adult.Mortality                         < 2e-16 ***
## Alcohol                                3.42e-07 ***
## Hepatitis.B                            5.63e-05 ***
## Measles                                6.32e-08 ***
## BMI                                     < 2e-16 ***
## Polio                                  1.53e-08 ***
## Total.expenditure                        0.0185 *
## Diphtheria                              < 2e-16 ***
```

```
## HIV.AIDS                                      < 2e-16 ***
## GDP                                          1.90e-13 ***
## Schooling                                     < 2e-16 ***
## Adult.Mortality:HIV.AIDS                      < 2e-16 ***
## Income.composition.of.resources:Schooling  < 2e-16 ***
## BMI:Schooling                                 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.912 on 2923 degrees of freedom
## Multiple R-squared:  0.8316, Adjusted R-squared:  0.8308
## F-statistic:  1031 on 14 and 2923 DF,  p-value: < 2.2e-16
```

Clearly by adding the above interaction terms cause 3%increase in our R^2 Value and R^2 value become 83. Also this cause a decrement of RSE value to 3.912.So It's a better model with respect to model2.

Diagnostic Check for model3

```
library(ggplot2)
library(ggfortify)
autoplot(model3)
```

Polynomial terms

```
library(ggplot2)
p1 = ggplot(data3 , aes(y = Life.expectancy, x = Adult.Mortality ))
f1 = p1 + geom_point() + ggtitle("Life Expectancy vs Adult.Mortality") + stat_smooth() + stat_smooth(me
```

```
p2 = ggplot(data3 , aes(y = Life.expectancy, x = Alcohol ))
f2 = p2 + geom_point() + ggtitle("Life Expectancy vs Alcohol") + stat_smooth() + stat_smooth(method = lm

p3 = ggplot(data3 , aes(y = Life.expectancy, x = Hepatitis.B ))
f3 = p3 + geom_point() + ggtitle("Life Expectancy vs Hepatitis.B") + stat_smooth() + stat_smooth(method

p4 = ggplot(data3 , aes(y = Life.expectancy, x = Measles ))
f4 = p4 + geom_point() + ggtitle("Life Expectancy vs Measles") + stat_smooth() + stat_smooth(method = lm

p5 = ggplot(data3 , aes(y = Life.expectancy, x = BMI ))
f5 = p5 + geom_point() + ggtitle("Life Expectancy vs BMI") + stat_smooth() + stat_smooth(method = lm , 

p6 = ggplot(data3 , aes(y = Life.expectancy, x = Polio ))
f6 = p6 + geom_point() + ggtitle("Life Expectancy vs Polio") + stat_smooth() + stat_smooth(method = lm

p7 = ggplot(data3 , aes(y = Life.expectancy, x = Total.expenditure ))
f7 = p7 + geom_point() + ggtitle("Life Expectancy vs Total.expenditure") + stat_smooth() + stat_smooth(

p8 = ggplot(data3 , aes(y = Life.expectancy, x = Diphtheria ))
f8 = p8 + geom_point() + ggtitle("Life Expectancy vs Diphtheria") + stat_smooth() + stat_smooth(method =

p9 = ggplot(data3 , aes(y = Life.expectancy, x = HIV.AIDS ))
f9 = p9 + geom_point() + ggtitle("Life Expectancy vs HIV/AIDS") + stat_smooth() +stat_smooth(method = lm

p10 = ggplot(data3 , aes(y = Life.expectancy, x = GDP ))
f10 = p10 + geom_point() + ggtitle("Life Expectancy vs GDP") + stat_smooth() + stat_smooth(method = lm

p11 = ggplot(data3 , aes(y = Life.expectancy, x = Schooling ))
f11 = p11 + geom_point() + ggtitle("Life Expectancy vs Schooling") + stat_smooth() + stat_smooth(method
```

```
library(gridExtra)
grid.arrange(f1,f2)
```

```
grid.arrange(f3,f4)
```

```
grid.arrange(f5,f6)
```

```
grid.arrange(f7,f8)
```

```
grid.arrange(f9,f10)
```

Life Expectancy vs HIV/AIDS
Life.expectancy
75
50
25
0
10
20
30
40
50
HIV.AIDS

Life Expectancy vs GDP
Life.expectancy
110
90
70
50
0
25000
50000
75000
100000
GDP

```
f11
```

```
test4 = lm(Life.expectancy~.-percentage.expenditure-thinness..1.19.years-thinness.5.9.years-Income.compo
summary(test4)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ . - percentage.expenditure - thinness..1.19.years -
##     thinness.5.9.years - Income.composition.of.resources - Population -
##     infant.deaths - under.five.deaths + Adult.Mortality:HIV.AIDS +
##     Schooling:Income.composition.of.resources + BMI:Schooling -
##     HIV.AIDS + poly(HIV.AIDS, 2) + poly(Polio, 7), data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.4817  -2.1095  -0.0122   2.1545  15.1688
##
## Coefficients: (1 not defined because of singularities)
##                                          Estimate Std. Error t value
## (Intercept)                             5.385e+01  7.017e-01  76.743
## Adult.Mortality                        -2.278e-02  8.519e-04 -26.741
## Alcohol                                 1.297e-01  2.112e-02   6.139
## Hepatitis.B                            -8.408e-03  3.457e-03  -2.432
## Measles                                -2.159e-05  6.115e-06  -3.531
## BMI                                     1.635e-01  1.539e-02  10.629
## Polio                                   4.424e-02  4.452e-03   9.936
## Total.expenditure                       7.150e-02  3.021e-02   2.367
## Diphtheria                              1.251e-02  4.506e-03   2.777
```

```
## GDP                                        5.160e-05  6.051e-06   8.528
## Schooling                                  6.561e-01  6.496e-02  10.101
## poly(HIV.AIDS, 2)1                        -2.311e+02  7.943e+00 -29.095
## poly(HIV.AIDS, 2)2                         5.474e+01  4.142e+00  13.217
## poly(Polio, 7)1                                   NA         NA      NA
## poly(Polio, 7)2                            4.310e+01  4.451e+00   9.682
## poly(Polio, 7)3                           -1.858e+01  3.782e+00  -4.913
## poly(Polio, 7)4                           -1.343e+01  3.698e+00  -3.630
## poly(Polio, 7)5                            2.457e+01  3.809e+00   6.452
## poly(Polio, 7)6                           -2.718e+01  3.756e+00  -7.238
## poly(Polio, 7)7                            1.392e+01  3.688e+00   3.776
## Adult.Mortality:HIV.AIDS                   8.323e-04  6.332e-05  13.145
## Income.composition.of.resources:Schooling  5.421e-01  4.484e-02  12.091
## BMI:Schooling                             -1.057e-02  1.186e-03  -8.917
##                                           Pr(>|t|)
## (Intercept)                                < 2e-16 ***
## Adult.Mortality                            < 2e-16 ***
## Alcohol                                   9.42e-10 ***
## Hepatitis.B                               0.015061 *
## Measles                                   0.000421 ***
## BMI                                        < 2e-16 ***
## Polio                                      < 2e-16 ***
## Total.expenditure                         0.018003 *
## Diphtheria                                0.005514 **
## GDP                                        < 2e-16 ***
## Schooling                                  < 2e-16 ***
## poly(HIV.AIDS, 2)1                         < 2e-16 ***
## poly(HIV.AIDS, 2)2                         < 2e-16 ***
## poly(Polio, 7)1                                 NA
## poly(Polio, 7)2                            < 2e-16 ***
## poly(Polio, 7)3                           9.46e-07 ***
## poly(Polio, 7)4                           0.000288 ***
## poly(Polio, 7)5                           1.28e-10 ***
## poly(Polio, 7)6                           5.80e-13 ***
## poly(Polio, 7)7                           0.000163 ***
## Adult.Mortality:HIV.AIDS                   < 2e-16 ***
## Income.composition.of.resources:Schooling  < 2e-16 ***
## BMI:Schooling                              < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.659 on 2916 degrees of freedom
## Multiple R-squared:  0.853,  Adjusted R-squared:  0.852
## F-statistic: 805.9 on 21 and 2916 DF,  p-value: < 2.2e-16
```

From the above plots and summary of testing model we can conclude that the model can be improved by adding the 2nd degree polynomial of HIV/AIDS and 7th degree polynomial of Polio as the R^2 value increase to 85.3.But on the stake of adding 6 more variables due to the polynomial of Polio the model is improved by only 1. So it will be better if the polynomial terms of Polio are dropped.

```
model4 = lm(Life.expectancy~.-percentage.expenditure-thinness..1.19.years-thinness.5.9.years-Income.comp
summary(model4)
```

```
##
```

```
## Call:
## lm(formula = Life.expectancy ~ . - percentage.expenditure - thinness..1.19.years -
##     thinness.5.9.years - Income.composition.of.resources - Population -
##     infant.deaths - under.five.deaths + Adult.Mortality:HIV.AIDS +
##     Schooling:Income.composition.of.resources + BMI:Schooling -
##     HIV.AIDS + poly(HIV.AIDS, 2), data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5067  -2.0497  -0.0117   2.1674  17.1798
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                              5.218e+01  7.114e-01  73.349
## Adult.Mortality                         -2.359e-02  8.746e-04 -26.968
## Alcohol                                  1.314e-01  2.169e-02   6.057
## Hepatitis.B                             -1.116e-02  3.438e-03  -3.247
## Measles                                 -3.221e-05  6.258e-06  -5.147
## BMI                                      1.943e-01  1.570e-02  12.379
## Polio                                    2.118e-02  4.166e-03   5.085
## Total.expenditure                        8.815e-02  3.103e-02   2.841
## Diphtheria                               3.566e-02  4.315e-03   8.264
## GDP                                      5.032e-05  6.252e-06   8.048
## Schooling                                8.188e-01  6.570e-02  12.464
## poly(HIV.AIDS, 2)1                      -2.414e+02  8.136e+00 -29.671
## poly(HIV.AIDS, 2)2                       6.078e+01  4.251e+00  14.297
## Adult.Mortality:HIV.AIDS                 9.081e-04  6.515e-05  13.939
## Income.composition.of.resources:Schooling  5.342e-01  4.599e-02  11.616
## BMI:Schooling                           -1.292e-02  1.210e-03 -10.681
##                                         Pr(>|t|)
## (Intercept)                              < 2e-16 ***
## Adult.Mortality                          < 2e-16 ***
## Alcohol                                 1.57e-09 ***
## Hepatitis.B                              0.00118 **
## Measles                                 2.82e-07 ***
## BMI                                      < 2e-16 ***
## Polio                                   3.92e-07 ***
## Total.expenditure                        0.00453 **
## Diphtheria                               < 2e-16 ***
## GDP                                     1.21e-15 ***
## Schooling                                < 2e-16 ***
## poly(HIV.AIDS, 2)1                       < 2e-16 ***
## poly(HIV.AIDS, 2)2                       < 2e-16 ***
## Adult.Mortality:HIV.AIDS                 < 2e-16 ***
## Income.composition.of.resources:Schooling  < 2e-16 ***
## BMI:Schooling                            < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.782 on 2922 degrees of freedom
## Multiple R-squared:  0.8426, Adjusted R-squared:  0.8418
## F-statistic:  1043 on 15 and 2922 DF,  p-value: < 2.2e-16
```

Diagnostic Check for model4:

```
library(ggfortify)
autoplot(model4)
```

conclusion of final model:
1). R^2 value of our final model is 84.18.
2). From the residual vs fitted graph we can see that the estimated error curve of our final model is almost converge to 0.
3). From the QQ-Plot we can see that the our model shows behave like normal except for the tail parts.
So our Final model is:
$LifeExpectancy = (52.18) + (-0.0235) * Adult.Mortality + (0.1314) * Alcohol + (-0.01116) * Hepatitis.B + (-0.000032) * Measles + (0.1943) * BMI + (0.02118) * Polio + (0.08815) * Total.expenditure + (0.0357) * Diphtheria + (0.00005032) * GDP + (0.8188) * Schooling + (-24.14) * HIV.AIDS + (60.78) * HIV.AIDS^2 + (0.000908) * (Adult.Mortality : HIV.AIDS) + (0.5342) * (Income.composition.of.resources : Schooling) + (-0.01292) * (BMI : Schooling)$