Life Expectancy Data Analysis

Jorania F. Alves and Sneha Verma

01/08/2021

This research focuses on predicting the life expectancy of a nation based on various health features. We started by conducting basic exploration and transformations of the features to account for skewness. Then, we conducted backward elimination and applied a first-order model to see which variables are significant and found out that hepatitis B is not a significant variable. To make a more reliable and a better predictive model, we conducted step-wise regression on centered interaction effects between all variables. We found out that many of the interaction effects of the variables are significant proving that certain health factors have an influence on life expectancy of a nation dependent on another health factor. However, our data set has a few clusters for various variables that also have large residuals and/or leverage points (points that have a large influence) suggesting that we should consider different transformations for these variables. Hence, while we have used the AIC criterion to make our predictive model more reliable, we would suggest conducting a few more tests and transformations to account for high leverage and residual points.

Note: Before running the data, change the location of the data

This data set describes the life expectancy of multiple countries over a period of 15 years. It provides various health and economic factors regarding the country for that year, such as BMI (Body Mass Index), GDP (Gross Domestic Product), Hepatitis B immunization coverage, etc. This research project will focus on the central question:

**Do health factors influence the life expectancy of a country?**

The researchers will look at the following predictor variables:

Adult Mortality: Adult mortality rates (probability of dying between 15 and 60 years per 1000 population).

infant deaths: Number of infant deaths per 1000 population.

Alcohol: recorded per capita (15+ years) consumption in liters of pure alcohol.

Hepatitis B: Hepatitis B immunization coverage among one-year olds in percentage.

Measles: number of reported cases per 1000 population.

BMI: Average body mass index of entire population (the units were not provided, hence, the researchers will use the most common unit for describing BMI: kilograms per meter squared).

Polio: polio immunization coverage among one-year olds in percentage.

Diphtheria: Diphtheria tetanus toxoid and pertussis immunization coverage among one-year olds in percentage.

thinness 1-19 years: prevalence of thinness among children and adolescents between the ages 10 and 19 in percentage.

thinness 5-9 years: prevalence of thinness among children and adolescents between the ages 5 and 9 in percentage.

Predictor variable: Life expectancy in age (years)

It should be noted that since the units of are complex, interpretations of parameters will be mostly qualitative to indicate the direction of the influence.

## Import the data set

lifeExpectancy = read.csv('/Users/sneha\_verma/Documents/MATH 327/Project1/Data/Life Expectancy Data.csv')   
  
# Remove any unnecessary variables:  
life\_expectancy0 = subset(lifeExpectancy, select = -c(Country, Year, Status, percentage.expenditure, under.five.deaths, Total.expenditure, HIV.AIDS, GDP, Population, Income.composition.of.resources, Schooling))  
  
# make dataset of complete observations for all columns.  
life\_expectancy = life\_expectancy0[complete.cases(life\_expectancy0), ]  
  
# Rename columns for convenience:  
colnames(life\_expectancy) = c("lifeExpectancy", "adult\_mortality", "infant\_deaths", "alcohol", "hepatitisB", "measles", "bmi", "polio", "diphtheria", "thin10to19", "thin5to9")

# Exploratory analysis

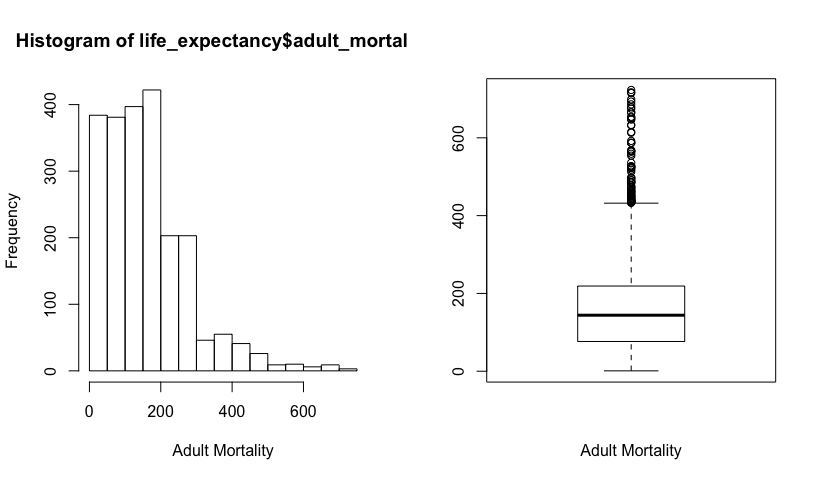
## adult\_mortality

summary(life\_expectancy$adult\_mortality)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.0 76.5 144.0 160.7 219.0 723.0

This shows that the minimum value of adult mortality is not 0, allowing us to make a log transformation, if necessary.

par(mfrow=c(1,2))  
hist(life\_expectancy$adult\_mortality, xlab = 'Adult Mortality')  
boxplot(life\_expectancy$adult\_mortality, xlab = 'Adult Mortality')



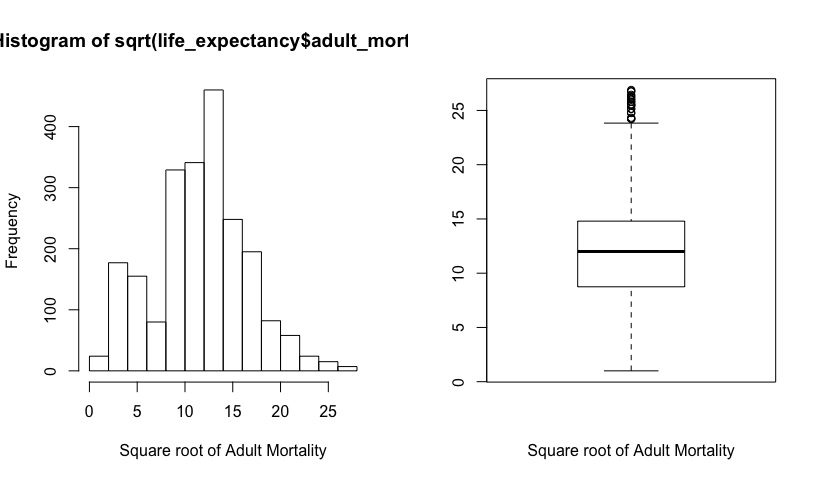
The distribution of adult mortality is clearly right-skewed. Let us try a log transformation to account for the skewness.

par(mfrow=c(1,2))  
hist(log(life\_expectancy$adult\_mortality), xlab = 'Log of Adult Mortality')  
boxplot(log(life\_expectancy$adult\_mortality), xlab = 'Log of Adult Mortality')



The log transformation over adjusts the skewness making the variable left-skewed. Thus, let us try a square-root transformation

par(mfrow=c(1,2))  
hist(sqrt(life\_expectancy$adult\_mortality), xlab = 'Square root of Adult Mortality')  
boxplot(sqrt(life\_expectancy$adult\_mortality), xlab = 'Square root of Adult Mortality')



The distribution of sqrt(adult-mortality) is more symmetric.

### infant\_deaths

summary(life\_expectancy$infant\_deaths)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 0.00 3.00 28.01 20.00 1600.00

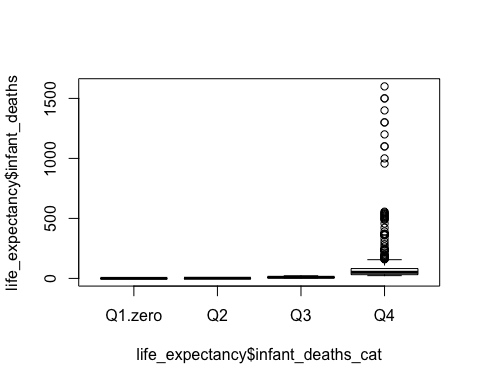
The minimum value of infant deaths is 0, indicating that a log transformation, if necessary, will not be possible since log(0) is undefined.

par(mfrow=c(1,2))  
hist(life\_expectancy$infant\_deaths, xlab = 'Number of infant deaths per 1000')  
boxplot(life\_expectancy$infant\_deaths, xlab = 'Number of infant deaths per 1000')



The distribution of the number of infant deaths per 1000 population is clearly right-skewed. Since this variable has the value zero, let us convert it into a categorical variable where each “bin” is set according to the quartiles found from the summary() function.

# Create a new variable where all values are "Q1.zero"  
life\_expectancy$infant\_deaths\_cat = "Q1.zero"  
  
# Change the new variable to "Q2" for the rows not in category 1  
life\_expectancy$infant\_deaths\_cat[life\_expectancy$infant\_deaths > 0] = "Q2"  
  
# Change the new variable to "Q3" for the rows not in category 1 or 2  
life\_expectancy$infant\_deaths\_cat[life\_expectancy$infant\_deaths > 3.0] = "Q3"  
  
# Change the new variable to "Q4" for the rows not in category 1, 2, or 3  
life\_expectancy$infant\_deaths\_cat[life\_expectancy$infant\_deaths > 22.0] = "Q4"  
  
# visualize the results  
boxplot(life\_expectancy$infant\_deaths ~ life\_expectancy$infant\_deaths\_cat)



# Count number of cases in each category  
table(life\_expectancy$infant\_deaths\_cat)

##   
## Q1.zero Q2 Q3 Q4   
## 613 574 491 517

### alcohol

summary(life\_expectancy$alcohol)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.010 0.690 3.500 4.445 7.540 17.870

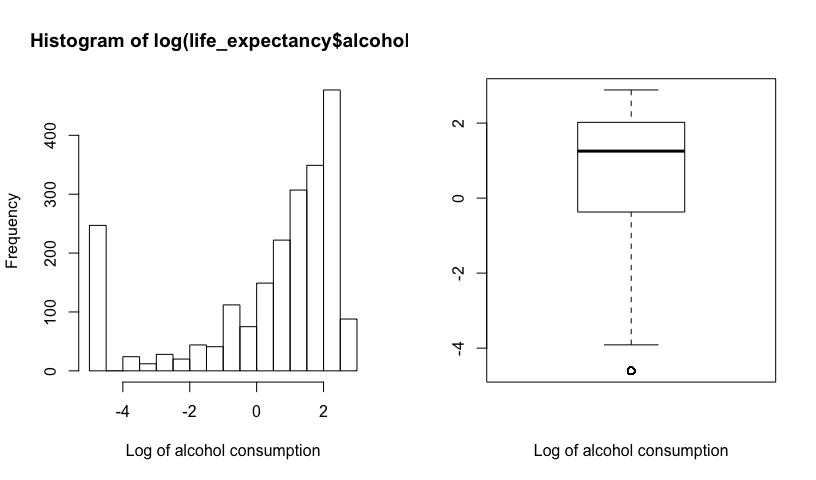
The minimum value of alcohol consumption is 0.01, indicating that a log transformation, if necessary, will be possible.

par(mfrow=c(1,2))  
hist(life\_expectancy$alcohol, xlab = 'Alcohol consumption')  
boxplot(life\_expectancy$alcohol, xlab = 'Alcohol consumption')



The distribution of alcohol consumption is clearly right-skewed. Let us try a log transformation to account for the skewness.

par(mfrow=c(1,2))  
hist(log(life\_expectancy$alcohol), xlab = 'Log of alcohol consumption')  
boxplot(log(life\_expectancy$alcohol), xlab = 'Log of alcohol consumption')



The log transformation over adjusts the skewness making the variable left-skewed. Thus, let us try a square-root transformation

par(mfrow=c(1,2))  
hist(sqrt(life\_expectancy$alcohol), xlab = 'Log of alcohol consumption')  
boxplot(sqrt(life\_expectancy$alcohol), xlab = 'Log of alcohol consumption')



The distribution of sqrt(alcohol) is more symmetric.

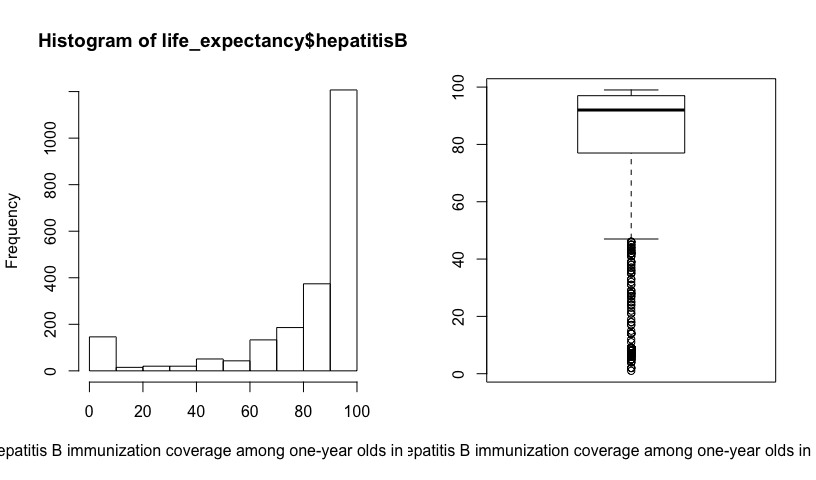
### hepatitsB

summary(life\_expectancy$hepatitisB)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 77.00 92.00 80.91 97.00 99.00

The minimum value of the percentage of one-year olds with HepatitisB immunization coverage is 1%.

par(mfrow=c(1,2))  
hist(life\_expectancy$hepatitisB, xlab = 'Hepatitis B immunization coverage among one-year olds in percent')  
boxplot(life\_expectancy$hepatitisB, xlab = 'Hepatitis B immunization coverage among one-year olds in percent')



The distribution of hepatits B among 1-year old in percent is clearly left-skewed. However, since the variable is in the units of percentage, let us leave the variable untouched and investigate its effects in the model.

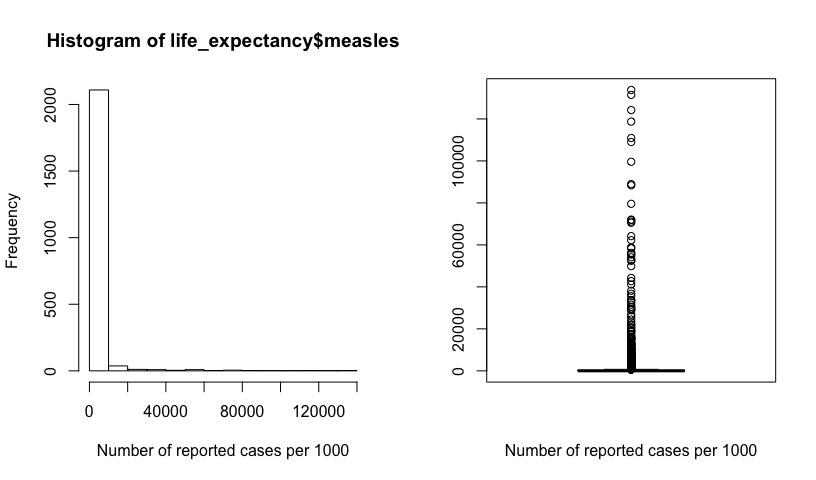
### measles

summary(life\_expectancy$measles)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0 0 13 1995 290 133802

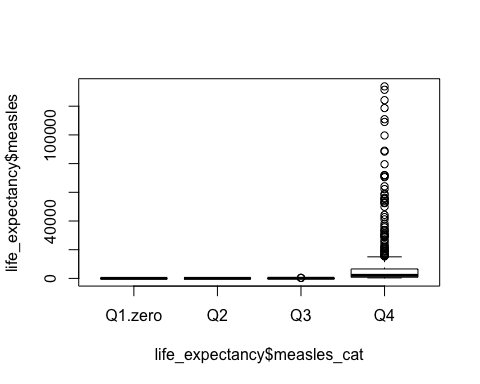
The minimum value of the number of reported cases of measles per 1000 is 0, indicating that a log transformation, if necessary, will not be possible since log(0) is undefined.

par(mfrow=c(1,2))  
hist(life\_expectancy$measles, xlab = 'Number of reported cases per 1000')  
boxplot(life\_expectancy$measles, xlab = 'Number of reported cases per 1000')



The distribution of the number of infant deaths per 1000 population is clearly right-skewed. Since this variable has the value zero, let us convert it into a categorical variable where each “bin” is set according to the quartiles found from the summary() function.

# Create a new variable where all values are "Q1.zero"  
life\_expectancy$measles\_cat = "Q1.zero"  
  
# Change the new variable to "Q2" for the rows not in category 1  
life\_expectancy$measles\_cat[life\_expectancy$measles > 0] = "Q2"  
  
# Change the new variable to "Q3" for the rows not in category 1 or 2  
life\_expectancy$measles\_cat[life\_expectancy$measles > 17.0] = "Q3"  
  
# Change the new variable to "Q4" for the rows not in category 1, 2, or 3  
life\_expectancy$measles\_cat[life\_expectancy$measles > 360.2] = "Q4"  
  
# visualize the results  
boxplot(life\_expectancy$measles ~ life\_expectancy$measles\_cat)



# Count number of cases in each category  
table(life\_expectancy$measles\_cat)

##   
## Q1.zero Q2 Q3 Q4   
## 762 379 548 506

### bmi

summary(life\_expectancy$bmi)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.40 21.30 44.70 39.02 56.30 77.10

The minimum value of body mass index is 1.40 kg/m^2, indicating that a log transformation, if necessary, will be possible.

par(mfrow=c(1,2))  
hist(life\_expectancy$bmi, xlab = 'Average body mass index of population')  
boxplot(life\_expectancy$bmi, xlab = 'Average body mass index of population')



The distribution of average bmi across populations is bi-modal, which is acceptable for this multiple linear regression analysis.

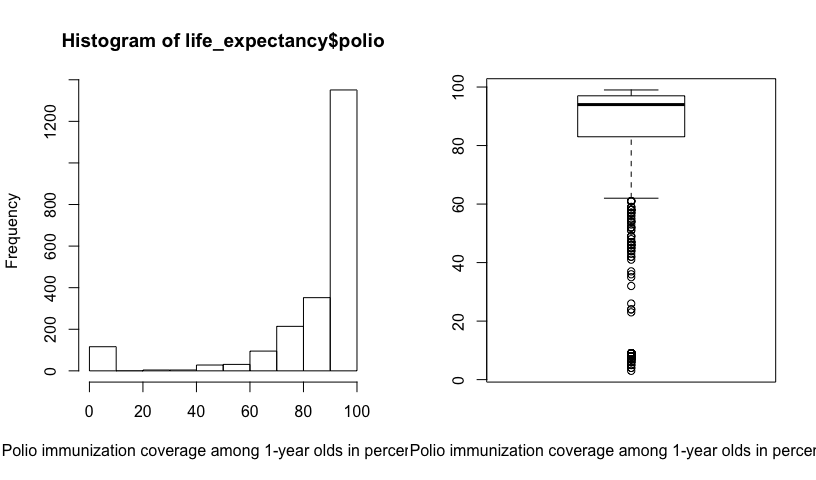
### polio

summary(life\_expectancy$polio)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3.00 83.00 94.00 84.97 97.00 99.00

The minimum value of one-year olds with polio immunization coverage is 3%, indicating that a log transformation, if necessary, will be possible.

par(mfrow=c(1,2))  
hist(life\_expectancy$polio, xlab = 'Polio immunization coverage among 1-year olds in percentage')  
boxplot(life\_expectancy$polio, xlab = 'Polio immunization coverage among 1-year olds in percentage')



The distribution of polio among 1 year olds cases is clearly left-skewed. However, since the variable is in the units of percentage, let us leave the variable untouched and investigate its effects in the model.

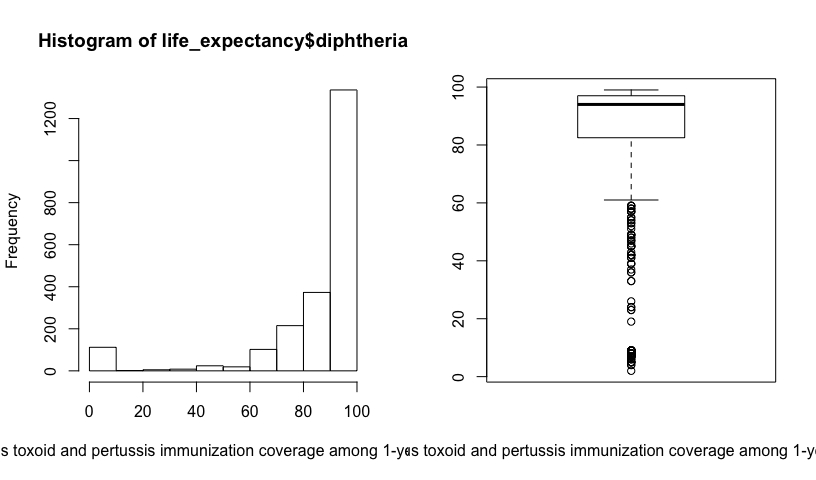
### diphtheria

summary(life\_expectancy$diphtheria)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 82.50 94.00 85.02 97.00 99.00

The minimum value of one-year olds with diphtheria immunization coverage is 2%, indicating that a log transformation, if necessary, will be possible.

par(mfrow=c(1,2))  
hist(life\_expectancy$diphtheria, xlab = 'Diphtheria tetanus toxoid and pertussis immunization coverage among 1-year olds in percentage')  
boxplot(life\_expectancy$diphtheria, xlab = 'Diphtheria tetanus toxoid and pertussis immunization coverage among 1-year olds in percentage')



The distribution of diphtheria cases cases is clearly left-skewed. However, since the variable is in the units of percentage, let us leave the variable untouched and investigate its effects in the model.

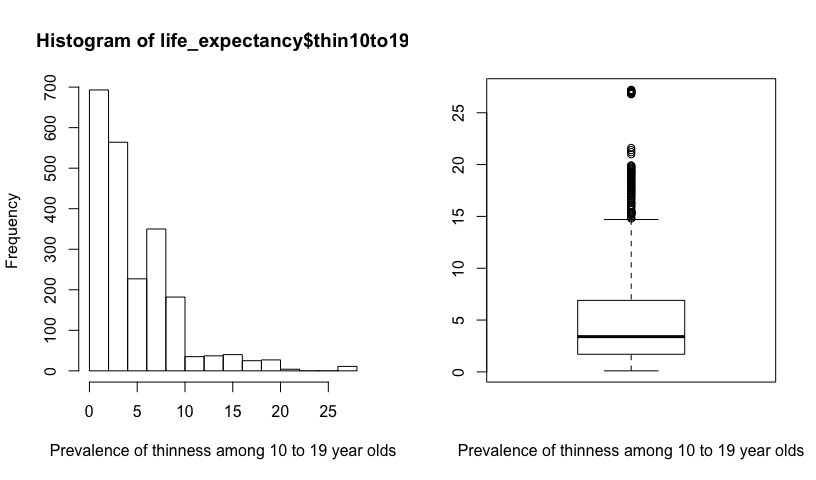
### thin10to19

summary(life\_expectancy$thin10to19)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.100 1.700 3.400 4.795 6.900 27.200

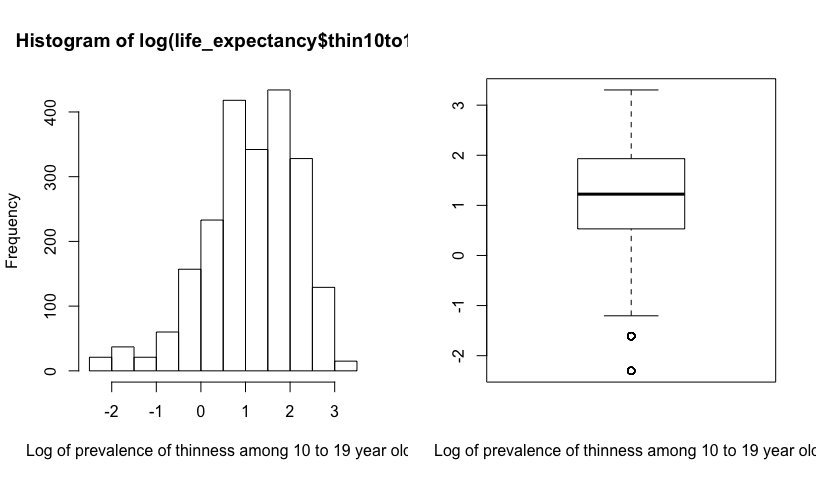
The minimum value of prevalence of thinness among 10 to 19 year olds is 0.1%, indicating that a log transformation, if necessary, will be possible.

par(mfrow=c(1,2))  
hist(life\_expectancy$thin10to19, xlab = 'Prevalence of thinness among 10 to 19 year olds')  
boxplot(life\_expectancy$thin10to19, xlab = 'Prevalence of thinness among 10 to 19 year olds')



The distribution of the prevalence of thinness among 10 to 19 year olds is clearly right-skewed. Let us try a log transformation to account for the skewness.

par(mfrow=c(1,2))  
hist(log(life\_expectancy$thin10to19), xlab = 'Log of prevalence of thinness among 10 to 19 year olds')  
boxplot(log(life\_expectancy$thin10to19), xlab = 'Log of prevalence of thinness among 10 to 19 year olds')



The log(thin10to19) is more symmetric.

### thin5to9

summary(life\_expectancy$thin5to9)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.100 1.700 3.400 4.822 6.900 28.200

The minimum value of prevalence of thinness among 5 to 9 year olds is 0.1%, indicating that a log transformation, if necessary, will be possible.

par(mfrow=c(1,2))  
hist(life\_expectancy$thin5to9, xlab = 'Prevalence of thinness among 5 to 9 year olds')  
boxplot(life\_expectancy$thin5to9, xlab = 'Prevalence of thinness among 5 to 9 year olds')



The distribution of the prevalence of thinness among 5 to 9 year olds is clearly right-skewed. Let us try a log transformation to account for the skewness.

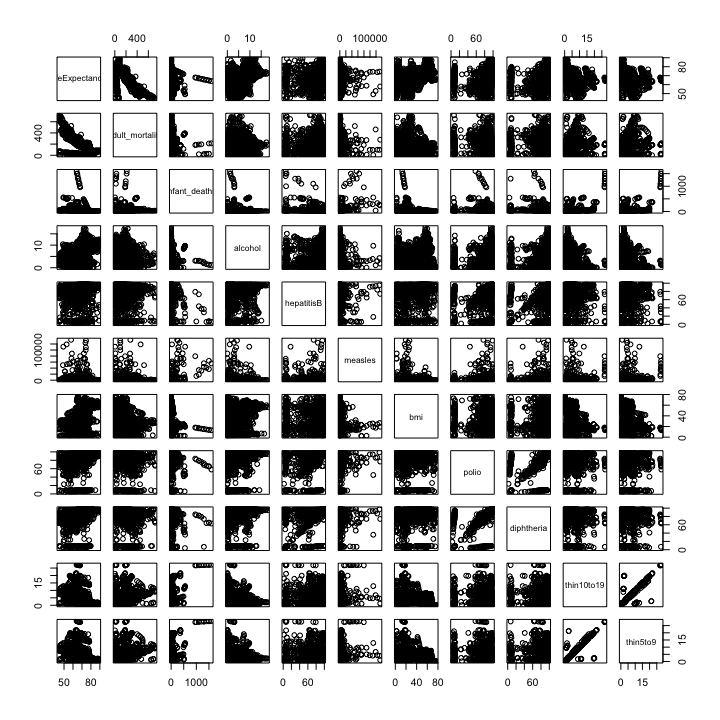
par(mfrow=c(1,2))  
hist(log(life\_expectancy$thin5to9), xlab = 'Log of prevalence of thinness among 5 to 9 year olds')  
boxplot(log(life\_expectancy$thin5to9), xlab = 'Log of prevalence of thinness among 5 to 9 year olds')



The distribution of log(thin5to9) is more symmetric.

## Scatter plot Matrix and Correlations

pairs(life\_expectancy[, 1:11])



Through the scatter plot matrix, it is clear that the relationship between the response variable (life expectancy) and the predictor variables is unclear because of the large number of the observations and the skewness in each variable (which has not been accounted for in this matrix).

Let us now look at pairwise correlations:

corr = cor(life\_expectancy[, 1:11], use = 'complete.obs')  
round(corr, 2)

## lifeExpectancy adult\_mortality infant\_deaths alcohol hepatitisB  
## lifeExpectancy 1.00 -0.72 -0.19 0.37 0.25  
## adult\_mortality -0.72 1.00 0.07 -0.14 -0.16  
## infant\_deaths -0.19 0.07 1.00 -0.10 -0.23  
## alcohol 0.37 -0.14 -0.10 1.00 0.09  
## hepatitisB 0.25 -0.16 -0.23 0.09 1.00  
## measles -0.10 0.03 0.52 -0.06 -0.13  
## bmi 0.51 -0.36 -0.23 0.28 0.15  
## polio 0.35 -0.23 -0.16 0.19 0.48  
## diphtheria 0.36 -0.23 -0.17 0.19 0.59  
## thin10to19 -0.44 0.27 0.44 -0.41 -0.13  
## thin5to9 -0.44 0.28 0.44 -0.39 -0.13  
## measles bmi polio diphtheria thin10to19 thin5to9  
## lifeExpectancy -0.10 0.51 0.35 0.36 -0.44 -0.44  
## adult\_mortality 0.03 -0.36 -0.23 -0.23 0.27 0.28  
## infant\_deaths 0.52 -0.23 -0.16 -0.17 0.44 0.44  
## alcohol -0.06 0.28 0.19 0.19 -0.41 -0.39  
## hepatitisB -0.13 0.15 0.48 0.59 -0.13 -0.13  
## measles 1.00 -0.16 -0.07 -0.08 0.16 0.16  
## bmi -0.16 1.00 0.20 0.18 -0.51 -0.52  
## polio -0.07 0.20 1.00 0.57 -0.16 -0.17  
## diphtheria -0.08 0.18 0.57 1.00 -0.17 -0.16  
## thin10to19 0.16 -0.51 -0.16 -0.17 1.00 0.94  
## thin5to9 0.16 -0.52 -0.17 -0.16 0.94 1.00

Observing the correlation matrix, it can be concluded that most predictor variables are not correlated to each other which shows that this data set does not display multicollinearity. However, the variables ‘thin5to9’ and ‘thin10to19’ have a high correlation of 0.94 indicating the presence of multicollinearity. Adult mortality, followed by body mass index, are the most correlated with the response variable, life expectancy.

Looking at the skewness and correlation, the variables indicating the percent of one-year olds with measles immunization coverage and the number of infant deaths should be categorical because they have large skewness with zero as values. Other variables that are left-skewed and have units in percentage have been left untouched to observe its significance in predicting life expectancy of a country.

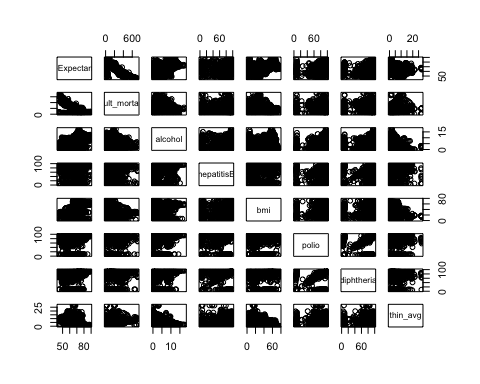
To deal with the high correlation between variables thin5to9 and thin10to19, let us create a third variable with the average of the first two:

thin = c('thin5to9', 'thin10to19')  
life\_expectancy$thin\_avg = rowMeans(life\_expectancy[thin])

We will use the thin\_avg variable in our first-order model.

Let us look at a scatter plot matrix with just the response variable and the predictor variables that will be used in the first-order model (non-transformed):

lst = c(1, 2, 4, 5, 7, 8, 9, 14)  
pairs(life\_expectancy[, lst])

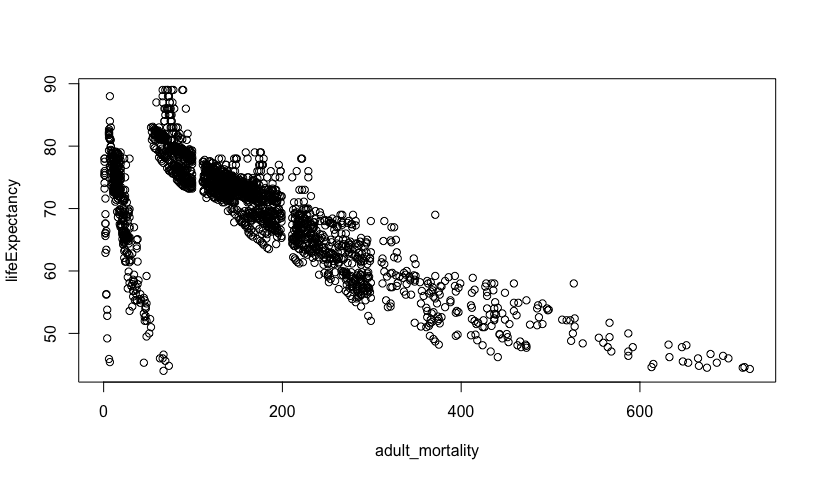


While this scatterplot matrix is largely similar to the previous scatterplot matrix in showing that the relationship between the response variable (life expectancy) and the predictor variables is unclear because of the large number of the observations and the skewness in each variable (which has not been accounted for in this matrix), the difference lies in the additional plot of ‘thin\_avg’ which is the average of the two thinness variables. The scatterplot between the response and the average of thinness, however, does not show strong correlations and rather appears to be a large cluster of points.

## Simple Linear Regression

Let us start with a simple linear regression model of the response variable and the predictor variable with the highest correlation with life expectancy, adult mortality:

plot(data = life\_expectancy, lifeExpectancy ~ adult\_mortality)



simpleFit = lm(data = life\_expectancy, lifeExpectancy ~ adult\_mortality)  
summary(simpleFit)

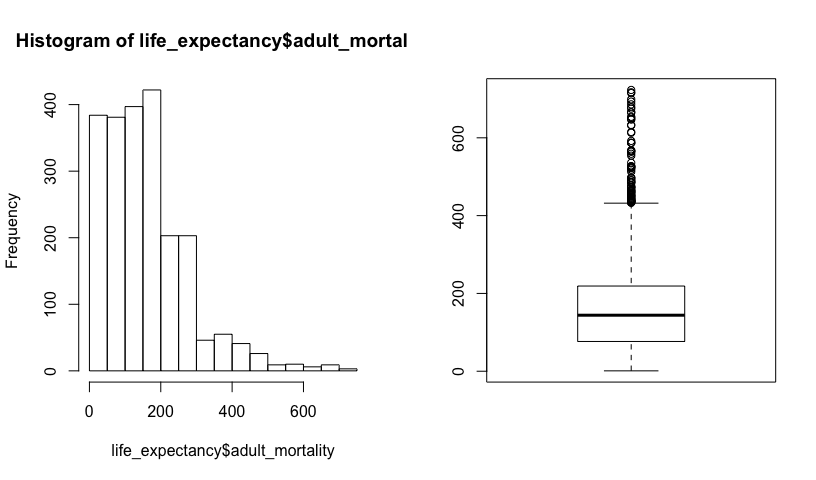
##   
## Call:  
## lm(formula = lifeExpectancy ~ adult\_mortality, data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -32.303 -2.449 1.113 3.524 15.457   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 78.058465 0.210921 370.1 <2e-16 \*\*\*  
## adult\_mortality -0.050737 0.001053 -48.2 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.904 on 2193 degrees of freedom  
## Multiple R-squared: 0.5144, Adjusted R-squared: 0.5142   
## F-statistic: 2323 on 1 and 2193 DF, p-value: < 2.2e-16

Looking at this plot, two or three clusters of points are visible indicating the possibility of errors in data entry or some unnatural events that created a large range of values of life expectancy for low adult mortality. We can explore the clusters with the following code, however, since multiple variables are displaying clusters, we will leave the variables untouched and continue with the regression analysis with the originally transformed variables.

# # create cluster 1  
# life\_expectancy$cluster1 = with(life\_expectancy, ifelse((lifeExpectancy < 70 & adult\_mortality < 100) | (lifeExpectancy >= 70 & adult\_mortality < 45), 1, 2))  
# plot(lifeExpectancy ~ adult\_mortality, data = life\_expectancy, col = cluster1)  
# life\_expectancy$adult\_mort2 = ifelse(life\_expectancy$cluster1 == 1, 10, 1) \* life\_expectancy$adult\_mortality  
# plot(lifeExpectancy ~ adult\_mort2, data = life\_expectancy, col = cluster1)  
#   
# # crate cluster 2  
# life\_expectancy$cluster2 = with(life\_expectancy, ifelse((lifeExpectancy < 90 - 0.3 \* adult\_mort2), 3, 4))  
# plot(lifeExpectancy ~ adult\_mort2, data = life\_expectancy, col = cluster2)  
# abline(90, -0.3)  
# life\_expectancy$adult\_mort3 = ifelse(life\_expectancy$cluster2 == 3, 10, 1) \* life\_expectancy$adult\_mort2  
# plot(lifeExpectancy ~ adult\_mort3, data = life\_expectancy, col = cluster2)

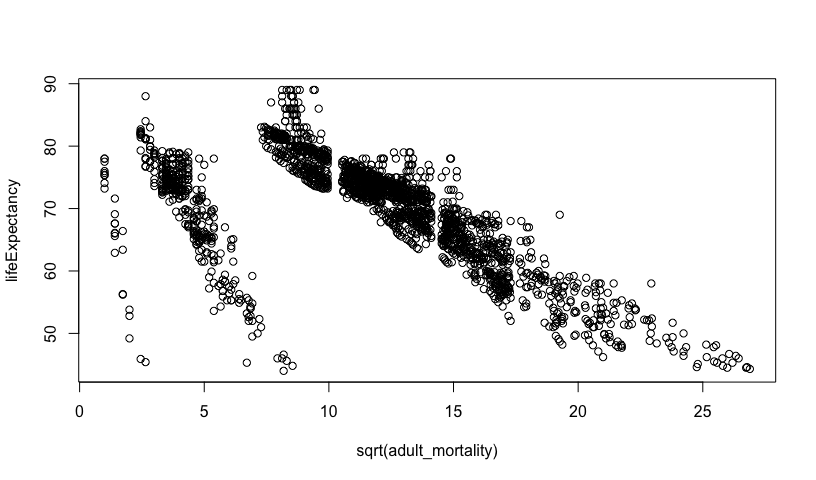
Let us have a second look at the distribution of adult mortality:

par(mfrow=c(1,2))  
hist(life\_expectancy$adult\_mortality)  
boxplot(life\_expectancy$adult\_mortality)



Since the variable is right-skewed, we transformed the variable with a square-root transformation. Let us use the transformed variable to fit a model:

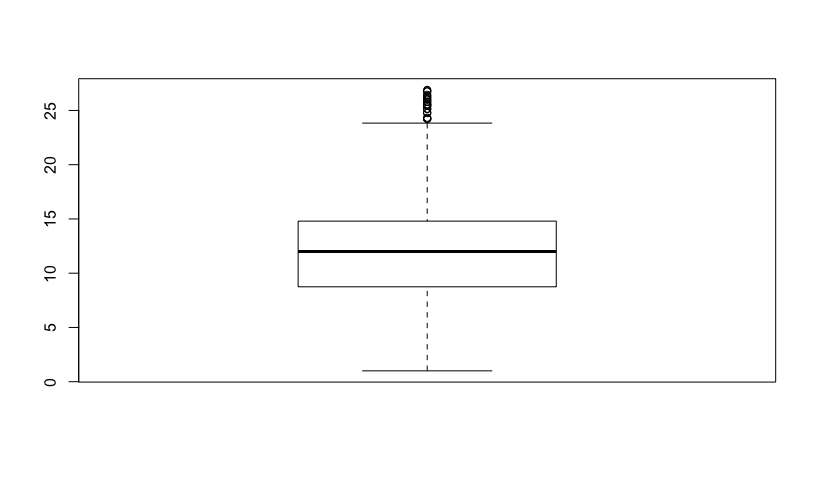
plot(data = life\_expectancy, lifeExpectancy ~ sqrt(adult\_mortality))



hist(sqrt(life\_expectancy$adult\_mortality))



boxplot(sqrt(life\_expectancy$adult\_mortality))



The initial plot shows that there are three clusters in adult\_mortality, not two (as we thought above). The sqrt(adult\_mortality) distribution is more symmetric.

simpleFit\_sqrt = lm(data = life\_expectancy, lifeExpectancy ~ sqrt(adult\_mortality))  
summary(simpleFit\_sqrt)

##   
## Call:  
## lm(formula = lifeExpectancy ~ sqrt(adult\_mortality), data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34.311 -3.695 1.596 4.475 16.632   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 82.57292 0.37052 222.86 <2e-16 \*\*\*  
## sqrt(adult\_mortality) -1.08173 0.02923 -37.01 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.647 on 2193 degrees of freedom  
## Multiple R-squared: 0.3844, Adjusted R-squared: 0.3842   
## F-statistic: 1370 on 1 and 2193 DF, p-value: < 2.2e-16

Looking at the regression results, one can observe that intercept is 82.6, indicating that when the square-root of adult mortality is 0, the life expectancy is 82.6 years. The slope is -1.08 indicating that when the square root of adult mortality increases by 1 square-root percentage per 1000 population, the life expectancy decreases by 1.08 years.

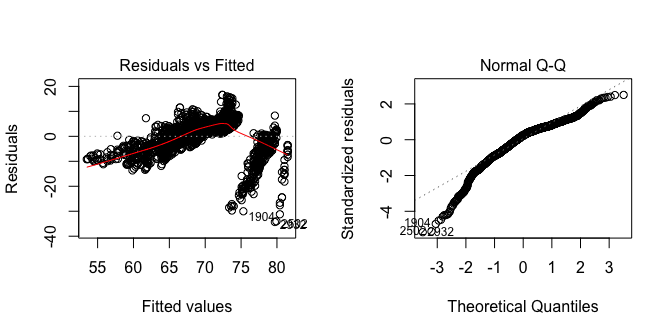
Let us now look at the confidence intervals and the residual plots:

confint(simpleFit\_sqrt)

## 2.5 % 97.5 %  
## (Intercept) 81.846316 83.299524  
## sqrt(adult\_mortality) -1.139054 -1.024413

The confidence interval of the slope is -1.14 to -1.02; this means that we are 95% confident that the mean life expectancy decreases between 1.02 to 1.14 years with a 1 square-root percent per 1000 increase in adult mortality.

par (mfrow = c(1,2))  
plot (simpleFit\_sqrt, which=1:2)



The residual plot shows non-linearity and non-constant variance. There are two clear clusters of points in the residual v/s fitted plot which were also reflected in the simple linear regression plot. There are more points beneath the 0.0 line and the red trend line shows curvature, indicating non-linearity. The quantile points in the center show a linear trend indicating normally-distributed residuals. However, there is deviation from the linear pattern at the left-tail indicating left-skewed residuals. Hence, the residuals do not follow the normal distribution condition.

## First Order Model

Next, we fit a first-order linear model will all nine predictors, with the transformations decided upon previously:

fit1 = lm(data = life\_expectancy, lifeExpectancy ~ sqrt(adult\_mortality) + infant\_deaths\_cat + hepatitisB + sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria + thin\_avg)  
summary(fit1)

##   
## Call:  
## lm(formula = lifeExpectancy ~ sqrt(adult\_mortality) + infant\_deaths\_cat +   
## hepatitisB + sqrt(alcohol) + measles\_cat + bmi + polio +   
## diphtheria + thin\_avg, data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31.048 -2.887 0.663 3.372 15.053   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.085e+01 8.374e-01 84.612 < 2e-16 \*\*\*  
## sqrt(adult\_mortality) -7.188e-01 2.595e-02 -27.702 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2 -7.815e-01 3.194e-01 -2.447 0.014477 \*   
## infant\_deaths\_catQ3 -2.530e+00 3.529e-01 -7.169 1.03e-12 \*\*\*  
## infant\_deaths\_catQ4 -5.722e+00 4.274e-01 -13.386 < 2e-16 \*\*\*  
## hepatitisB -1.078e-05 5.786e-03 -0.002 0.998514   
## sqrt(alcohol) 7.350e-01 1.185e-01 6.205 6.54e-10 \*\*\*  
## measles\_catQ2 1.456e-01 3.390e-01 0.429 0.667606   
## measles\_catQ3 7.102e-01 3.155e-01 2.251 0.024483 \*   
## measles\_catQ4 1.523e+00 3.712e-01 4.102 4.24e-05 \*\*\*  
## bmi 7.260e-02 7.082e-03 10.251 < 2e-16 \*\*\*  
## polio 2.260e-02 6.742e-03 3.353 0.000814 \*\*\*  
## diphtheria 4.630e-02 7.323e-03 6.323 3.10e-10 \*\*\*  
## thin\_avg -2.038e-01 3.470e-02 -5.873 4.95e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.317 on 2181 degrees of freedom  
## Multiple R-squared: 0.6084, Adjusted R-squared: 0.606   
## F-statistic: 260.6 on 13 and 2181 DF, p-value: < 2.2e-16

From the regression analysis, the following variables are significant with a 0.001 significance level: the intercept, sqrt(adult\_mort3), infant\_deaths\_cat, sqrt(alcohol), measles\_cat, bmi, polio, and diphtheria. HepatitisB is not significant. The (adjusted) R-squared is 60.6% which implies that this model explains 60.6% of the variation in the response variable. The residual standard error is 5.317 years.

### Interpretation of parameters:

NOTE: Due to the complexity and unclarity of the units, while we will be interpreting the parameter effects with units, we will also interpret it in a qualitative manner by describing the direction of effect.

The intercept 70.85 is the estimated mean life expectancy in age (years), when all the other predictors are 0.

When the square root of the probability of dying between 15 and 60 years per 1000 population increases by 1 square root percent, the life expectancy decreases by 0.7188 years. Hence, when adult mortality increases, the life expectancy decreases, while holding other predictors as constant.

If the number of infant deaths are greater than 0 and less than 3 life expectancy will decrease by 0.7815 years relative to when there are 0 infant deaths (first group), while holding other predictors constant. If the number of infant deaths are greater than 3 and less than 22 life expectancy will decrease by 2.530 years relative to when there are 0 infant deaths (first group), while holding other predictors constant. If the number of infant deaths are greater than 22 life expectancy will decrease by 5.722 years relative to when there are 0 infant deaths (first group), while holding other predictors constant. In overall, it can be seen as the number of infant deaths increase, the life expectancy keeps decreasing.

If hepatitis B immunization coverage among one-year olds increases by 1%, life expectancy will decrease by 1.078e-05 years, while holding other predictors constant.

As alcohol consumption increases by the square root of one square root litre of pure alcohol, the life expectancy increases by 0.735 years, while holding other predictors as constant.

If the number of reported measles cases increases by 0 to 17 cases, the life expectancy will increase by 0.1456 years, while holding other predictors as constant. If the number of reported measles cases increases by 17 to 360.2 cases, the life expectancy will increase by 0.7102 years, while holding other predictors as constant. If the number of reported measles cases increases by more than 360.2 cases, the life expectancy will increase by 1.523 years, while holding other predictors as constant. In general, it appears that as the number of measles cases increases, the life expectancy increases.

An increase in bmi by 1 kg/m^2 will increase life expectancy by 0.0726 years, while keeping other predictors constant.

An increase in polio immunization coverage among 1 year old by 1% will increase life expectancy by 0.02260 years, keeping other predictors constant.

An increase in diphtheria immunization coverage among 1 year old by 1%, will increase life expectancy by 0.0463, keeping other predictors constant.

An increase in the average thinness among 5 to 19 year olds will decrease life expectancy by 0.2038 years, while keeping other predictors constant.

anova(fit1)

## Analysis of Variance Table  
##   
## Response: lifeExpectancy  
## Df Sum Sq Mean Sq F value Pr(>F)   
## sqrt(adult\_mortality) 1 60514 60514 2140.8845 < 2.2e-16 \*\*\*  
## infant\_deaths\_cat 3 21496 7165 253.5002 < 2.2e-16 \*\*\*  
## hepatitisB 1 1283 1283 45.4043 2.044e-11 \*\*\*  
## sqrt(alcohol) 1 3974 3974 140.5982 < 2.2e-16 \*\*\*  
## measles\_cat 3 235 78 2.7719 0.04017 \*   
## bmi 1 5143 5143 181.9525 < 2.2e-16 \*\*\*  
## polio 1 1003 1003 35.4838 2.990e-09 \*\*\*  
## diphtheria 1 1137 1137 40.2316 2.734e-10 \*\*\*  
## thin\_avg 1 975 975 34.4878 4.947e-09 \*\*\*  
## Residuals 2181 61648 28   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The analysis of variance table (ANOVA table) suggests that all variables except the categorical measles variable are significant predictors.

Let us check the variance inflation factor to look for correlation among variables:

library(car)

## Loading required package: carData

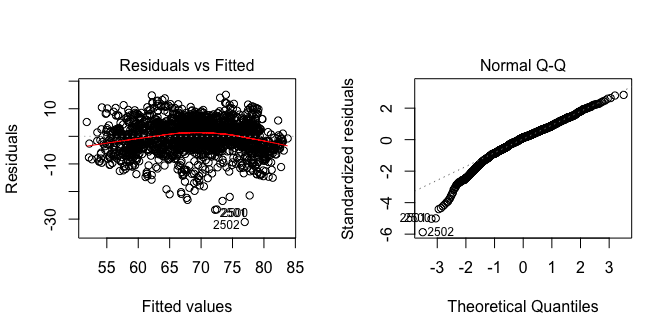
vif(fit1)

## GVIF Df GVIF^(1/(2\*Df))  
## sqrt(adult\_mortality) 1.231999 1 1.109954  
## infant\_deaths\_cat 1.999629 3 1.122427  
## hepatitisB 1.631066 1 1.277132  
## sqrt(alcohol) 1.311486 1 1.145201  
## measles\_cat 1.548812 3 1.075639  
## bmi 1.526217 1 1.235402  
## polio 1.636722 1 1.279344  
## diphtheria 1.890459 1 1.374940  
## thin\_avg 1.700353 1 1.303976

The VIF values of the variables (including the categorical variables) show that while there is some correlation between the predictors, it is low enough (less than 5) to not be a concern.

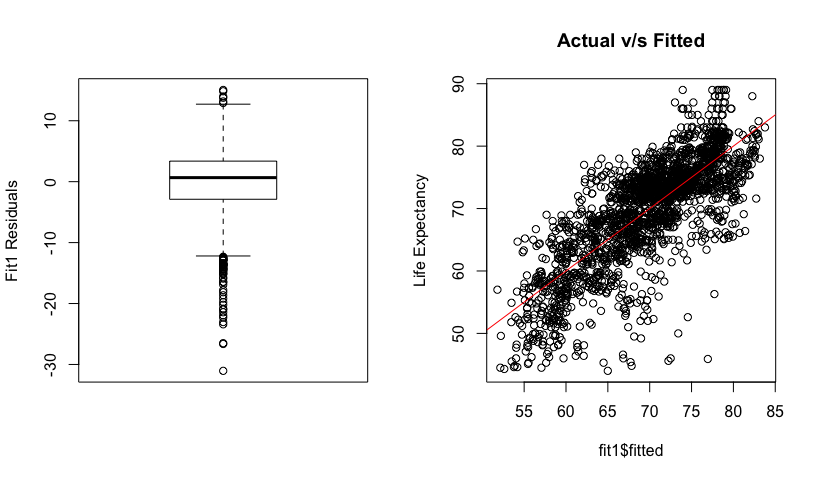
### Residual Analysis

par (mfrow = c(1,2))  
plot (fit1, which=1:2)



The residual v/s fitted value plot is useful for assessing linearity and constant variance conditions of a simple linear model. The residuals are scattered across the plot with high residuals and there appears to bee some curvature in the trend line, however, there does not appear to be an extreme non-linear trend in the residuals. Hence, the residuals are consistent with the linearity condition. Further, the residuals appear to have constant variance with points being scattered above and below the 0.0 line. The normal quantile plot is a scatterplot that displays the observed data v/s the values that would be expected from a normal sample of the same size. In this quantile plot, there does not appear to be any major deviation from a straight line. However, there are some left-skewed residuals that should be looked and might be improved with significant interaction effects.

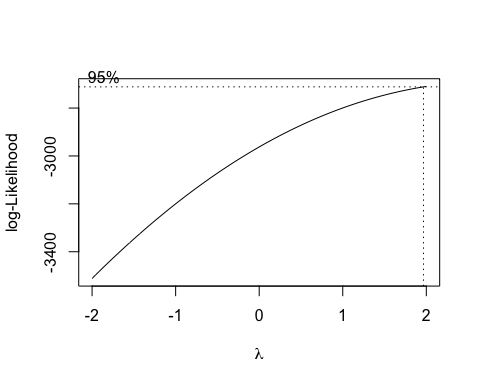
par(mfrow = c(1,2))  
boxplot(fit1$residuals, ylab = 'Fit1 Residuals')  
plot(lifeExpectancy~fit1$fitted, main = "Actual v/s Fitted", ylab = "Life Expectancy", data = life\_expectancy)  
abline(0,1, col = "red")



The box plot shows skewness in either ends of the residuals indicating some deviation from a normal pattern. The Actual v/s Fitted plot shows a moderate linear trend.

## Box-Cox Analysis

library(MASS)  
boxcox(fit1)



The box-cox analysis suggests a squared power transformation, with lambda = 2, on the response variable.

fit2 = lm(data = life\_expectancy, lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat + hepatitisB + sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria + thin\_avg)  
summary(fit2)

##   
## Call:  
## lm(formula = lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat +   
## hepatitisB + sqrt(alcohol) + measles\_cat + bmi + polio +   
## diphtheria + thin\_avg, data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3844.7 -414.7 63.4 442.9 2401.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5051.3667 111.4773 45.313 < 2e-16 \*\*\*  
## sqrt(adult\_mortality) -94.7458 3.4544 -27.427 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2 -95.1125 42.5148 -2.237 0.025377 \*   
## infant\_deaths\_catQ3 -376.0221 46.9719 -8.005 1.92e-15 \*\*\*  
## infant\_deaths\_catQ4 -752.2956 56.9019 -13.221 < 2e-16 \*\*\*  
## hepatitisB -0.2626 0.7703 -0.341 0.733202   
## sqrt(alcohol) 119.7871 15.7695 7.596 4.50e-14 \*\*\*  
## measles\_catQ2 48.1528 45.1290 1.067 0.286089   
## measles\_catQ3 148.8318 41.9993 3.544 0.000403 \*\*\*  
## measles\_catQ4 236.5485 49.4208 4.786 1.81e-06 \*\*\*  
## bmi 9.4200 0.9428 9.992 < 2e-16 \*\*\*  
## polio 3.0310 0.8975 3.377 0.000745 \*\*\*  
## diphtheria 6.2238 0.9749 6.384 2.10e-10 \*\*\*  
## thin\_avg -30.3409 4.6191 -6.569 6.33e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 707.8 on 2181 degrees of freedom  
## Multiple R-squared: 0.6127, Adjusted R-squared: 0.6104   
## F-statistic: 265.4 on 13 and 2181 DF, p-value: < 2.2e-16

From the regression analysis, all variables, except hepatitisB, are significant at the 0.001 level. The (adjusted) R-squared is 61.04% which means that this model explains 61.04% of the variation in the response variable. The residual standard error is 707.8 squared years.

anova(fit2)

## Analysis of Variance Table  
##   
## Response: lifeExpectancy^2  
## Df Sum Sq Mean Sq F value Pr(>F)   
## sqrt(adult\_mortality) 1 1071435008 1071435008 2138.9702 < 2.2e-16 \*\*\*  
## infant\_deaths\_cat 3 385944250 128648083 256.8279 < 2.2e-16 \*\*\*  
## hepatitisB 1 20385373 20385373 40.6965 2.164e-10 \*\*\*  
## sqrt(alcohol) 1 92461996 92461996 184.5874 < 2.2e-16 \*\*\*  
## measles\_cat 3 6812518 2270839 4.5334 0.003568 \*\*   
## bmi 1 91136603 91136603 181.9415 < 2.2e-16 \*\*\*  
## polio 1 17930225 17930225 35.7952 2.555e-09 \*\*\*  
## diphtheria 1 20557872 20557872 41.0409 1.821e-10 \*\*\*  
## thin\_avg 1 21612408 21612408 43.1462 6.332e-11 \*\*\*  
## Residuals 2181 1092488238 500912   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The analysis of variance table (ANOVA table) suggests that all variables are significant at the 0.001 level, except the measles\_cat variable which is significant at the 0.01 level (which is higher than the first fitted model).

Let us check the variance inflation factor to check for correlation among variables:

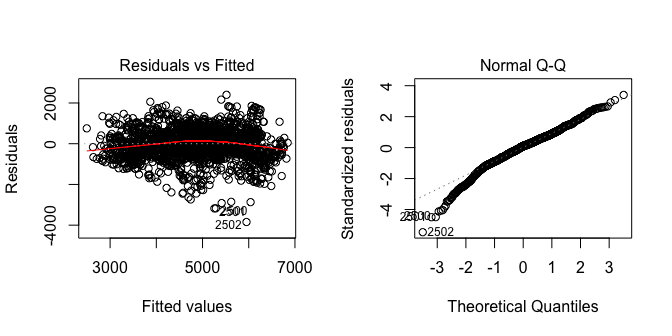
library(car)  
vif(fit2)

## GVIF Df GVIF^(1/(2\*Df))  
## sqrt(adult\_mortality) 1.231999 1 1.109954  
## infant\_deaths\_cat 1.999629 3 1.122427  
## hepatitisB 1.631066 1 1.277132  
## sqrt(alcohol) 1.311486 1 1.145201  
## measles\_cat 1.548812 3 1.075639  
## bmi 1.526217 1 1.235402  
## polio 1.636722 1 1.279344  
## diphtheria 1.890459 1 1.374940  
## thin\_avg 1.700353 1 1.303976

The VIF values of the variables (including the categorical variables) show that the variables in the model have low correlation with each other, since all values are below 5.

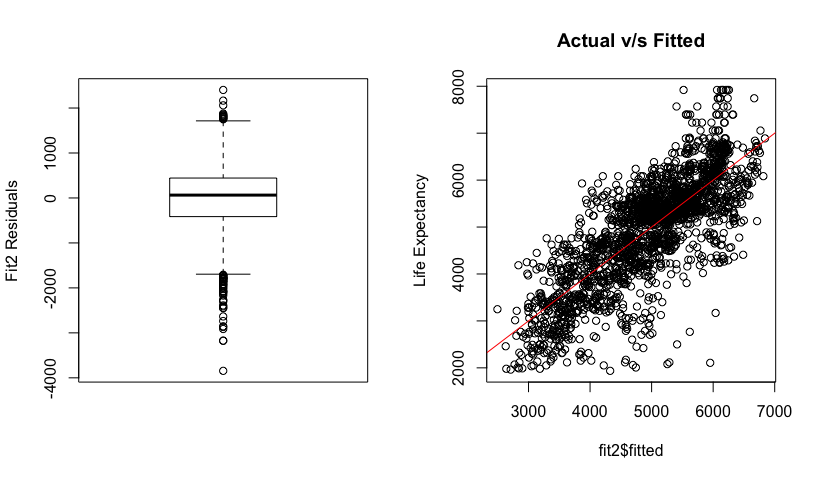
## Residual Analysis

par (mfrow = c(1,2))  
plot (fit2, which=1:2)



The residual v/s fitted value plot is useful for assessing linearity and constant variance conditions of a simple linear model. The residuals are scattered across the plot with high residuals at the bottom of the plot, however, the curvature is less than the previous model hence, there does not appear to be an extreme non-linear trend in the residuals. Thus, the residuals are consistent with the linearity condition. Further, there does not appear to be a major deviation from constant variance with points being scattered above and below the 0.0 line. The normal quantile plot is a scatterplot that displays the observed data v/s the values that would be expected from a normal sample of the same size. In this quantile plot, there does not appear to be any major deviation from a straight line. While there are no right-skewed residuals, there are some left-skewed residuals but the trend is largely linear indicating that the deviation might be a small variation that can be ignored.

par(mfrow = c(1,2))  
boxplot(fit2$residuals, ylab = 'Fit2 Residuals')  
plot(lifeExpectancy^2~fit2$fitted, main = "Actual v/s Fitted", ylab = "Life Expectancy", data = life\_expectancy)  
abline(0,1, col = "red")



The box plot shows slight skewness at the right end with large skewness on the left end, which was shown in the quantile plot, indicating some deviation from a normal pattern. The Actual v/s Fitted plot shows a stronger linear trend than the originally-fitted model.

## Backward Elimination

Looking at the regression results of the second fitted model, it can be concluded that Hepatitis B is not a significant predictor. Let us remove the variable and refit the model:

fit3 = lm(data = life\_expectancy, lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat + sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria + thin\_avg)  
summary(fit3)

##   
## Call:  
## lm(formula = lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat +   
## sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria +   
## thin\_avg, data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3826.1 -413.2 63.3 443.4 2399.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5047.0004 110.7166 45.585 < 2e-16 \*\*\*  
## sqrt(adult\_mortality) -94.7856 3.4518 -27.460 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2 -95.0049 42.5050 -2.235 0.025509 \*   
## infant\_deaths\_catQ3 -375.4832 46.9358 -8.000 2.00e-15 \*\*\*  
## infant\_deaths\_catQ4 -751.7965 56.8715 -13.219 < 2e-16 \*\*\*  
## sqrt(alcohol) 120.2154 15.7161 7.649 3.01e-14 \*\*\*  
## measles\_catQ2 48.0180 45.1182 1.064 0.287323   
## measles\_catQ3 148.8659 41.9907 3.545 0.000401 \*\*\*  
## measles\_catQ4 237.6588 49.3034 4.820 1.53e-06 \*\*\*  
## bmi 9.4118 0.9423 9.989 < 2e-16 \*\*\*  
## polio 2.9698 0.8792 3.378 0.000743 \*\*\*  
## diphtheria 6.0799 0.8786 6.920 5.93e-12 \*\*\*  
## thin\_avg -30.3288 4.6180 -6.567 6.38e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 707.6 on 2182 degrees of freedom  
## Multiple R-squared: 0.6127, Adjusted R-squared: 0.6105   
## F-statistic: 287.6 on 12 and 2182 DF, p-value: < 2.2e-16

From the regression analysis, all variables are significant with a 0.001 significance level. The (adjusted) R-squared is 61.05% which implies that this model explains 61.05% of the variation in the response variable. This is not significantly larger than the R-squared obtained before removing hepatitsB. The residual standard error is 707.6.

### Interpretation of parameters:

NOTE: Due to the complexity and unclarity of the units, we will be interpreting the parameter effects in a qualitative and quantitative manner by focusing on the direction of effect.

The intercept 5047 is the estimated mean life expectancy in years squared, when all the other predictors are 0.

When the square root of the probability of dying between 15 and 60 years per 1000 population increases by 1 square root percent, the life expectancy decreases by 94.79 years squared. Hence, when adult mortality increases, the life expectancy decreases, while holding other predictors constant.

If the number of infant deaths are greater than 0 and less than 3, life expectancy will decrease by 95.0049 years squared, while holding other predictors constant. If the number of infant deaths are greater than 3 and less than 22, life expectancy will decrease by 375.48 years squared, while holding other predictors as constant. If the number of infant deaths are greater than 22, life expectancy will decrease by 751.8 years squared, while holding other predictors constant. In overall, it can be seen as the number of infant deaths increase, the life expectancy will decrease.

As alcohol consumption increases by one square root litre of pure alcohol, the life expectancy increases by 120.22 squared years, while holding other predictors constant. Hence, when alcohol consumption increases, so does life expectancy.

If the number of reported measles cases increases by 0 to 17 cases, the life expectancy will increase by 48.02 squared years, while holding other predictors as constant. If the number of reported measles cases increases by 17 to 360.2 cases, the life expectancy will increase by 148.87 squared years, while holding other predictors as constant. If the number of reported measles cases increases by more than 360.2 cases, the life expectancy will increase by 237.67 squared years, while holding other predictors as constant. In general, it appears that as the number of measles cases increases, the life expectancy increases.

An increase in bmi by 1 kg/m^2 will increase life expectancy by 9.41 squared years, while keeping other predictors constant. Hence, as bmi increases, the life expectancy also increases.

An increase in polio immunization coverage among 1 year old by 1% will increase life expectancy by 2.97 squared years, keeping other predictors constant.

An increase in diphtheria immunization coverage among 1 year old by 1%, will increase life expectancy by 6.08 squared years, keeping other predictors constant.

An increase in the average thinness among 5 to 19 year olds will decrease life expectancy by 30.33 squared years, while keeping other predictors the constant.

The surprising parameters are alcohol and measles because as alcohol consumption increases and measles cases increase, the life expectancy increases instead of decreasing which is what the researchers expected.

anova(fit3)

## Analysis of Variance Table  
##   
## Response: lifeExpectancy^2  
## Df Sum Sq Mean Sq F value Pr(>F)   
## sqrt(adult\_mortality) 1 1071435008 1071435008 2139.8369 < 2.2e-16 \*\*\*  
## infant\_deaths\_cat 3 385944250 128648083 256.9320 < 2.2e-16 \*\*\*  
## sqrt(alcohol) 1 92742699 92742699 185.2229 < 2.2e-16 \*\*\*  
## measles\_cat 3 5425859 1808620 3.6121 0.01279 \*   
## bmi 1 95030472 95030472 189.7919 < 2.2e-16 \*\*\*  
## polio 1 31727915 31727915 63.3660 2.734e-15 \*\*\*  
## diphtheria 1 24315374 24315374 48.5619 4.222e-12 \*\*\*  
## thin\_avg 1 21596458 21596458 43.1318 6.377e-11 \*\*\*  
## Residuals 2182 1092546455 500709   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The analysis of variance table (ANOVA table) suggests that all variables except the categorical measles variable are significant predictors at the 0.001 level.

Let us check the variance inflation factor to check for correlation among variables:

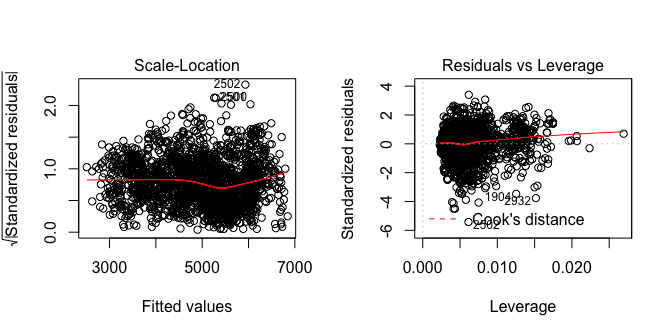
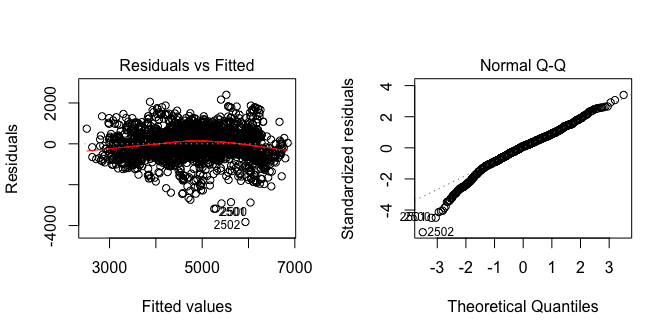
library(car)  
vif(fit3)

## GVIF Df GVIF^(1/(2\*Df))  
## sqrt(adult\_mortality) 1.230588 1 1.109319  
## infant\_deaths\_cat 1.996855 3 1.122168  
## sqrt(alcohol) 1.303162 1 1.141561  
## measles\_cat 1.539060 3 1.074507  
## bmi 1.525217 1 1.234997  
## polio 1.571279 1 1.253507  
## diphtheria 1.536260 1 1.239460  
## thin\_avg 1.700253 1 1.303937

The VIF values of the variables (including the categorical variables) show that the variables in the model have low correlation with each other, since all values are below 5.

### Residual Analysis

par (mfrow = c(1,2))  
plot(fit3)



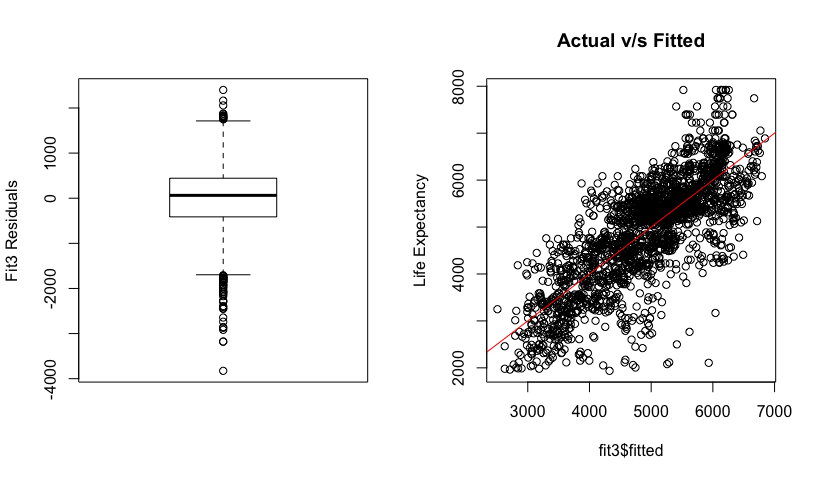
The residual v/s fitted value plot is useful for assessing linearity and constant variance conditions of a simple linear model. The residuals are scattered across the plot with high residuals at the bottom of the plot, however, there does not appear to be an extreme non-linear trend in the residuals. Hence, the residuals are consistent with the linearity condition. Further, there does not appear to be a major deviation from constant variance with points being scattered above and below the 0.0 line. The normal quantile plot is a scatterplot that displays the observed data v/s the values that would be expected from a normal sample of the same size. In this quantile plot, there does not appear to be any major deviation from a straight line. While there are no right-skewed residuals, there are some left-skewed residuals but the trend is largely linear indicating that the deviation might be a small variation that can be ignored.

The Scale-Location plot does not indicate any obvious concerns; however, there is some curvature in the trend line in the middle but the residuals appear to be largely clustered.

The Residuals vs. Leverage plot appears to show a large cluster of points to the left of the graph with some points being scattered to the right causing the trend line to bend upwards. Further, the standardized residuals are greater than 3 on both ends. This might be a cause of concern.

These plots are extremely similar to the plots produced by fit2 indicating that there has not been a large change by the removal of hepatitsB.

par(mfrow = c(1,2))  
boxplot(fit3$residuals, ylab = 'Fit3 Residuals')  
plot(lifeExpectancy^2~fit3$fitted, main = "Actual v/s Fitted", ylab = "Life Expectancy", data = life\_expectancy)  
abline(0,1, col = "red")

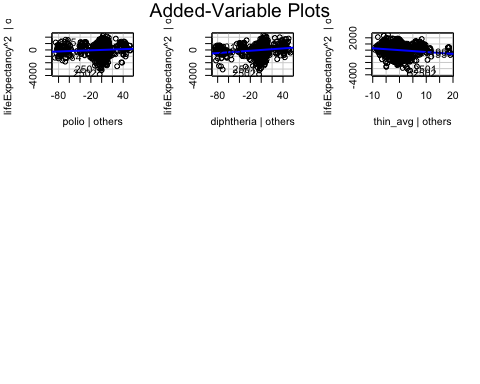
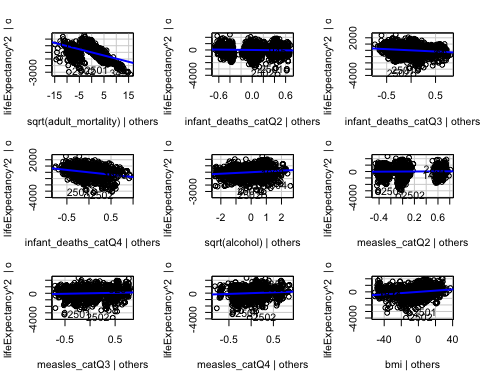


The box plot shows slight skewness at the right end with large skewness on the left end, which was shown in the quantile plot, indicating some deviation from a normal pattern. The Actual v/s Fitted plot shows a similar linear trend as the fit2 model.

## Added Variable Plots

Added variable plots produce all of the added variable plots - one for each quantitative or coded predictor variable - with one function call.

library(car)  
avPlots(fit3)



## Step-wise Regression

Next, let us apply step-wise regression to the fit3 model that has been obtained after applying a Box Cox plot and backward elimination.

fit3aic = step(fit3, direction = 'both')

## Start: AIC=28819.66  
## lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat +   
## sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria +   
## thin\_avg  
##   
## Df Sum of Sq RSS AIC  
## <none> 1092546455 28820  
## - polio 1 5713522 1098259977 28829  
## - measles\_cat 3 13187946 1105734400 28840  
## - thin\_avg 1 21596458 1114142913 28861  
## - diphtheria 1 23974693 1116521148 28865  
## - sqrt(alcohol) 1 29296335 1121842790 28876  
## - bmi 1 49956326 1142502781 28916  
## - infant\_deaths\_cat 3 101229563 1193776018 29008  
## - sqrt(adult\_mortality) 1 377563273 1470109728 29469

summary(fit3aic)

##   
## Call:  
## lm(formula = lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat +   
## sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria +   
## thin\_avg, data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3826.1 -413.2 63.3 443.4 2399.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5047.0004 110.7166 45.585 < 2e-16 \*\*\*  
## sqrt(adult\_mortality) -94.7856 3.4518 -27.460 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2 -95.0049 42.5050 -2.235 0.025509 \*   
## infant\_deaths\_catQ3 -375.4832 46.9358 -8.000 2.00e-15 \*\*\*  
## infant\_deaths\_catQ4 -751.7965 56.8715 -13.219 < 2e-16 \*\*\*  
## sqrt(alcohol) 120.2154 15.7161 7.649 3.01e-14 \*\*\*  
## measles\_catQ2 48.0180 45.1182 1.064 0.287323   
## measles\_catQ3 148.8659 41.9907 3.545 0.000401 \*\*\*  
## measles\_catQ4 237.6588 49.3034 4.820 1.53e-06 \*\*\*  
## bmi 9.4118 0.9423 9.989 < 2e-16 \*\*\*  
## polio 2.9698 0.8792 3.378 0.000743 \*\*\*  
## diphtheria 6.0799 0.8786 6.920 5.93e-12 \*\*\*  
## thin\_avg -30.3288 4.6180 -6.567 6.38e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 707.6 on 2182 degrees of freedom  
## Multiple R-squared: 0.6127, Adjusted R-squared: 0.6105   
## F-statistic: 287.6 on 12 and 2182 DF, p-value: < 2.2e-16

According to this result, there are no steps that can be taken to decrease the value of AIC.

Let us now try step-wise regression with the BIC criterion:

fit3bic = step(fit3, direction = "both", k = log(fit3$rank + fit3$df.residual))

## Start: AIC=28893.68  
## lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat +   
## sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria +   
## thin\_avg  
##   
## Df Sum of Sq RSS AIC  
## <none> 1092546455 28894  
## - measles\_cat 3 13187946 1105734400 28897  
## - polio 1 5713522 1098259977 28897  
## - thin\_avg 1 21596458 1114142913 28929  
## - diphtheria 1 23974693 1116521148 28934  
## - sqrt(alcohol) 1 29296335 1121842790 28944  
## - bmi 1 49956326 1142502781 28984  
## - infant\_deaths\_cat 3 101229563 1193776018 29065  
## - sqrt(adult\_mortality) 1 377563273 1470109728 29538

summary(fit3bic)

##   
## Call:  
## lm(formula = lifeExpectancy^2 ~ sqrt(adult\_mortality) + infant\_deaths\_cat +   
## sqrt(alcohol) + measles\_cat + bmi + polio + diphtheria +   
## thin\_avg, data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3826.1 -413.2 63.3 443.4 2399.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5047.0004 110.7166 45.585 < 2e-16 \*\*\*  
## sqrt(adult\_mortality) -94.7856 3.4518 -27.460 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2 -95.0049 42.5050 -2.235 0.025509 \*   
## infant\_deaths\_catQ3 -375.4832 46.9358 -8.000 2.00e-15 \*\*\*  
## infant\_deaths\_catQ4 -751.7965 56.8715 -13.219 < 2e-16 \*\*\*  
## sqrt(alcohol) 120.2154 15.7161 7.649 3.01e-14 \*\*\*  
## measles\_catQ2 48.0180 45.1182 1.064 0.287323   
## measles\_catQ3 148.8659 41.9907 3.545 0.000401 \*\*\*  
## measles\_catQ4 237.6588 49.3034 4.820 1.53e-06 \*\*\*  
## bmi 9.4118 0.9423 9.989 < 2e-16 \*\*\*  
## polio 2.9698 0.8792 3.378 0.000743 \*\*\*  
## diphtheria 6.0799 0.8786 6.920 5.93e-12 \*\*\*  
## thin\_avg -30.3288 4.6180 -6.567 6.38e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 707.6 on 2182 degrees of freedom  
## Multiple R-squared: 0.6127, Adjusted R-squared: 0.6105   
## F-statistic: 287.6 on 12 and 2182 DF, p-value: < 2.2e-16

The result is the same as the AIC step-wise regression indicating that no more steps can be taken to lower the BIC and AIC values.

Since, we conducted residual analysis previously on this fit, we will not be repeating that information.

## Centered Interaction Effects

Now, let us add all two-way interaction effects possible.

life\_expectancy$sqrt.ad.mort = sqrt(life\_expectancy$adult\_mortality)  
  
life\_expectancy$sqrt.ad.mort.c = life\_expectancy$sqrt.ad.mort - mean(life\_expectancy$sqrt.ad.mort)  
  
life\_expectancy$sqrt.alc = sqrt(life\_expectancy$alcohol)  
  
life\_expectancy$sqrt.alc.c = life\_expectancy$sqrt.alc - mean(life\_expectancy$sqrt.alc)  
  
life\_expectancy$bmi.c = life\_expectancy$bmi - mean(life\_expectancy$bmi)  
  
life\_expectancy$polio.c = life\_expectancy$polio - mean(life\_expectancy$polio)  
  
life\_expectancy$diphtheria.c = life\_expectancy$diphtheria - mean(life\_expectancy$diphtheria)  
  
life\_expectancy$thin.c = life\_expectancy$thin\_avg - mean(life\_expectancy$thin\_avg)

Linear model using the ‘centered\_fit’

centered\_fit = lm(data = life\_expectancy, lifeExpectancy^2 ~ (sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c + measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c)^2)  
summary(centered\_fit)

##   
## Call:  
## lm(formula = lifeExpectancy^2 ~ (sqrt.ad.mort.c + infant\_deaths\_cat +   
## sqrt.alc.c + measles\_cat + bmi.c + polio.c + diphtheria.c +   
## thin.c)^2, data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3068.49 -362.43 24.58 336.89 2108.76   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.972e+03 4.105e+01 121.127 < 2e-16 \*\*\*  
## sqrt.ad.mort.c -3.988e+01 8.063e+00 -4.946 8.16e-07 \*\*\*  
## infant\_deaths\_catQ2 -1.840e+02 6.002e+01 -3.065 0.002205 \*\*   
## infant\_deaths\_catQ3 -1.719e+02 6.695e+01 -2.567 0.010329 \*   
## infant\_deaths\_catQ4 -2.636e+02 1.262e+02 -2.089 0.036865 \*   
## sqrt.alc.c 1.484e+02 2.936e+01 5.053 4.73e-07 \*\*\*  
## measles\_catQ2 2.802e+02 7.363e+01 3.805 0.000146 \*\*\*  
## measles\_catQ3 6.038e+02 7.724e+01 7.817 8.42e-15 \*\*\*  
## measles\_catQ4 6.528e+02 1.922e+02 3.396 0.000697 \*\*\*  
## bmi.c -8.188e+00 1.923e+00 -4.257 2.16e-05 \*\*\*  
## polio.c 6.987e+00 2.151e+00 3.248 0.001179 \*\*   
## diphtheria.c 2.350e+00 1.868e+00 1.258 0.208605   
## thin.c -9.357e+01 1.241e+01 -7.539 6.97e-14 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ2 -4.423e+01 9.989e+00 -4.428 1.00e-05 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ3 -3.837e+01 1.090e+01 -3.520 0.000441 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ4 -1.800e+01 1.210e+01 -1.488 0.136920   
## sqrt.ad.mort.c:sqrt.alc.c -3.635e+00 3.343e+00 -1.088 0.276927   
## sqrt.ad.mort.c:measles\_catQ2 1.499e+00 9.659e+00 0.155 0.876722   
## sqrt.ad.mort.c:measles\_catQ3 -1.651e+00 9.264e+00 -0.178 0.858563   
## sqrt.ad.mort.c:measles\_catQ4 -2.782e+01 1.014e+01 -2.743 0.006132 \*\*   
## sqrt.ad.mort.c:bmi.c 8.042e-01 2.053e-01 3.917 9.26e-05 \*\*\*  
## sqrt.ad.mort.c:polio.c -1.991e-01 1.782e-01 -1.118 0.263888   
## sqrt.ad.mort.c:diphtheria.c -5.212e-01 1.738e-01 -2.999 0.002738 \*\*   
## sqrt.ad.mort.c:thin.c 1.612e+00 9.005e-01 1.790 0.073654 .   
## infant\_deaths\_catQ2:sqrt.alc.c -1.073e+02 4.088e+01 -2.624 0.008763 \*\*   
## infant\_deaths\_catQ3:sqrt.alc.c 2.342e+01 4.705e+01 0.498 0.618675   
## infant\_deaths\_catQ4:sqrt.alc.c -6.450e+01 5.707e+01 -1.130 0.258534   
## infant\_deaths\_catQ2:measles\_catQ2 -2.353e+02 9.784e+01 -2.405 0.016258 \*   
## infant\_deaths\_catQ3:measles\_catQ2 -4.205e+02 1.194e+02 -3.523 0.000436 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ2 -5.407e+02 1.739e+02 -3.110 0.001895 \*\*   
## infant\_deaths\_catQ2:measles\_catQ3 -4.084e+02 9.742e+01 -4.193 2.87e-05 \*\*\*  
## infant\_deaths\_catQ3:measles\_catQ3 -7.547e+02 1.133e+02 -6.664 3.39e-11 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ3 -1.010e+03 1.561e+02 -6.467 1.24e-10 \*\*\*  
## infant\_deaths\_catQ2:measles\_catQ4 -4.005e+02 2.062e+02 -1.943 0.052160 .   
## infant\_deaths\_catQ3:measles\_catQ4 -7.537e+02 2.127e+02 -3.543 0.000404 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ4 -7.302e+02 2.371e+02 -3.080 0.002095 \*\*   
## infant\_deaths\_catQ2:bmi.c 9.945e+00 2.223e+00 4.473 8.12e-06 \*\*\*  
## infant\_deaths\_catQ3:bmi.c 1.614e+01 2.590e+00 6.232 5.53e-10 \*\*\*  
## infant\_deaths\_catQ4:bmi.c 2.606e+01 3.020e+00 8.630 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2:polio.c -6.072e+00 2.633e+00 -2.306 0.021217 \*   
## infant\_deaths\_catQ3:polio.c -1.885e+00 2.837e+00 -0.664 0.506554   
## infant\_deaths\_catQ4:polio.c -8.468e+00 3.243e+00 -2.612 0.009073 \*\*   
## infant\_deaths\_catQ2:diphtheria.c 4.977e+00 2.409e+00 2.066 0.038968 \*   
## infant\_deaths\_catQ3:diphtheria.c 2.671e+00 2.700e+00 0.989 0.322617   
## infant\_deaths\_catQ4:diphtheria.c 9.167e+00 3.077e+00 2.979 0.002925 \*\*   
## infant\_deaths\_catQ2:thin.c -7.317e+01 1.559e+01 -4.695 2.84e-06 \*\*\*  
## infant\_deaths\_catQ3:thin.c 3.449e+01 1.709e+01 2.018 0.043753 \*   
## infant\_deaths\_catQ4:thin.c 5.136e+00 1.711e+01 0.300 0.764150   
## sqrt.alc.c:measles\_catQ2 -1.333e+02 4.289e+01 -3.109 0.001902 \*\*   
## sqrt.alc.c:measles\_catQ3 -9.492e+01 4.188e+01 -2.266 0.023533 \*   
## sqrt.alc.c:measles\_catQ4 -8.859e+01 4.961e+01 -1.786 0.074285 .   
## sqrt.alc.c:bmi.c -2.668e+00 8.068e-01 -3.307 0.000960 \*\*\*  
## sqrt.alc.c:polio.c -2.393e-02 9.074e-01 -0.026 0.978961   
## sqrt.alc.c:diphtheria.c -1.548e+00 8.633e-01 -1.793 0.073147 .   
## sqrt.alc.c:thin.c -4.943e+01 4.578e+00 -10.797 < 2e-16 \*\*\*  
## measles\_catQ2:bmi.c 7.134e+00 2.476e+00 2.881 0.004006 \*\*   
## measles\_catQ3:bmi.c 8.887e+00 2.238e+00 3.970 7.41e-05 \*\*\*  
## measles\_catQ4:bmi.c -1.796e+00 2.906e+00 -0.618 0.536535   
## measles\_catQ2:polio.c -1.060e+00 2.576e+00 -0.411 0.680894   
## measles\_catQ3:polio.c -2.742e-01 2.491e+00 -0.110 0.912383   
## measles\_catQ4:polio.c 7.268e+00 2.845e+00 2.554 0.010707 \*   
## measles\_catQ2:diphtheria.c 3.262e+00 2.998e+00 1.088 0.276667   
## measles\_catQ3:diphtheria.c 1.072e+00 2.217e+00 0.483 0.628929   
## measles\_catQ4:diphtheria.c -2.700e+00 2.681e+00 -1.007 0.314091   
## measles\_catQ2:thin.c 9.211e-01 1.672e+01 0.055 0.956072   
## measles\_catQ3:thin.c 2.146e+01 1.534e+01 1.399 0.162061   
## measles\_catQ4:thin.c 1.247e+01 1.645e+01 0.758 0.448653   
## bmi.c:polio.c 4.233e-02 5.262e-02 0.804 0.421233   
## bmi.c:diphtheria.c -2.060e-01 5.089e-02 -4.048 5.36e-05 \*\*\*  
## bmi.c:thin.c -1.902e+00 3.040e-01 -6.255 4.79e-10 \*\*\*  
## polio.c:diphtheria.c 1.242e-01 2.387e-02 5.202 2.16e-07 \*\*\*  
## polio.c:thin.c 6.833e-01 2.468e-01 2.769 0.005680 \*\*   
## diphtheria.c:thin.c -3.979e-01 2.548e-01 -1.562 0.118543   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 609.9 on 2122 degrees of freedom  
## Multiple R-squared: 0.7201, Adjusted R-squared: 0.7106   
## F-statistic: 75.84 on 72 and 2122 DF, p-value: < 2.2e-16

From the regression analysis, all non-interacted variables are significant with a 0.01 significance level. Multiple interaction effects are also significant. The (adjusted) R-squared is 71.06% which implies that this model explains 71.06% of the variation in the response variable. This is significantly larger than the R-squared obtained before centering and interacting the variables, where the adjusted R-squared was 61.05%. The residual standard error has decreased from 707.6 to 609.1.

### Step-wise Regression

Let us conduct the step-wise regression for the centered fit.

aic\_cent = step(centered\_fit, direction = 'both')

## Start: AIC=28226.4  
## lifeExpectancy^2 ~ (sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c)^2  
##   
## Df Sum of Sq RSS AIC  
## - measles\_cat:thin.c 3 934897 790369956 28223  
## - sqrt.alc.c:polio.c 1 259 789435318 28224  
## - bmi.c:polio.c 1 240745 789675805 28225  
## - measles\_cat:diphtheria.c 3 1720716 791155776 28225  
## - sqrt.ad.mort.c:sqrt.alc.c 1 439999 789875058 28226  
## - sqrt.ad.mort.c:polio.c 1 464623 789899683 28226  
## <none> 789435059 28226  
## - diphtheria.c:thin.c 1 907156 790342215 28227  
## - sqrt.ad.mort.c:thin.c 1 1191526 790626586 28228  
## - sqrt.alc.c:diphtheria.c 1 1195740 790630799 28228  
## - sqrt.alc.c:measles\_cat 3 4079995 793515055 28232  
## - infant\_deaths\_cat:sqrt.alc.c 3 4199510 793634569 28232  
## - polio.c:thin.c 1 2851509 792286568 28232  
## - infant\_deaths\_cat:diphtheria.c 3 4449524 793884583 28233  
## - infant\_deaths\_cat:polio.c 3 4484120 793919180 28233  
## - sqrt.ad.mort.c:measles\_cat 3 4519483 793954542 28233  
## - measles\_cat:polio.c 3 4642536 794077595 28233  
## - sqrt.ad.mort.c:diphtheria.c 1 3346512 792781572 28234  
## - sqrt.alc.c:bmi.c 1 4067338 793502398 28236  
## - sqrt.ad.mort.c:bmi.c 1 5706762 795141821 28240  
## - bmi.c:diphtheria.c 1 6094730 795529790 28241  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 9847866 799282925 28248  
## - measles\_cat:bmi.c 3 10027153 799462212 28248  
## - polio.c:diphtheria.c 1 10066995 799502054 28252  
## - bmi.c:thin.c 1 14556062 803991121 28264  
## - infant\_deaths\_cat:thin.c 3 19428098 808863157 28274  
## - infant\_deaths\_cat:measles\_cat 9 28473875 817908934 28286  
## - infant\_deaths\_cat:bmi.c 3 29717319 819152378 28302  
## - sqrt.alc.c:thin.c 1 43370899 832805958 28342  
##   
## Step: AIC=28222.99  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + measles\_cat:polio.c + measles\_cat:diphtheria.c +   
## bmi.c:polio.c + bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.alc.c:polio.c 1 1256 790371212 28221  
## - bmi.c:polio.c 1 209696 790579652 28222  
## - measles\_cat:diphtheria.c 3 1682986 792052942 28222  
## - sqrt.ad.mort.c:sqrt.alc.c 1 465034 790834990 28222  
## - sqrt.ad.mort.c:polio.c 1 484917 790854873 28222  
## <none> 790369956 28223  
## - diphtheria.c:thin.c 1 920557 791290513 28224  
## - sqrt.ad.mort.c:thin.c 1 1059610 791429566 28224  
## - sqrt.alc.c:diphtheria.c 1 1136491 791506447 28224  
## + measles\_cat:thin.c 3 934897 789435059 28226  
## - infant\_deaths\_cat:sqrt.alc.c 3 4007388 794377344 28228  
## - polio.c:thin.c 1 2896323 793266279 28229  
## - sqrt.ad.mort.c:measles\_cat 3 4362317 794732273 28229  
## - infant\_deaths\_cat:diphtheria.c 3 4417681 794787637 28229  
## - infant\_deaths\_cat:polio.c 3 4509333 794879289 28230  
## - measles\_cat:polio.c 3 4648150 795018106 28230  
## - sqrt.ad.mort.c:diphtheria.c 1 3388184 793758140 28230  
## - sqrt.alc.c:measles\_cat 3 5449764 795819720 28232  
## - sqrt.alc.c:bmi.c 1 4291474 794661430 28233  
## - sqrt.ad.mort.c:bmi.c 1 5891832 796261788 28237  
## - bmi.c:diphtheria.c 1 6121147 796491102 28238  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10026799 800396755 28245  
## - measles\_cat:bmi.c 3 10512288 800882244 28246  
## - polio.c:diphtheria.c 1 10313656 800683612 28250  
## - bmi.c:thin.c 1 15653449 806023405 28264  
## - infant\_deaths\_cat:thin.c 3 21121234 811491190 28275  
## - infant\_deaths\_cat:measles\_cat 9 28711172 819081128 28283  
## - infant\_deaths\_cat:bmi.c 3 31814954 822184910 28304  
## - sqrt.alc.c:thin.c 1 43012128 833382084 28337  
##   
## Step: AIC=28221  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## measles\_cat:polio.c + measles\_cat:diphtheria.c + bmi.c:polio.c +   
## bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - bmi.c:polio.c 1 212264 790583476 28220  
## - measles\_cat:diphtheria.c 3 1715735 792086947 28220  
## - sqrt.ad.mort.c:sqrt.alc.c 1 463870 790835082 28220  
## - sqrt.ad.mort.c:polio.c 1 488252 790859464 28220  
## <none> 790371212 28221  
## - diphtheria.c:thin.c 1 924851 791296064 28222  
## - sqrt.ad.mort.c:thin.c 1 1062921 791434133 28222  
## + sqrt.alc.c:polio.c 1 1256 790369956 28223  
## - sqrt.alc.c:diphtheria.c 1 1625147 791996359 28224  
## + measles\_cat:thin.c 3 935894 789435318 28224  
## - infant\_deaths\_cat:sqrt.alc.c 3 4014983 794386195 28226  
## - sqrt.ad.mort.c:measles\_cat 3 4361284 794732496 28227  
## - polio.c:thin.c 1 2966803 793338015 28227  
## - infant\_deaths\_cat:diphtheria.c 3 4425289 794796501 28227  
## - infant\_deaths\_cat:polio.c 3 4524750 794895962 28228  
## - measles\_cat:polio.c 3 4750260 795121473 28228  
## - sqrt.ad.mort.c:diphtheria.c 1 3387726 793758938 28228  
## - sqrt.alc.c:measles\_cat 3 5453053 795824266 28230  
## - sqrt.alc.c:bmi.c 1 4292073 794663286 28231  
## - sqrt.ad.mort.c:bmi.c 1 5897873 796269085 28235  
## - bmi.c:diphtheria.c 1 6129915 796501127 28236  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10032231 800403444 28243  
## - measles\_cat:bmi.c 3 10511464 800882676 28244  
## - polio.c:diphtheria.c 1 10398909 800770122 28248  
## - bmi.c:thin.c 1 15719222 806090435 28262  
## - infant\_deaths\_cat:thin.c 3 21251679 811622892 28273  
## - infant\_deaths\_cat:measles\_cat 9 28760269 819131482 28282  
## - infant\_deaths\_cat:bmi.c 3 31814599 822185811 28302  
## - sqrt.alc.c:thin.c 1 43015911 833387124 28335  
##   
## Step: AIC=28219.59  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## measles\_cat:polio.c + measles\_cat:diphtheria.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - measles\_cat:diphtheria.c 3 1646646 792230123 28218  
## - sqrt.ad.mort.c:sqrt.alc.c 1 487639 791071115 28219  
## - sqrt.ad.mort.c:polio.c 1 601254 791184730 28219  
## <none> 790583476 28220  
## - diphtheria.c:thin.c 1 829049 791412525 28220  
## - sqrt.ad.mort.c:thin.c 1 979713 791563189 28220  
## + bmi.c:polio.c 1 212264 790371212 28221  
## + sqrt.alc.c:polio.c 1 3824 790579652 28222  
## - sqrt.alc.c:diphtheria.c 1 1632997 792216474 28222  
## + measles\_cat:thin.c 3 905678 789677798 28223  
## - infant\_deaths\_cat:sqrt.alc.c 3 4020618 794604095 28225  
## - polio.c:thin.c 1 2765359 793348836 28225  
## - sqrt.ad.mort.c:measles\_cat 3 4345137 794928614 28226  
## - measles\_cat:polio.c 3 4617808 795201284 28226  
## - sqrt.ad.mort.c:diphtheria.c 1 3297337 793880813 28227  
## - infant\_deaths\_cat:diphtheria.c 3 4775542 795359018 28227  
## - infant\_deaths\_cat:polio.c 3 5003426 795586902 28227  
## - sqrt.alc.c:measles\_cat 3 5442688 796026165 28229  
## - sqrt.alc.c:bmi.c 1 4323745 794907221 28230  
## - sqrt.ad.mort.c:bmi.c 1 5847840 796431316 28234  
## - bmi.c:diphtheria.c 1 6860238 797443714 28237  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 9952419 800535896 28241  
## - measles\_cat:bmi.c 3 10590025 801173501 28243  
## - polio.c:diphtheria.c 1 10256518 800839994 28246  
## - bmi.c:thin.c 1 15799734 806383211 28261  
## - infant\_deaths\_cat:thin.c 3 21045709 811629186 28271  
## - infant\_deaths\_cat:measles\_cat 9 28661404 819244881 28280  
## - infant\_deaths\_cat:bmi.c 3 31923891 822507367 28300  
## - sqrt.alc.c:thin.c 1 43197076 833780553 28334  
##   
## Step: AIC=28218.15  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## measles\_cat:polio.c + bmi.c:diphtheria.c + bmi.c:thin.c +   
## polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.ad.mort.c:sqrt.alc.c 1 540736 792770859 28218  
## - sqrt.ad.mort.c:polio.c 1 548463 792778586 28218  
## <none> 792230123 28218  
## - sqrt.ad.mort.c:thin.c 1 951537 793181660 28219  
## - diphtheria.c:thin.c 1 1180225 793410348 28219  
## + measles\_cat:diphtheria.c 3 1646646 790583476 28220  
## + bmi.c:polio.c 1 143176 792086947 28220  
## + sqrt.alc.c:polio.c 1 41696 792188427 28220  
## - sqrt.alc.c:diphtheria.c 1 1837031 794067154 28221  
## - measles\_cat:polio.c 3 3292380 795522502 28221  
## + measles\_cat:thin.c 3 880862 791349261 28222  
## - infant\_deaths\_cat:sqrt.alc.c 3 3899014 796129137 28223  
## - sqrt.ad.mort.c:measles\_cat 3 4090112 796320234 28224  
## - polio.c:thin.c 1 2976492 795206615 28224  
## - infant\_deaths\_cat:diphtheria.c 3 4850933 797081056 28226  
## - infant\_deaths\_cat:polio.c 3 4881479 797111602 28226  
## - sqrt.ad.mort.c:diphtheria.c 1 3464835 795694958 28226  
## - sqrt.alc.c:measles\_cat 3 5258893 797489016 28227  
## - sqrt.alc.c:bmi.c 1 4454286 796684408 28228  
## - sqrt.ad.mort.c:bmi.c 1 5776295 798006418 28232  
## - bmi.c:diphtheria.c 1 7074320 799304443 28236  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10009558 802239681 28240  
## - measles\_cat:bmi.c 3 10733949 802964072 28242  
## - polio.c:diphtheria.c 1 9899385 802129508 28243  
## - bmi.c:thin.c 1 15826884 808057007 28260  
## - infant\_deaths\_cat:thin.c 3 20701998 812932121 28269  
## - infant\_deaths\_cat:measles\_cat 9 28718915 820949038 28278  
## - infant\_deaths\_cat:bmi.c 3 32393920 824624043 28300  
## - sqrt.alc.c:thin.c 1 43364408 835594531 28333  
##   
## Step: AIC=28217.65  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:measles\_cat + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c +   
## sqrt.ad.mort.c:diphtheria.c + sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c +   
## infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## measles\_cat:polio.c + bmi.c:diphtheria.c + bmi.c:thin.c +   
## polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.ad.mort.c:polio.c 1 645145 793416004 28217  
## <none> 792770859 28218  
## + sqrt.ad.mort.c:sqrt.alc.c 1 540736 792230123 28218  
## + measles\_cat:diphtheria.c 3 1699744 791071115 28219  
## - diphtheria.c:thin.c 1 1223004 793993863 28219  
## + bmi.c:polio.c 1 162661 792608198 28219  
## - sqrt.ad.mort.c:thin.c 1 1287379 794058239 28219  
## + sqrt.alc.c:polio.c 1 32433 792738426 28220  
## - sqrt.alc.c:diphtheria.c 1 1650049 794420908 28220  
## - measles\_cat:polio.c 3 3248900 796019759 28221  
## + measles\_cat:thin.c 3 904452 791866407 28221  
## - infant\_deaths\_cat:sqrt.alc.c 3 4082663 796853522 28223  
## - sqrt.ad.mort.c:measles\_cat 3 4281889 797052748 28224  
## - polio.c:thin.c 1 3063145 795834004 28224  
## - infant\_deaths\_cat:polio.c 3 4711396 797482256 28225  
## - infant\_deaths\_cat:diphtheria.c 3 4877422 797648281 28225  
## - sqrt.ad.mort.c:diphtheria.c 1 3443140 796213999 28225  
## - sqrt.alc.c:measles\_cat 3 5101871 797872730 28226  
## - sqrt.alc.c:bmi.c 1 4283039 797053899 28228  
## - sqrt.ad.mort.c:bmi.c 1 5439552 798210411 28231  
## - bmi.c:diphtheria.c 1 7200566 799971425 28236  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 9711380 802482240 28238  
## - measles\_cat:bmi.c 3 10772462 803543321 28241  
## - polio.c:diphtheria.c 1 9743390 802514249 28242  
## - bmi.c:thin.c 1 16831740 809602599 28262  
## - infant\_deaths\_cat:thin.c 3 20799877 813570736 28268  
## - infant\_deaths\_cat:measles\_cat 9 28689256 821460115 28278  
## - infant\_deaths\_cat:bmi.c 3 33090999 825861858 28301  
## - sqrt.alc.c:thin.c 1 46078258 838849117 28340  
##   
## Step: AIC=28217.44  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:measles\_cat + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## measles\_cat:polio.c + bmi.c:diphtheria.c + bmi.c:thin.c +   
## polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## <none> 793416004 28217  
## + sqrt.ad.mort.c:polio.c 1 645145 792770859 28218  
## + sqrt.ad.mort.c:sqrt.alc.c 1 637418 792778586 28218  
## - diphtheria.c:thin.c 1 1014426 794430430 28218  
## + bmi.c:polio.c 1 270972 793145032 28219  
## - sqrt.ad.mort.c:thin.c 1 1220537 794636542 28219  
## + measles\_cat:diphtheria.c 3 1646458 791769546 28219  
## + sqrt.alc.c:polio.c 1 48300 793367704 28219  
## - sqrt.alc.c:diphtheria.c 1 1678111 795094115 28220  
## - measles\_cat:polio.c 3 3207854 796623858 28220  
## + measles\_cat:thin.c 3 920729 792495275 28221  
## - infant\_deaths\_cat:sqrt.alc.c 3 3919441 797335446 28222  
## - polio.c:thin.c 1 2610582 796026586 28223  
## - sqrt.ad.mort.c:measles\_cat 3 4114496 797530500 28223  
## - sqrt.alc.c:measles\_cat 3 5035552 798451557 28225  
## - infant\_deaths\_cat:polio.c 3 5195124 798611128 28226  
## - infant\_deaths\_cat:diphtheria.c 3 5349745 798765749 28226  
## - sqrt.alc.c:bmi.c 1 4147395 797563400 28227  
## - sqrt.ad.mort.c:bmi.c 1 5250998 798667002 28230  
## - bmi.c:diphtheria.c 1 7283238 800699243 28236  
## - sqrt.ad.mort.c:diphtheria.c 1 7739869 801155873 28237  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 9799593 803215598 28238  
## - measles\_cat:bmi.c 3 10594886 804010890 28241  
## - polio.c:diphtheria.c 1 10025066 803441070 28243  
## - bmi.c:thin.c 1 16523910 809939915 28261  
## - infant\_deaths\_cat:thin.c 3 20695739 814111743 28268  
## - infant\_deaths\_cat:measles\_cat 9 28520807 821936811 28277  
## - infant\_deaths\_cat:bmi.c 3 33224814 826640818 28302  
## - sqrt.alc.c:thin.c 1 45795307 839211312 28339

summary(aic\_cent)

##   
## Call:  
## lm(formula = lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat +   
## sqrt.alc.c + measles\_cat + bmi.c + polio.c + diphtheria.c +   
## thin.c + sqrt.ad.mort.c:infant\_deaths\_cat + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c + sqrt.ad.mort.c:thin.c +   
## infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## measles\_cat:polio.c + bmi.c:diphtheria.c + bmi.c:thin.c +   
## polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c,   
## data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3121.02 -361.30 26.41 338.49 2170.86   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4969.14988 40.50417 122.682 < 2e-16 \*\*\*  
## sqrt.ad.mort.c -41.44248 8.02353 -5.165 2.63e-07 \*\*\*  
## infant\_deaths\_catQ2 -171.76705 59.47794 -2.888 0.003917 \*\*   
## infant\_deaths\_catQ3 -157.09138 64.93156 -2.419 0.015632 \*   
## infant\_deaths\_catQ4 -261.37859 125.80766 -2.078 0.037865 \*   
## sqrt.alc.c 153.87422 28.90459 5.324 1.12e-07 \*\*\*  
## measles\_catQ2 283.39491 72.48703 3.910 9.53e-05 \*\*\*  
## measles\_catQ3 588.66405 76.53825 7.691 2.21e-14 \*\*\*  
## measles\_catQ4 649.24380 191.75384 3.386 0.000723 \*\*\*  
## bmi.c -7.79370 1.86703 -4.174 3.11e-05 \*\*\*  
## polio.c 7.21560 2.08600 3.459 0.000553 \*\*\*  
## diphtheria.c 2.50705 1.73159 1.448 0.147811   
## thin.c -88.60409 11.05874 -8.012 1.84e-15 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ2 -43.84784 9.96212 -4.401 1.13e-05 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ3 -36.16116 10.62534 -3.403 0.000678 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ4 -15.95806 11.94276 -1.336 0.181622   
## sqrt.ad.mort.c:measles\_catQ2 0.19007 9.38168 0.020 0.983838   
## sqrt.ad.mort.c:measles\_catQ3 -0.86723 9.06236 -0.096 0.923771   
## sqrt.ad.mort.c:measles\_catQ4 -26.27850 9.96842 -2.636 0.008445 \*\*   
## sqrt.ad.mort.c:bmi.c 0.76159 0.20275 3.756 0.000177 \*\*\*  
## sqrt.ad.mort.c:diphtheria.c -0.65428 0.14347 -4.560 5.39e-06 \*\*\*  
## sqrt.ad.mort.c:thin.c 1.58262 0.87389 1.811 0.070281 .   
## infant\_deaths\_catQ2:sqrt.alc.c -105.78635 39.99118 -2.645 0.008223 \*\*   
## infant\_deaths\_catQ3:sqrt.alc.c 17.20538 45.25386 0.380 0.703837   
## infant\_deaths\_catQ4:sqrt.alc.c -63.17887 52.63899 -1.200 0.230184   
## infant\_deaths\_catQ2:measles\_catQ2 -246.98862 97.54090 -2.532 0.011408 \*   
## infant\_deaths\_catQ3:measles\_catQ2 -419.72612 117.89951 -3.560 0.000379 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ2 -542.08317 173.46006 -3.125 0.001801 \*\*   
## infant\_deaths\_catQ2:measles\_catQ3 -407.13945 97.24623 -4.187 2.95e-05 \*\*\*  
## infant\_deaths\_catQ3:measles\_catQ3 -751.25541 112.33472 -6.688 2.89e-11 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ3 -993.06890 155.11524 -6.402 1.88e-10 \*\*\*  
## infant\_deaths\_catQ2:measles\_catQ4 -404.38071 205.46790 -1.968 0.049186 \*   
## infant\_deaths\_catQ3:measles\_catQ4 -758.18427 211.00898 -3.593 0.000334 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ4 -723.07984 236.70055 -3.055 0.002280 \*\*   
## infant\_deaths\_catQ2:bmi.c 9.89863 2.21391 4.471 8.19e-06 \*\*\*  
## infant\_deaths\_catQ3:bmi.c 16.10524 2.57217 6.261 4.60e-10 \*\*\*  
## infant\_deaths\_catQ4:bmi.c 26.39537 2.86276 9.220 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2:polio.c -6.66727 2.57793 -2.586 0.009767 \*\*   
## infant\_deaths\_catQ3:polio.c -2.49541 2.66541 -0.936 0.349265   
## infant\_deaths\_catQ4:polio.c -8.86005 2.92068 -3.034 0.002446 \*\*   
## infant\_deaths\_catQ2:diphtheria.c 5.42753 2.32091 2.339 0.019452 \*   
## infant\_deaths\_catQ3:diphtheria.c 2.59763 2.47497 1.050 0.294040   
## infant\_deaths\_catQ4:diphtheria.c 8.63787 2.54463 3.395 0.000700 \*\*\*  
## infant\_deaths\_catQ2:thin.c -70.13276 14.78884 -4.742 2.25e-06 \*\*\*  
## infant\_deaths\_catQ3:thin.c 37.54738 16.09370 2.333 0.019738 \*   
## infant\_deaths\_catQ4:thin.c 9.99155 13.38752 0.746 0.455549   
## sqrt.alc.c:measles\_catQ2 -123.53456 38.91409 -3.175 0.001522 \*\*   
## sqrt.alc.c:measles\_catQ3 -116.67668 38.22054 -3.053 0.002296 \*\*   
## sqrt.alc.c:measles\_catQ4 -100.64745 45.55086 -2.210 0.027242 \*   
## sqrt.alc.c:bmi.c -2.67705 0.80191 -3.338 0.000857 \*\*\*  
## sqrt.alc.c:diphtheria.c -1.54079 0.72559 -2.124 0.033826 \*   
## sqrt.alc.c:thin.c -49.96785 4.50440 -11.093 < 2e-16 \*\*\*  
## measles\_catQ2:bmi.c 7.12168 2.27703 3.128 0.001786 \*\*   
## measles\_catQ3:bmi.c 7.69327 2.09306 3.676 0.000243 \*\*\*  
## measles\_catQ4:bmi.c -2.61219 2.67112 -0.978 0.328217   
## measles\_catQ2:polio.c 0.63405 2.17896 0.291 0.771088   
## measles\_catQ3:polio.c 0.15348 2.14213 0.072 0.942887   
## measles\_catQ4:polio.c 5.50880 2.44166 2.256 0.024161 \*   
## bmi.c:diphtheria.c -0.19160 0.04331 -4.424 1.02e-05 \*\*\*  
## bmi.c:thin.c -1.97897 0.29699 -6.663 3.39e-11 \*\*\*  
## polio.c:diphtheria.c 0.12166 0.02344 5.190 2.30e-07 \*\*\*  
## polio.c:thin.c 0.59579 0.22495 2.649 0.008143 \*\*   
## diphtheria.c:thin.c -0.40375 0.24455 -1.651 0.098881 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 610 on 2132 degrees of freedom  
## Multiple R-squared: 0.7187, Adjusted R-squared: 0.7105   
## F-statistic: 87.87 on 62 and 2132 DF, p-value: < 2.2e-16

The AIC criterion dropped the following predictors from the model: measles\_cat:thin.c sqrt.alc.c:polio.c bmi.c:polio.c measles\_cat:diphtheria.c sqrt.ad.mort.c:thin.c sqrt.ad.mort.c:polio.c

From the regression analysis, all variables are significant with a 0.01 significance level. The (adjusted) R-squared is 71.05% which implies that this model explains 71.05% of the variation in the response variable. This is smaller than the R-squared obtained before applying the step-wise regression. The residual standard error is 610; it is larger than the residual standard error obtained previously by one.

bic\_cent = step(centered\_fit, direction = "both", k = log(centered\_fit$rank + centered\_fit$df.residual))

## Start: AIC=28642.05  
## lifeExpectancy^2 ~ (sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c)^2  
##   
## Df Sum of Sq RSS AIC  
## - measles\_cat:thin.c 3 934897 790369956 28622  
## - measles\_cat:diphtheria.c 3 1720716 791155776 28624  
## - sqrt.alc.c:measles\_cat 3 4079995 793515055 28630  
## - infant\_deaths\_cat:sqrt.alc.c 3 4199510 793634569 28631  
## - infant\_deaths\_cat:diphtheria.c 3 4449524 793884583 28631  
## - infant\_deaths\_cat:polio.c 3 4484120 793919180 28631  
## - sqrt.ad.mort.c:measles\_cat 3 4519483 793954542 28632  
## - measles\_cat:polio.c 3 4642536 794077595 28632  
## - sqrt.alc.c:polio.c 1 259 789435318 28634  
## - bmi.c:polio.c 1 240745 789675805 28635  
## - sqrt.ad.mort.c:sqrt.alc.c 1 439999 789875058 28636  
## - sqrt.ad.mort.c:polio.c 1 464623 789899683 28636  
## - diphtheria.c:thin.c 1 907156 790342215 28637  
## - sqrt.ad.mort.c:thin.c 1 1191526 790626586 28638  
## - sqrt.alc.c:diphtheria.c 1 1195740 790630799 28638  
## <none> 789435059 28642  
## - polio.c:thin.c 1 2851509 792286568 28642  
## - sqrt.ad.mort.c:diphtheria.c 1 3346512 792781572 28644  
## - sqrt.alc.c:bmi.c 1 4067338 793502398 28646  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 9847866 799282925 28646  
## - measles\_cat:bmi.c 3 10027153 799462212 28647  
## - sqrt.ad.mort.c:bmi.c 1 5706762 795141821 28650  
## - infant\_deaths\_cat:measles\_cat 9 28473875 817908934 28651  
## - bmi.c:diphtheria.c 1 6094730 795529790 28651  
## - polio.c:diphtheria.c 1 10066995 799502054 28662  
## - infant\_deaths\_cat:thin.c 3 19428098 808863157 28672  
## - bmi.c:thin.c 1 14556062 803991121 28674  
## - infant\_deaths\_cat:bmi.c 3 29717319 819152378 28700  
## - sqrt.alc.c:thin.c 1 43370899 832805958 28752  
##   
## Step: AIC=28621.57  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + measles\_cat:polio.c + measles\_cat:diphtheria.c +   
## bmi.c:polio.c + bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - measles\_cat:diphtheria.c 3 1682986 792052942 28603  
## - infant\_deaths\_cat:sqrt.alc.c 3 4007388 794377344 28610  
## - sqrt.ad.mort.c:measles\_cat 3 4362317 794732273 28611  
## - infant\_deaths\_cat:diphtheria.c 3 4417681 794787637 28611  
## - infant\_deaths\_cat:polio.c 3 4509333 794879289 28611  
## - measles\_cat:polio.c 3 4648150 795018106 28611  
## - sqrt.alc.c:measles\_cat 3 5449764 795819720 28614  
## - sqrt.alc.c:polio.c 1 1256 790371212 28614  
## - bmi.c:polio.c 1 209696 790579652 28614  
## - sqrt.ad.mort.c:sqrt.alc.c 1 465034 790834990 28615  
## - sqrt.ad.mort.c:polio.c 1 484917 790854873 28615  
## - diphtheria.c:thin.c 1 920557 791290513 28616  
## - sqrt.ad.mort.c:thin.c 1 1059610 791429566 28617  
## - sqrt.alc.c:diphtheria.c 1 1136491 791506447 28617  
## <none> 790369956 28622  
## - polio.c:thin.c 1 2896323 793266279 28622  
## - sqrt.ad.mort.c:diphtheria.c 1 3388184 793758140 28623  
## - sqrt.alc.c:bmi.c 1 4291474 794661430 28626  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10026799 800396755 28626  
## - measles\_cat:bmi.c 3 10512288 800882244 28628  
## - sqrt.ad.mort.c:bmi.c 1 5891832 796261788 28630  
## - infant\_deaths\_cat:measles\_cat 9 28711172 819081128 28631  
## - bmi.c:diphtheria.c 1 6121147 796491102 28631  
## + measles\_cat:thin.c 3 934897 789435059 28642  
## - polio.c:diphtheria.c 1 10313656 800683612 28642  
## - infant\_deaths\_cat:thin.c 3 21121234 811491190 28656  
## - bmi.c:thin.c 1 15653449 806023405 28657  
## - infant\_deaths\_cat:bmi.c 3 31814954 822184910 28685  
## - sqrt.alc.c:thin.c 1 43012128 833382084 28730  
##   
## Step: AIC=28603.16  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + measles\_cat:polio.c + bmi.c:polio.c +   
## bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - measles\_cat:polio.c 3 3282751 795335693 28589  
## - infant\_deaths\_cat:sqrt.alc.c 3 3872365 795925307 28591  
## - sqrt.ad.mort.c:measles\_cat 3 4114371 796167314 28591  
## - infant\_deaths\_cat:polio.c 3 4417429 796470372 28592  
## - infant\_deaths\_cat:diphtheria.c 3 4466755 796519698 28592  
## - sqrt.alc.c:measles\_cat 3 5249689 797302631 28595  
## - sqrt.alc.c:polio.c 1 34005 792086947 28596  
## - bmi.c:polio.c 1 135485 792188427 28596  
## - sqrt.ad.mort.c:polio.c 1 445376 792498318 28597  
## - sqrt.ad.mort.c:sqrt.alc.c 1 529845 792582787 28597  
## - sqrt.ad.mort.c:thin.c 1 1006492 793059434 28598  
## - sqrt.alc.c:diphtheria.c 1 1113092 793166034 28598  
## - diphtheria.c:thin.c 1 1244051 793296994 28599  
## <none> 792052942 28603  
## - polio.c:thin.c 1 2930826 794983768 28604  
## - sqrt.ad.mort.c:diphtheria.c 1 3556332 795609274 28605  
## - sqrt.alc.c:bmi.c 1 4426646 796479589 28608  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10046990 802099932 28608  
## - measles\_cat:bmi.c 3 10679673 802732615 28610  
## - sqrt.ad.mort.c:bmi.c 1 5805012 797857954 28612  
## - infant\_deaths\_cat:measles\_cat 9 28697935 820750877 28612  
## - bmi.c:diphtheria.c 1 6069823 798122765 28612  
## + measles\_cat:diphtheria.c 3 1682986 790369956 28622  
## - polio.c:diphtheria.c 1 9996603 802049545 28623  
## + measles\_cat:thin.c 3 897167 791155776 28624  
## - infant\_deaths\_cat:thin.c 3 20637281 812690224 28636  
## - bmi.c:thin.c 1 15769797 807822739 28639  
## - infant\_deaths\_cat:bmi.c 3 32092406 824145349 28667  
## - sqrt.alc.c:thin.c 1 43258837 835311780 28712  
##   
## Step: AIC=28589.15  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:polio.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - infant\_deaths\_cat:polio.c 3 3186769 798522463 28575  
## - infant\_deaths\_cat:diphtheria.c 3 4215634 799551328 28578  
## - infant\_deaths\_cat:sqrt.alc.c 3 4474507 799810200 28578  
## - sqrt.ad.mort.c:measles\_cat 3 5067874 800403567 28580  
## - sqrt.alc.c:measles\_cat 3 5149854 800485547 28580  
## - bmi.c:polio.c 1 68532 795404225 28582  
## - sqrt.alc.c:polio.c 1 108803 795444496 28582  
## - sqrt.ad.mort.c:polio.c 1 425427 795761120 28583  
## - sqrt.ad.mort.c:sqrt.alc.c 1 499080 795834773 28583  
## - sqrt.alc.c:diphtheria.c 1 810945 796146638 28584  
## - diphtheria.c:thin.c 1 1116128 796451821 28584  
## - sqrt.ad.mort.c:thin.c 1 1145178 796480871 28585  
## <none> 795335693 28589  
## - polio.c:thin.c 1 3538909 798874603 28591  
## - sqrt.ad.mort.c:diphtheria.c 1 3729328 799065021 28592  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10051475 805387168 28594  
## - sqrt.alc.c:bmi.c 1 4458297 799793991 28594  
## - measles\_cat:bmi.c 3 10114267 805449960 28594  
## - bmi.c:diphtheria.c 1 5705618 801041311 28597  
## - infant\_deaths\_cat:measles\_cat 9 28523065 823858758 28597  
## - sqrt.ad.mort.c:bmi.c 1 6224841 801560534 28599  
## + measles\_cat:polio.c 3 3282751 792052942 28603  
## + measles\_cat:thin.c 3 950680 794385013 28610  
## - polio.c:diphtheria.c 1 10732282 806067975 28611  
## + measles\_cat:diphtheria.c 3 317587 795018106 28611  
## - infant\_deaths\_cat:thin.c 3 21307063 816642756 28624  
## - bmi.c:thin.c 1 15618464 810954157 28624  
## - infant\_deaths\_cat:bmi.c 3 31597361 826933054 28652  
## - sqrt.alc.c:thin.c 1 43228822 838564515 28698  
##   
## Step: AIC=28574.85  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:diphtheria.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - infant\_deaths\_cat:diphtheria.c 3 2323115 800845578 28558  
## - sqrt.alc.c:measles\_cat 3 4945579 803468042 28565  
## - infant\_deaths\_cat:sqrt.alc.c 3 5012629 803535092 28566  
## - sqrt.ad.mort.c:measles\_cat 3 5254160 803776623 28566  
## - sqrt.alc.c:polio.c 1 284317 798806780 28568  
## - bmi.c:polio.c 1 301124 798823587 28568  
## - sqrt.ad.mort.c:sqrt.alc.c 1 370082 798892544 28568  
## - sqrt.alc.c:diphtheria.c 1 593389 799115851 28569  
## - sqrt.ad.mort.c:polio.c 1 668049 799190512 28569  
## - diphtheria.c:thin.c 1 1022721 799545183 28570  
## - sqrt.ad.mort.c:thin.c 1 1200012 799722475 28570  
## <none> 798522463 28575  
## - polio.c:thin.c 1 3057728 801580190 28576  
## - sqrt.ad.mort.c:diphtheria.c 1 3492589 802015051 28577  
## - sqrt.alc.c:bmi.c 1 4289928 802812390 28579  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10198490 808720952 28580  
## - measles\_cat:bmi.c 3 10319534 808841996 28580  
## - infant\_deaths\_cat:measles\_cat 9 28915668 827438130 28584  
## - sqrt.ad.mort.c:bmi.c 1 6145822 804668284 28584  
## - bmi.c:diphtheria.c 1 7040871 805563334 28586  
## + infant\_deaths\_cat:polio.c 3 3186769 795335693 28589  
## + measles\_cat:polio.c 3 2052091 796470372 28592  
## - polio.c:diphtheria.c 1 10164228 808686690 28595  
## + measles\_cat:thin.c 3 936776 797585687 28595  
## + measles\_cat:diphtheria.c 3 370469 798151994 28597  
## - bmi.c:thin.c 1 15301080 813823543 28609  
## - infant\_deaths\_cat:thin.c 3 21570767 820093230 28610  
## - infant\_deaths\_cat:bmi.c 3 32214180 830736643 28639  
## - sqrt.alc.c:thin.c 1 43652514 842174977 28684  
##   
## Step: AIC=28558.14  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:sqrt.alc.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat +   
## sqrt.alc.c:bmi.c + sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c +   
## sqrt.alc.c:thin.c + measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - infant\_deaths\_cat:sqrt.alc.c 3 4637929 805483507 28548  
## - sqrt.alc.c:measles\_cat 3 5048553 805894131 28549  
## - sqrt.ad.mort.c:measles\_cat 3 5356783 806202361 28550  
## - diphtheria.c:thin.c 1 312639 801158217 28551  
## - sqrt.alc.c:polio.c 1 347738 801193316 28551  
## - bmi.c:polio.c 1 354369 801199947 28551  
## - sqrt.ad.mort.c:sqrt.alc.c 1 459064 801304643 28552  
## - sqrt.ad.mort.c:polio.c 1 710130 801555708 28552  
## - sqrt.alc.c:diphtheria.c 1 721934 801567513 28552  
## - sqrt.ad.mort.c:thin.c 1 1286570 802132148 28554  
## <none> 800845578 28558  
## - polio.c:thin.c 1 2812610 803658188 28558  
## - sqrt.ad.mort.c:diphtheria.c 1 2816692 803662270 28558  
## - sqrt.alc.c:bmi.c 1 4461002 805306580 28563  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10573839 811419418 28564  
## - measles\_cat:bmi.c 3 10883328 811728907 28565  
## - sqrt.ad.mort.c:bmi.c 1 5920396 806765974 28567  
## - infant\_deaths\_cat:measles\_cat 9 29168770 830014348 28567  
## + measles\_cat:polio.c 3 2698752 798146827 28574  
## - bmi.c:diphtheria.c 1 8779807 809625385 28574  
## + infant\_deaths\_cat:diphtheria.c 3 2323115 798522463 28575  
## - polio.c:diphtheria.c 1 9607567 810453145 28577  
## + measles\_cat:diphtheria.c 3 1298917 799546661 28578  
## + infant\_deaths\_cat:polio.c 3 1294250 799551328 28578  
## + measles\_cat:thin.c 3 944335 799901243 28579  
## - bmi.c:thin.c 1 16216568 817062146 28594  
## - infant\_deaths\_cat:thin.c 3 22415430 823261008 28596  
## - infant\_deaths\_cat:bmi.c 3 34512974 835358552 28628  
## - sqrt.alc.c:thin.c 1 44410316 845255894 28669  
##   
## Step: AIC=28547.74  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:measles\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.ad.mort.c:measles\_cat 3 5186764 810670272 28539  
## - diphtheria.c:thin.c 1 290378 805773885 28541  
## - sqrt.alc.c:polio.c 1 403265 805886772 28541  
## - bmi.c:polio.c 1 406370 805889877 28541  
## - sqrt.ad.mort.c:polio.c 1 479587 805963094 28541  
## - sqrt.ad.mort.c:sqrt.alc.c 1 615380 806098888 28542  
## - sqrt.alc.c:diphtheria.c 1 748236 806231744 28542  
## - sqrt.ad.mort.c:thin.c 1 1078614 806562122 28543  
## - sqrt.alc.c:measles\_cat 3 6841572 812325079 28543  
## - polio.c:thin.c 1 2676314 808159821 28547  
## - sqrt.ad.mort.c:diphtheria.c 1 2706871 808190379 28547  
## <none> 805483507 28548  
## - sqrt.alc.c:bmi.c 1 4285353 809768860 28552  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 10839257 816322764 28554  
## - measles\_cat:bmi.c 3 10993952 816477460 28554  
## - sqrt.ad.mort.c:bmi.c 1 6133750 811617257 28557  
## + infant\_deaths\_cat:sqrt.alc.c 3 4637929 800845578 28558  
## - infant\_deaths\_cat:measles\_cat 9 29813781 835297289 28558  
## + measles\_cat:polio.c 3 2946936 802536571 28563  
## - bmi.c:diphtheria.c 1 8531572 814015079 28563  
## + infant\_deaths\_cat:diphtheria.c 3 1948415 803535092 28566  
## + infant\_deaths\_cat:polio.c 3 1573410 803910097 28566  
## + measles\_cat:diphtheria.c 3 1184637 804298870 28568  
## - polio.c:diphtheria.c 1 10264170 815747677 28568  
## + measles\_cat:thin.c 3 689713 804793794 28569  
## - infant\_deaths\_cat:thin.c 3 19338563 824822070 28577  
## - bmi.c:thin.c 1 16332743 821816251 28584  
## - infant\_deaths\_cat:bmi.c 3 34588381 840071888 28617  
## - sqrt.alc.c:thin.c 1 47820430 853303938 28667  
##   
## Step: AIC=28538.74  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c +   
## sqrt.ad.mort.c:diphtheria.c + sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat +   
## sqrt.alc.c:bmi.c + sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c +   
## sqrt.alc.c:thin.c + measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c + diphtheria.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - diphtheria.c:thin.c 1 243104 810913376 28532  
## - sqrt.ad.mort.c:polio.c 1 309972 810980243 28532  
## - sqrt.ad.mort.c:thin.c 1 347750 811018022 28532  
## - sqrt.alc.c:polio.c 1 372763 811043035 28532  
## - bmi.c:polio.c 1 374391 811044663 28532  
## - sqrt.ad.mort.c:sqrt.alc.c 1 835797 811506068 28533  
## - sqrt.alc.c:measles\_cat 3 6580113 817250385 28533  
## - sqrt.alc.c:diphtheria.c 1 891599 811561871 28534  
## - polio.c:thin.c 1 2376071 813046343 28538  
## <none> 810670272 28539  
## - sqrt.ad.mort.c:diphtheria.c 1 3094818 813765089 28539  
## - measles\_cat:bmi.c 3 9313486 819983758 28541  
## - sqrt.alc.c:bmi.c 1 4272082 814942353 28543  
## - sqrt.ad.mort.c:bmi.c 1 5373349 816043620 28546  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 11819108 822489379 28547  
## + sqrt.ad.mort.c:measles\_cat 3 5186764 805483507 28548  
## + infant\_deaths\_cat:sqrt.alc.c 3 4467911 806202361 28550  
## + measles\_cat:polio.c 3 3783543 806886729 28552  
## - bmi.c:diphtheria.c 1 9116031 819786302 28556  
## + infant\_deaths\_cat:diphtheria.c 3 2044329 808625943 28556  
## + infant\_deaths\_cat:polio.c 3 1623389 809046883 28557  
## - infant\_deaths\_cat:measles\_cat 9 33205814 843876085 28558  
## + measles\_cat:diphtheria.c 3 1461391 809208880 28558  
## - polio.c:diphtheria.c 1 10375896 821046168 28559  
## + measles\_cat:thin.c 3 536734 810133538 28560  
## - infant\_deaths\_cat:thin.c 3 21177766 831848038 28572  
## - bmi.c:thin.c 1 16726934 827397206 28576  
## - infant\_deaths\_cat:bmi.c 3 32788079 843458351 28603  
## - sqrt.alc.c:thin.c 1 46793138 857463410 28654  
##   
## Step: AIC=28531.71  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:polio.c +   
## sqrt.ad.mort.c:diphtheria.c + sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat +   
## sqrt.alc.c:bmi.c + sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c +   
## sqrt.alc.c:thin.c + measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.ad.mort.c:polio.c 1 236027 811149402 28525  
## - bmi.c:polio.c 1 276756 811190131 28525  
## - sqrt.ad.mort.c:thin.c 1 324229 811237604 28525  
## - sqrt.alc.c:polio.c 1 444290 811357665 28525  
## - sqrt.alc.c:diphtheria.c 1 748832 811662208 28526  
## - sqrt.ad.mort.c:sqrt.alc.c 1 861302 811774678 28526  
## - sqrt.alc.c:measles\_cat 3 6630624 817544000 28526  
## - polio.c:thin.c 1 2376095 813289471 28530  
## <none> 810913376 28532  
## - sqrt.ad.mort.c:diphtheria.c 1 3341824 814255200 28533  
## - measles\_cat:bmi.c 3 9298150 820211525 28534  
## - sqrt.alc.c:bmi.c 1 4318242 815231618 28536  
## - sqrt.ad.mort.c:bmi.c 1 5382127 816295503 28538  
## + diphtheria.c:thin.c 1 243104 810670272 28539  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 11694615 822607991 28540  
## + sqrt.ad.mort.c:measles\_cat 3 5139491 805773885 28541  
## + infant\_deaths\_cat:sqrt.alc.c 3 4446659 806466717 28543  
## + measles\_cat:polio.c 3 3625534 807287841 28545  
## - bmi.c:diphtheria.c 1 9623158 820536534 28550  
## + infant\_deaths\_cat:polio.c 3 1618386 809294990 28550  
## - infant\_deaths\_cat:measles\_cat 9 33183565 844096941 28550  
## + infant\_deaths\_cat:diphtheria.c 3 1447959 809465417 28551  
## + measles\_cat:diphtheria.c 3 1072218 809841157 28552  
## - polio.c:diphtheria.c 1 10801071 821714446 28553  
## + measles\_cat:thin.c 3 546575 810366801 28553  
## - infant\_deaths\_cat:thin.c 3 21961397 832874773 28567  
## - bmi.c:thin.c 1 16767880 827681255 28569  
## - infant\_deaths\_cat:bmi.c 3 32737236 843650612 28596  
## - sqrt.alc.c:thin.c 1 46605598 857518973 28647  
##   
## Step: AIC=28524.65  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c +   
## sqrt.ad.mort.c:thin.c + infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.ad.mort.c:thin.c 1 311838 811461241 28518  
## - sqrt.alc.c:polio.c 1 415416 811564818 28518  
## - bmi.c:polio.c 1 451772 811601175 28518  
## - sqrt.alc.c:diphtheria.c 1 799378 811948780 28519  
## - sqrt.alc.c:measles\_cat 3 6555182 817704585 28519  
## - sqrt.ad.mort.c:sqrt.alc.c 1 921956 812071359 28520  
## - polio.c:thin.c 1 2223760 813373162 28523  
## <none> 811149402 28525  
## - measles\_cat:bmi.c 3 9217950 820367353 28526  
## - sqrt.alc.c:bmi.c 1 4256563 815405965 28528  
## - sqrt.ad.mort.c:bmi.c 1 5282032 816431435 28531  
## + sqrt.ad.mort.c:polio.c 1 236027 810913376 28532  
## + diphtheria.c:thin.c 1 169159 810980243 28532  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 11738000 822887402 28533  
## - sqrt.ad.mort.c:diphtheria.c 1 6231726 817381129 28534  
## + sqrt.ad.mort.c:measles\_cat 3 4996096 806153306 28534  
## + infant\_deaths\_cat:sqrt.alc.c 3 4280443 806868959 28536  
## + measles\_cat:polio.c 3 3513086 807636316 28538  
## + infant\_deaths\_cat:polio.c 3 1617277 809532125 28543  
## + infant\_deaths\_cat:diphtheria.c 3 1521019 809628383 28544  
## - infant\_deaths\_cat:measles\_cat 9 33290102 844439505 28544  
## + measles\_cat:diphtheria.c 3 1110270 810039133 28545  
## - bmi.c:diphtheria.c 1 10575310 821724712 28545  
## - polio.c:diphtheria.c 1 10868878 822018280 28546  
## + measles\_cat:thin.c 3 556633 810592770 28546  
## - infant\_deaths\_cat:thin.c 3 21959601 833109003 28560  
## - bmi.c:thin.c 1 16564556 827713959 28561  
## - infant\_deaths\_cat:bmi.c 3 33175843 844325245 28590  
## - sqrt.alc.c:thin.c 1 46393369 857542771 28639  
##   
## Step: AIC=28517.8  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c +   
## infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + bmi.c:polio.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - bmi.c:polio.c 1 385573 811846814 28511  
## - sqrt.alc.c:polio.c 1 437819 811899059 28511  
## - sqrt.alc.c:measles\_cat 3 6431262 817892502 28512  
## - sqrt.alc.c:diphtheria.c 1 802211 812263452 28512  
## - sqrt.ad.mort.c:sqrt.alc.c 1 1142963 812604203 28513  
## - polio.c:thin.c 1 1971120 813432361 28515  
## <none> 811461241 28518  
## - measles\_cat:bmi.c 3 8998020 820459260 28519  
## - sqrt.alc.c:bmi.c 1 4252471 815713712 28522  
## - sqrt.ad.mort.c:bmi.c 1 5015316 816476557 28524  
## + sqrt.ad.mort.c:thin.c 1 311838 811149402 28525  
## + sqrt.ad.mort.c:polio.c 1 223636 811237604 28525  
## + diphtheria.c:thin.c 1 151870 811309371 28525  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 11670457 823131698 28526  
## - sqrt.ad.mort.c:diphtheria.c 1 6421131 817882371 28527  
## + sqrt.ad.mort.c:measles\_cat 3 4313963 807147278 28529  
## + infant\_deaths\_cat:sqrt.alc.c 3 4172524 807288717 28530  
## + measles\_cat:polio.c 3 3519150 807942091 28531  
## + infant\_deaths\_cat:polio.c 3 1600713 809860528 28537  
## + infant\_deaths\_cat:diphtheria.c 3 1573140 809888101 28537  
## - infant\_deaths\_cat:measles\_cat 9 33478151 844939392 28537  
## + measles\_cat:diphtheria.c 3 1160388 810300853 28538  
## - polio.c:diphtheria.c 1 10717124 822178365 28539  
## - bmi.c:diphtheria.c 1 10767182 822228422 28539  
## + measles\_cat:thin.c 3 507263 810953977 28540  
## - infant\_deaths\_cat:thin.c 3 21847435 833308676 28553  
## - bmi.c:thin.c 1 16669107 828130347 28555  
## - infant\_deaths\_cat:bmi.c 3 34215668 845676909 28585  
## - sqrt.alc.c:thin.c 1 46683311 858144552 28633  
##   
## Step: AIC=28511.15  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c +   
## infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:polio.c + sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c +   
## measles\_cat:bmi.c + bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.alc.c:polio.c 1 357116 812203929 28504  
## - sqrt.alc.c:measles\_cat 3 6431040 818277854 28505  
## - sqrt.alc.c:diphtheria.c 1 926798 812773612 28506  
## - sqrt.ad.mort.c:sqrt.alc.c 1 1147675 812994489 28507  
## - polio.c:thin.c 1 1585593 813432406 28508  
## <none> 811846814 28511  
## - measles\_cat:bmi.c 3 9106602 820953416 28513  
## - sqrt.alc.c:bmi.c 1 4212023 816058836 28515  
## - sqrt.ad.mort.c:bmi.c 1 4944867 816791681 28517  
## + bmi.c:polio.c 1 385573 811461241 28518  
## + sqrt.ad.mort.c:polio.c 1 380816 811465997 28518  
## + sqrt.ad.mort.c:thin.c 1 245639 811601175 28518  
## + diphtheria.c:thin.c 1 54535 811792279 28519  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 11617810 823464624 28519  
## - sqrt.ad.mort.c:diphtheria.c 1 6461911 818308725 28521  
## + sqrt.ad.mort.c:measles\_cat 3 4307856 807538958 28523  
## + infant\_deaths\_cat:sqrt.alc.c 3 4191961 807654853 28523  
## + measles\_cat:polio.c 3 3094601 808752213 28526  
## + infant\_deaths\_cat:diphtheria.c 3 1729213 810117601 28530  
## + infant\_deaths\_cat:polio.c 3 1666526 810180288 28530  
## - infant\_deaths\_cat:measles\_cat 9 33209195 845056009 28530  
## + measles\_cat:diphtheria.c 3 1238030 810608784 28531  
## - polio.c:diphtheria.c 1 10500865 822347679 28532  
## + measles\_cat:thin.c 3 491722 811355091 28533  
## - bmi.c:diphtheria.c 1 11024550 822871364 28533  
## - infant\_deaths\_cat:thin.c 3 21461889 833308702 28545  
## - bmi.c:thin.c 1 16553493 828400306 28548  
## - infant\_deaths\_cat:bmi.c 3 34115726 845962539 28578  
## - sqrt.alc.c:thin.c 1 46887284 858734098 28627  
##   
## Step: AIC=28504.42  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c +   
## infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:measles\_cat + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.alc.c:measles\_cat 3 6270826 818474756 28498  
## - sqrt.ad.mort.c:sqrt.alc.c 1 1081615 813285544 28500  
## - polio.c:thin.c 1 2076581 814280510 28502  
## - sqrt.alc.c:diphtheria.c 1 2108472 814312401 28502  
## <none> 812203929 28504  
## - measles\_cat:bmi.c 3 9112926 821316855 28506  
## - sqrt.alc.c:bmi.c 1 4342108 816546037 28508  
## - sqrt.ad.mort.c:bmi.c 1 4992232 817196161 28510  
## + sqrt.alc.c:polio.c 1 357116 811846814 28511  
## + sqrt.ad.mort.c:polio.c 1 326059 811877871 28511  
## + bmi.c:polio.c 1 304870 811899059 28511  
## + sqrt.ad.mort.c:thin.c 1 269979 811933950 28511  
## + diphtheria.c:thin.c 1 96858 812107072 28512  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 11639531 823843460 28513  
## - sqrt.ad.mort.c:diphtheria.c 1 6471676 818675606 28514  
## + sqrt.ad.mort.c:measles\_cat 3 4267794 807936135 28516  
## + infant\_deaths\_cat:sqrt.alc.c 3 4250652 807953277 28516  
## + measles\_cat:polio.c 3 3353653 808850276 28518  
## + infant\_deaths\_cat:polio.c 3 1828185 810375744 28523  
## + infant\_deaths\_cat:diphtheria.c 3 1734820 810469109 28523  
## - infant\_deaths\_cat:measles\_cat 9 33217318 845421248 28523  
## - polio.c:diphtheria.c 1 10251327 822455257 28524  
## + measles\_cat:diphtheria.c 3 1179552 811024377 28524  
## - bmi.c:diphtheria.c 1 10853247 823057176 28526  
## + measles\_cat:thin.c 3 519199 811684730 28526  
## - infant\_deaths\_cat:thin.c 3 21711808 833915737 28539  
## - bmi.c:thin.c 1 16460084 828664013 28541  
## - infant\_deaths\_cat:bmi.c 3 34320848 846524777 28572  
## - sqrt.alc.c:thin.c 1 46556564 858760493 28619  
##   
## Step: AIC=28498.22  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:sqrt.alc.c + sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c +   
## infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:bmi.c + sqrt.alc.c:diphtheria.c +   
## sqrt.alc.c:thin.c + measles\_cat:bmi.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.ad.mort.c:sqrt.alc.c 1 1189062 819663818 28494  
## - measles\_cat:bmi.c 3 7460674 825935430 28495  
## - sqrt.alc.c:diphtheria.c 1 1880049 820354804 28496  
## - polio.c:thin.c 1 2165545 820640301 28496  
## <none> 818474756 28498  
## - sqrt.alc.c:bmi.c 1 3577326 822052082 28500  
## - sqrt.ad.mort.c:bmi.c 1 4533396 823008151 28503  
## + sqrt.alc.c:measles\_cat 3 6270826 812203929 28504  
## + bmi.c:polio.c 1 323734 818151021 28505  
## + sqrt.ad.mort.c:polio.c 1 256381 818218375 28505  
## + sqrt.alc.c:polio.c 1 196902 818277854 28505  
## + sqrt.ad.mort.c:thin.c 1 153456 818321300 28506  
## + diphtheria.c:thin.c 1 124278 818350478 28506  
## + infant\_deaths\_cat:sqrt.alc.c 3 5809856 812664899 28506  
## - sqrt.ad.mort.c:diphtheria.c 1 6218116 824692872 28507  
## - infant\_deaths\_cat:measles\_cat 9 30299482 848774238 28509  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 13039055 831513810 28510  
## + sqrt.ad.mort.c:measles\_cat 3 4147161 814327595 28510  
## + measles\_cat:polio.c 3 3261075 815213681 28512  
## + infant\_deaths\_cat:polio.c 3 1841637 816633118 28516  
## + measles\_cat:thin.c 3 1728099 816746657 28517  
## + infant\_deaths\_cat:diphtheria.c 3 1654732 816820024 28517  
## - polio.c:diphtheria.c 1 10460872 828935627 28518  
## - bmi.c:diphtheria.c 1 10499264 828974019 28518  
## + measles\_cat:diphtheria.c 3 959998 817514757 28519  
## - bmi.c:thin.c 1 13178322 831653077 28526  
## - infant\_deaths\_cat:thin.c 3 20153696 838628451 28528  
## - infant\_deaths\_cat:bmi.c 3 34659826 853134582 28566  
## - sqrt.alc.c:thin.c 1 53291357 871766113 28629  
##   
## Step: AIC=28493.72  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:diphtheria.c + sqrt.alc.c:thin.c + measles\_cat:bmi.c +   
## bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.alc.c:diphtheria.c 1 1512513 821176331 28490  
## - measles\_cat:bmi.c 3 7394869 827058686 28490  
## - polio.c:thin.c 1 2173804 821837622 28492  
## <none> 819663818 28494  
## - sqrt.alc.c:bmi.c 1 3174809 822838627 28494  
## - sqrt.ad.mort.c:bmi.c 1 3791167 823454984 28496  
## + sqrt.ad.mort.c:sqrt.alc.c 1 1189062 818474756 28498  
## + sqrt.alc.c:measles\_cat 3 6378274 813285544 28500  
## + infant\_deaths\_cat:sqrt.alc.c 3 6240924 813422894 28500  
## + bmi.c:polio.c 1 337132 819326686 28500  
## + sqrt.ad.mort.c:polio.c 1 326730 819337088 28500  
## + sqrt.ad.mort.c:thin.c 1 316661 819347157 28501  
## + sqrt.alc.c:polio.c 1 142761 819521057 28501  
## + diphtheria.c:thin.c 1 126930 819536888 28501  
## - sqrt.ad.mort.c:diphtheria.c 1 6494038 826157855 28503  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 12589995 832253813 28504  
## - infant\_deaths\_cat:measles\_cat 9 30740201 850404019 28505  
## + sqrt.ad.mort.c:measles\_cat 3 4232200 815431618 28505  
## + measles\_cat:polio.c 3 3260501 816403317 28508  
## + infant\_deaths\_cat:diphtheria.c 3 1731803 817932015 28512  
## + infant\_deaths\_cat:polio.c 3 1698382 817965436 28512  
## + measles\_cat:thin.c 3 1633440 818030378 28512  
## - polio.c:diphtheria.c 1 10207577 829871395 28513  
## + measles\_cat:diphtheria.c 3 1004792 818659026 28514  
## - bmi.c:diphtheria.c 1 10867560 830531378 28515  
## - bmi.c:thin.c 1 13816254 833480072 28523  
## - infant\_deaths\_cat:thin.c 3 20519299 840183117 28525  
## - infant\_deaths\_cat:bmi.c 3 35759877 855423695 28564  
## - sqrt.alc.c:thin.c 1 60649586 880313403 28643  
##   
## Step: AIC=28490.07  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:thin.c + measles\_cat:bmi.c + bmi.c:diphtheria.c +   
## bmi.c:thin.c + polio.c:diphtheria.c + polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - measles\_cat:bmi.c 3 7659697 828836028 28487  
## - polio.c:thin.c 1 2689451 823865782 28490  
## <none> 821176331 28490  
## - sqrt.alc.c:bmi.c 1 3629785 824806116 28492  
## - sqrt.ad.mort.c:bmi.c 1 3785449 824961780 28492  
## + sqrt.alc.c:diphtheria.c 1 1512513 819663818 28494  
## + sqrt.alc.c:polio.c 1 916165 820260166 28495  
## + sqrt.ad.mort.c:sqrt.alc.c 1 821527 820354804 28496  
## + bmi.c:polio.c 1 411227 820765104 28497  
## + sqrt.ad.mort.c:polio.c 1 393685 820782646 28497  
## + sqrt.alc.c:measles\_cat 3 6123588 815052743 28497  
## + sqrt.ad.mort.c:thin.c 1 316545 820859786 28497  
## + infant\_deaths\_cat:sqrt.alc.c 3 5965773 815210558 28497  
## + diphtheria.c:thin.c 1 16476 821159855 28498  
## - sqrt.ad.mort.c:diphtheria.c 1 6527249 827703580 28500  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 12487745 833664076 28500  
## + sqrt.ad.mort.c:measles\_cat 3 4330578 816845753 28502  
## - infant\_deaths\_cat:measles\_cat 9 31696597 852872928 28504  
## + measles\_cat:polio.c 3 3019097 818157234 28505  
## - polio.c:diphtheria.c 1 9147346 830323677 28507  
## + infant\_deaths\_cat:diphtheria.c 3 2202906 818973425 28507  
## + infant\_deaths\_cat:polio.c 3 1765522 819410809 28508  
## + measles\_cat:thin.c 3 1495462 819680869 28509  
## + measles\_cat:diphtheria.c 3 1427691 819748640 28509  
## - bmi.c:diphtheria.c 1 11999279 833175610 28514  
## - bmi.c:thin.c 1 13677544 834853875 28519  
## - infant\_deaths\_cat:thin.c 3 20097555 841273886 28520  
## - infant\_deaths\_cat:bmi.c 3 35978396 857154727 28561  
## - sqrt.alc.c:thin.c 1 59140938 880317269 28635  
##   
## Step: AIC=28487.37  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:thin.c + bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c +   
## polio.c:thin.c  
##   
## Df Sum of Sq RSS AIC  
## - polio.c:thin.c 1 2737841 831573869 28487  
## - sqrt.ad.mort.c:bmi.c 1 2877852 831713880 28487  
## <none> 828836028 28487  
## - sqrt.alc.c:bmi.c 1 3063514 831899542 28488  
## + measles\_cat:bmi.c 3 7659697 821176331 28490  
## + sqrt.alc.c:diphtheria.c 1 1777342 827058686 28490  
## + sqrt.alc.c:polio.c 1 1011286 827824742 28492  
## + sqrt.ad.mort.c:sqrt.alc.c 1 740897 828095131 28493  
## + bmi.c:polio.c 1 520716 828315312 28494  
## + sqrt.ad.mort.c:polio.c 1 345770 828490257 28494  
## + infant\_deaths\_cat:sqrt.alc.c 3 6106204 822729823 28494  
## + sqrt.ad.mort.c:thin.c 1 108387 828727641 28495  
## + diphtheria.c:thin.c 1 8188 828827840 28495  
## - sqrt.ad.mort.c:diphtheria.c 1 6486090 835322118 28497  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 13048524 841884552 28499  
## + sqrt.alc.c:measles\_cat 3 4401994 824434034 28499  
## + sqrt.ad.mort.c:measles\_cat 3 2922403 825913625 28503  
## - polio.c:diphtheria.c 1 8912362 837748390 28503  
## + infant\_deaths\_cat:diphtheria.c 3 2726937 826109091 28503  
## + measles\_cat:polio.c 3 2013589 826822439 28505  
## + infant\_deaths\_cat:polio.c 3 1764446 827071581 28506  
## + measles\_cat:diphtheria.c 3 1506295 827329733 28506  
## + measles\_cat:thin.c 3 1072655 827763373 28508  
## - bmi.c:diphtheria.c 1 10723471 839559499 28508  
## - bmi.c:thin.c 1 14202776 843038804 28517  
## - infant\_deaths\_cat:thin.c 3 20348711 849184739 28518  
## - infant\_deaths\_cat:measles\_cat 9 40623835 869459863 28523  
## - infant\_deaths\_cat:bmi.c 3 42236782 871072810 28573  
## - sqrt.alc.c:thin.c 1 57714894 886550922 28627  
##   
## Step: AIC=28486.91  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:bmi.c + sqrt.ad.mort.c:diphtheria.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:thin.c + bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c  
##   
## Df Sum of Sq RSS AIC  
## - sqrt.ad.mort.c:bmi.c 1 2741924 834315793 28486  
## <none> 831573869 28487  
## + polio.c:thin.c 1 2737841 828836028 28487  
## - sqrt.alc.c:bmi.c 1 3261833 834835702 28488  
## + sqrt.alc.c:diphtheria.c 1 2336717 829237152 28488  
## + sqrt.alc.c:polio.c 1 1969006 829604863 28489  
## + measles\_cat:bmi.c 3 7708087 823865782 28490  
## + sqrt.ad.mort.c:sqrt.alc.c 1 701647 830872222 28493  
## + diphtheria.c:thin.c 1 663147 830910723 28493  
## + infant\_deaths\_cat:sqrt.alc.c 3 6117854 825456016 28494  
## + bmi.c:polio.c 1 15537 831558332 28495  
## + sqrt.ad.mort.c:polio.c 1 13133 831560736 28495  
## + sqrt.ad.mort.c:thin.c 1 119 831573751 28495  
## - sqrt.ad.mort.c:diphtheria.c 1 6572641 838146510 28496  
## - polio.c:diphtheria.c 1 7131977 838705846 28498  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 13096301 844670170 28498  
## + sqrt.alc.c:measles\_cat 3 4439879 827133990 28498  
## + measles\_cat:polio.c 3 3607160 827966710 28500  
## + infant\_deaths\_cat:diphtheria.c 3 3463409 828110460 28501  
## + sqrt.ad.mort.c:measles\_cat 3 2974305 828599565 28502  
## + measles\_cat:diphtheria.c 3 2241933 829331937 28504  
## + infant\_deaths\_cat:polio.c 3 2124371 829449498 28504  
## + measles\_cat:thin.c 3 1020116 830553754 28507  
## - infant\_deaths\_cat:thin.c 3 18462033 850035902 28512  
## - bmi.c:thin.c 1 14500479 846074348 28517  
## - bmi.c:diphtheria.c 1 15398200 846972070 28520  
## - infant\_deaths\_cat:measles\_cat 9 39959329 871533198 28521  
## - infant\_deaths\_cat:bmi.c 3 40786677 872360546 28569  
## - sqrt.alc.c:thin.c 1 56633517 888207386 28624  
##   
## Step: AIC=28486.44  
## lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat + sqrt.alc.c +   
## measles\_cat + bmi.c + polio.c + diphtheria.c + thin.c + sqrt.ad.mort.c:infant\_deaths\_cat +   
## sqrt.ad.mort.c:diphtheria.c + infant\_deaths\_cat:measles\_cat +   
## infant\_deaths\_cat:bmi.c + infant\_deaths\_cat:thin.c + sqrt.alc.c:bmi.c +   
## sqrt.alc.c:thin.c + bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c  
##   
## Df Sum of Sq RSS AIC  
## <none> 834315793 28486  
## - sqrt.alc.c:bmi.c 1 2991205 837306998 28487  
## + sqrt.ad.mort.c:bmi.c 1 2741924 831573869 28487  
## + polio.c:thin.c 1 2601913 831713880 28487  
## + sqrt.alc.c:diphtheria.c 1 2298466 832017327 28488  
## + sqrt.alc.c:polio.c 1 2040002 832275791 28489  
## + measles\_cat:bmi.c 3 6834100 827481693 28492  
## + diphtheria.c:thin.c 1 559605 833756188 28493  
## + infant\_deaths\_cat:sqrt.alc.c 3 6202325 828113468 28493  
## + sqrt.ad.mort.c:thin.c 1 305123 834010670 28493  
## + sqrt.ad.mort.c:sqrt.alc.c 1 236193 834079600 28494  
## + bmi.c:polio.c 1 24233 834291561 28494  
## + sqrt.ad.mort.c:polio.c 1 1030 834314763 28494  
## - sqrt.ad.mort.c:diphtheria.c 1 5982357 840298150 28494  
## - polio.c:diphtheria.c 1 7145225 841461018 28498  
## + sqrt.alc.c:measles\_cat 3 4338776 829977017 28498  
## + measles\_cat:polio.c 3 3783715 830532078 28500  
## + infant\_deaths\_cat:diphtheria.c 3 3013552 831302241 28502  
## + sqrt.ad.mort.c:measles\_cat 3 2914160 831401633 28502  
## + measles\_cat:diphtheria.c 3 2249168 832066625 28504  
## + infant\_deaths\_cat:polio.c 3 2194338 832121455 28504  
## - sqrt.ad.mort.c:infant\_deaths\_cat 3 16194493 850510286 28506  
## + measles\_cat:thin.c 3 1320603 832995190 28506  
## - bmi.c:thin.c 1 12784015 847099808 28512  
## - infant\_deaths\_cat:thin.c 3 20031327 854347120 28515  
## - infant\_deaths\_cat:measles\_cat 9 39918243 874234037 28520  
## - bmi.c:diphtheria.c 1 17228860 851544653 28524  
## - infant\_deaths\_cat:bmi.c 3 43725655 878041448 28576  
## - sqrt.alc.c:thin.c 1 56041630 890357423 28621

summary(bic\_cent)

##   
## Call:  
## lm(formula = lifeExpectancy^2 ~ sqrt.ad.mort.c + infant\_deaths\_cat +   
## sqrt.alc.c + measles\_cat + bmi.c + polio.c + diphtheria.c +   
## thin.c + sqrt.ad.mort.c:infant\_deaths\_cat + sqrt.ad.mort.c:diphtheria.c +   
## infant\_deaths\_cat:measles\_cat + infant\_deaths\_cat:bmi.c +   
## infant\_deaths\_cat:thin.c + sqrt.alc.c:bmi.c + sqrt.alc.c:thin.c +   
## bmi.c:diphtheria.c + bmi.c:thin.c + polio.c:diphtheria.c,   
## data = life\_expectancy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3196.9 -377.9 19.1 355.1 2292.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.007e+03 3.807e+01 131.508 < 2e-16 \*\*\*  
## sqrt.ad.mort.c -3.581e+01 7.320e+00 -4.892 1.07e-06 \*\*\*  
## infant\_deaths\_catQ2 -1.953e+02 5.857e+01 -3.334 0.000871 \*\*\*  
## infant\_deaths\_catQ3 -1.942e+02 6.298e+01 -3.084 0.002070 \*\*   
## infant\_deaths\_catQ4 -2.534e+02 1.244e+02 -2.038 0.041701 \*   
## sqrt.alc.c 5.379e+01 1.454e+01 3.701 0.000220 \*\*\*  
## measles\_catQ2 2.936e+02 6.526e+01 4.500 7.17e-06 \*\*\*  
## measles\_catQ3 6.285e+02 7.082e+01 8.875 < 2e-16 \*\*\*  
## measles\_catQ4 7.309e+02 1.916e+02 3.814 0.000141 \*\*\*  
## bmi.c -5.458e+00 1.657e+00 -3.293 0.001006 \*\*   
## polio.c 3.484e+00 8.481e-01 4.108 4.13e-05 \*\*\*  
## diphtheria.c 7.088e+00 8.583e-01 8.258 2.54e-16 \*\*\*  
## thin.c -8.507e+01 1.032e+01 -8.244 2.87e-16 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ2 -5.697e+01 9.521e+00 -5.983 2.56e-09 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ3 -5.247e+01 9.879e+00 -5.311 1.20e-07 \*\*\*  
## sqrt.ad.mort.c:infant\_deaths\_catQ4 -4.565e+01 9.330e+00 -4.893 1.07e-06 \*\*\*  
## sqrt.ad.mort.c:diphtheria.c -5.506e-01 1.400e-01 -3.934 8.63e-05 \*\*\*  
## infant\_deaths\_catQ2:measles\_catQ2 -2.936e+02 9.667e+01 -3.037 0.002416 \*\*   
## infant\_deaths\_catQ3:measles\_catQ2 -4.223e+02 1.097e+02 -3.849 0.000122 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ2 -5.519e+02 1.634e+02 -3.377 0.000745 \*\*\*  
## infant\_deaths\_catQ2:measles\_catQ3 -4.702e+02 9.701e+01 -4.847 1.34e-06 \*\*\*  
## infant\_deaths\_catQ3:measles\_catQ3 -7.907e+02 1.056e+02 -7.489 1.00e-13 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ3 -1.059e+03 1.472e+02 -7.196 8.49e-13 \*\*\*  
## infant\_deaths\_catQ2:measles\_catQ4 -5.219e+02 2.080e+02 -2.509 0.012186 \*   
## infant\_deaths\_catQ3:measles\_catQ4 -8.469e+02 2.087e+02 -4.058 5.13e-05 \*\*\*  
## infant\_deaths\_catQ4:measles\_catQ4 -8.238e+02 2.293e+02 -3.593 0.000334 \*\*\*  
## infant\_deaths\_catQ2:bmi.c 1.106e+01 2.207e+00 5.013 5.81e-07 \*\*\*  
## infant\_deaths\_catQ3:bmi.c 1.758e+01 2.524e+00 6.966 4.31e-12 \*\*\*  
## infant\_deaths\_catQ4:bmi.c 2.650e+01 2.575e+00 10.292 < 2e-16 \*\*\*  
## infant\_deaths\_catQ2:thin.c -5.097e+01 1.330e+01 -3.832 0.000131 \*\*\*  
## infant\_deaths\_catQ3:thin.c 3.988e+01 1.487e+01 2.681 0.007398 \*\*   
## infant\_deaths\_catQ4:thin.c 1.690e+01 1.206e+01 1.401 0.161316   
## sqrt.alc.c:bmi.c -2.176e+00 7.822e-01 -2.782 0.005457 \*\*   
## sqrt.alc.c:thin.c -5.019e+01 4.168e+00 -12.040 < 2e-16 \*\*\*  
## bmi.c:diphtheria.c -2.422e-01 3.629e-02 -6.676 3.12e-11 \*\*\*  
## bmi.c:thin.c -1.553e+00 2.701e-01 -5.750 1.02e-08 \*\*\*  
## polio.c:diphtheria.c 9.580e-02 2.228e-02 4.299 1.79e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 621.8 on 2158 degrees of freedom  
## Multiple R-squared: 0.7042, Adjusted R-squared: 0.6993   
## F-statistic: 142.7 on 36 and 2158 DF, p-value: < 2.2e-16

Similar to the AIC model, the BIC dropped a few variables.

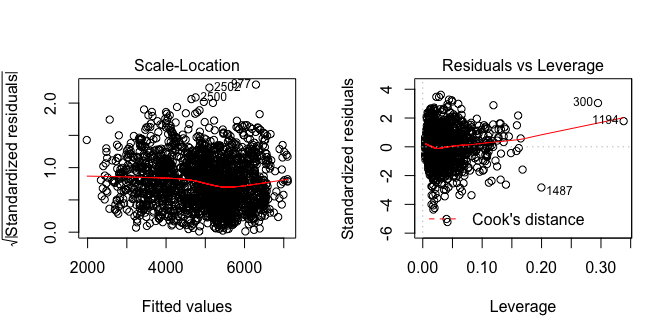
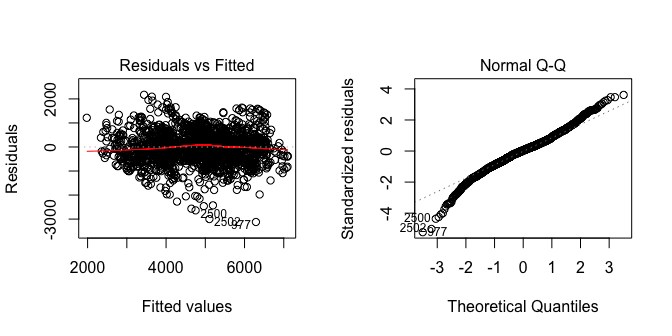
From the regression analysis, all variables are significant with a 0.001 significance level. The (adjusted) R-squared is 69.93% which implies that this model explains 69.93% of the variation in the response variable. This is smaller than the R-squared obtained before applying the step-wise regression and the AIC criterion. The residual standard error is 621; it is larger than the residual standard error obtained previously by 21 points.

Based on these results, it appears that the AIC model is better than the BIC criterion because it gives a higher R-squared and lower standard residuals.

## Residual Analysis

We will use the AIC model (based on the explanations given above) and conduct residual analysis.

par (mfrow = c(1,2))  
plot(aic\_cent)



The residual v/s fitted value plot is useful for assessing linearity and constant variance conditions of a simple linear model. The residuals are scattered across the plot with high residuals at the bottom of the plot, however, there does not appear to be an extreme non-linear trend in the residuals. Hence, the residuals are consistent with the linearity condition. Further, there does not appear to be a major deviation from constant variance with points being scattered above and below the 0.0 line. However, there are a few points that might be outliers.

The normal quantile plot is a scatterplot that displays the observed data v/s the values that would be expected from a normal sample of the same size. In this quantile plot, there does not appear to be any major deviation from a straight line. While there are some right-skewed residuals and some highly left-skewed residuals but the trend is largely linear indicating that the deviation might be a small variation that can be ignored.

The Scale-Location plot does not indicate any obvious concerns; however, there is some curvature in the trendline in the middle but the residuals appear to be mainly clustered.

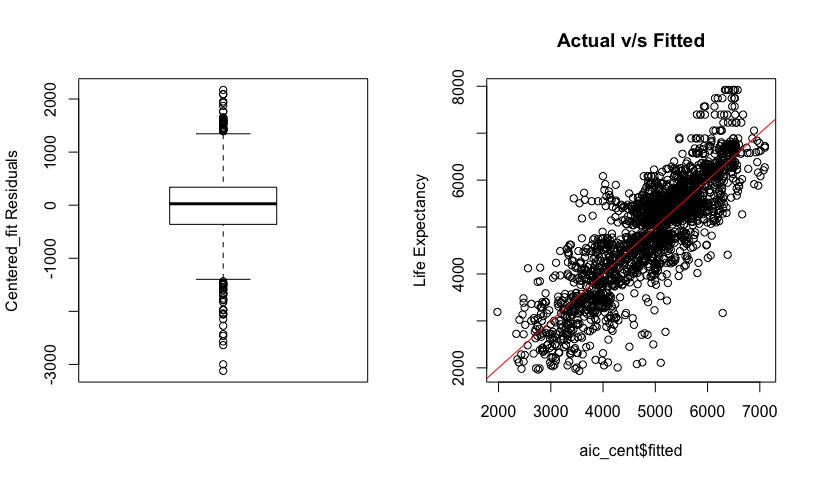
The Residuals vs. Leverage plot appears to show a large cluster of points to the left of the graph (indicating small leverage/ amount of influence) with some points being scattered to the right causing the trend line to bend upwards. The standard of leverage for this model is

which is equal to 0.03 (63 predictors, n = 2195). Most of the points have a higher leverage than 0.03. The points have a somewhat high leverage as they are bigger than

. However, most of the points appear to be clustered with a leverage less than 0.10 which means that they do not have high leverage ((2(k +1)) / n = 0.09). There are a few points that have extremely high leverage indicating that these points have a large influence. Further, there are standardized residuals greater than 3 on both ends (especially the left-end), indicating outliers. This might be a cause of concern.

As shown in the graph, there seem to be no other red dash-line, which means that none of the data has a high Cook’s Distance value, there are no unusual points

par(mfrow = c(1,2))  
boxplot(aic\_cent$residuals, ylab = 'Centered\_fit Residuals')  
plot(lifeExpectancy^2~aic\_cent$fitted, main = "Actual v/s Fitted", ylab = "Life Expectancy", data = life\_expectancy)  
abline(0,1, col = "red")



The box plot shows slight skewness at the right end with large skewness on the left end, which was shown in the quantile plot, indicating some deviation from a normal pattern. The Actual v/s Fitted plot shows a moderate linear trend.

##Variance Inflation Factor

vif(aic\_cent)

## GVIF Df GVIF^(1/(2\*Df))  
## sqrt.ad.mort.c 8.946115 1 2.991006  
## infant\_deaths\_cat 151.617488 3 2.309182  
## sqrt.alc.c 5.930774 1 2.435318  
## measles\_cat 722.482891 3 2.995513  
## bmi.c 8.056859 1 2.838461  
## polio.c 11.901665 1 3.449879  
## diphtheria.c 8.027833 1 2.833343  
## thin.c 13.118433 1 3.621938  
## sqrt.ad.mort.c:infant\_deaths\_cat 21.957460 3 1.673389  
## sqrt.ad.mort.c:measles\_cat 13.617160 3 1.545306  
## sqrt.ad.mort.c:bmi.c 1.995541 1 1.412636  
## sqrt.ad.mort.c:diphtheria.c 1.482176 1 1.217447  
## sqrt.ad.mort.c:thin.c 1.859580 1 1.363664  
## infant\_deaths\_cat:sqrt.alc.c 17.552188 3 1.612087  
## infant\_deaths\_cat:measles\_cat 87395.188898 9 1.881599  
## infant\_deaths\_cat:bmi.c 19.812359 3 1.644963  
## infant\_deaths\_cat:polio.c 49.862446 3 1.918502  
## infant\_deaths\_cat:diphtheria.c 28.800928 3 1.750791  
## infant\_deaths\_cat:thin.c 57.745550 3 1.966013  
## sqrt.alc.c:measles\_cat 11.290440 3 1.497793  
## sqrt.alc.c:bmi.c 1.775249 1 1.332385  
## sqrt.alc.c:diphtheria.c 1.548073 1 1.244216  
## sqrt.alc.c:thin.c 2.402788 1 1.550093  
## measles\_cat:bmi.c 13.201151 3 1.537335  
## measles\_cat:polio.c 13.413392 3 1.541428  
## bmi.c:diphtheria.c 1.885495 1 1.373133  
## bmi.c:thin.c 3.652990 1 1.911280  
## polio.c:diphtheria.c 2.129759 1 1.459370  
## polio.c:thin.c 2.930304 1 1.711813  
## diphtheria.c:thin.c 3.429576 1 1.851912

It seems that the VIF values are all less than 5, which means that the predictors have low correlation with one another.

## Influence Analysis

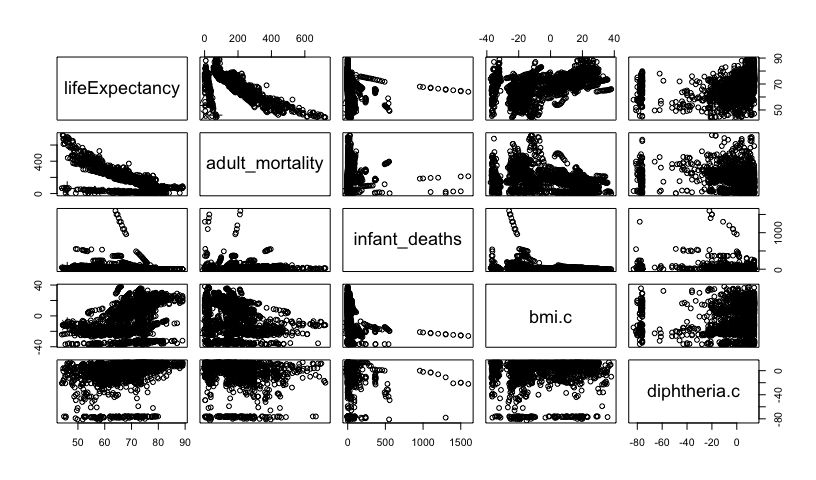
##Scatter Plot Matrix

Locate large residuals:

names(life\_expectancy)

## [1] "lifeExpectancy" "adult\_mortality" "infant\_deaths"   
## [4] "alcohol" "hepatitisB" "measles"   
## [7] "bmi" "polio" "diphtheria"   
## [10] "thin10to19" "thin5to9" "infant\_deaths\_cat"  
## [13] "measles\_cat" "thin\_avg" "sqrt.ad.mort"   
## [16] "sqrt.ad.mort.c" "sqrt.alc" "sqrt.alc.c"   
## [19] "bmi.c" "polio.c" "diphtheria.c"   
## [22] "thin.c"

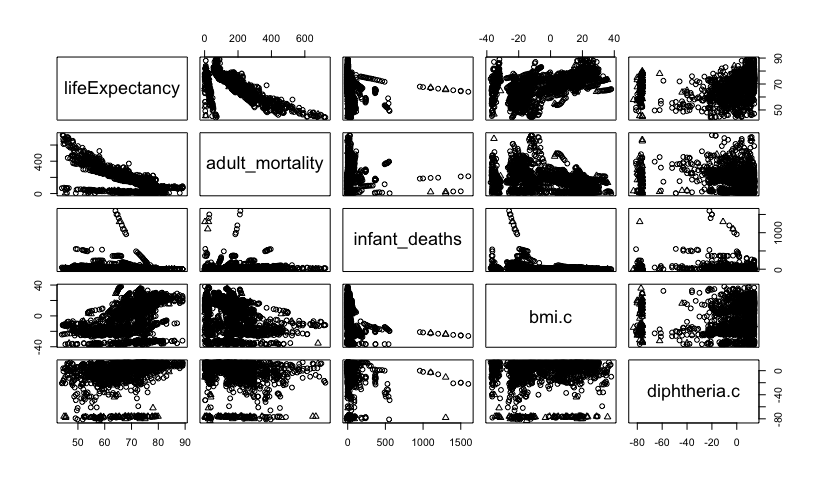
nrows = dim (life\_expectancy) [1]  
  
shapes = rep (1, nrows)  
shapes [c(740, 1867, 1869)] = 2:4  
sizes = rep (1, nrows)  
sizes [c(740, 1867, 1869)] = 2  
  
plot (life\_expectancy [,c(1:3,19,21)], pch=shapes, cex=sizes)



##Identify High Points Leverage  
  
life\_expectancy$leverage = hatvalues (aic\_cent)  
(lev.cut = 3 \* aic\_cent$rank / (aic\_cent$rank + aic\_cent$df.residual))

## [1] 0.08610478

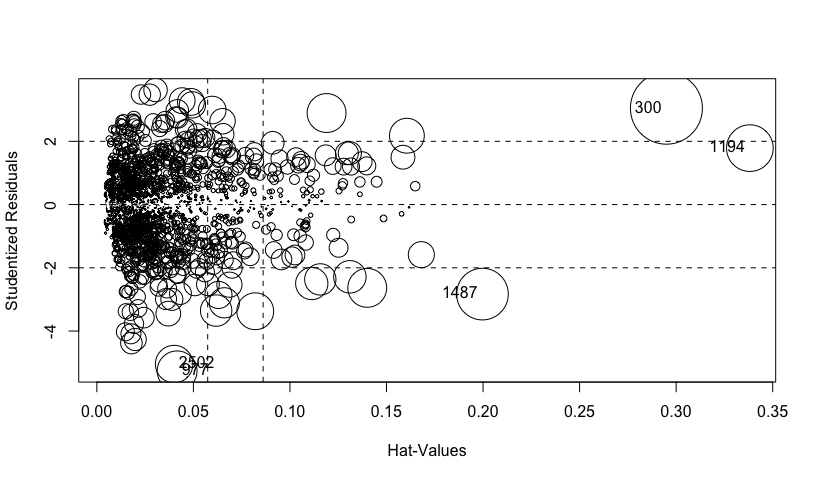
plot (life\_expectancy [,c(1:3,19,21)], pch=ifelse (life\_expectancy$leverage > lev.cut, 2, 1))



As shown in the scatter plot matrix, the influence points can be seen in the lower left corner on the relationship between life expectancy and adult mortality and life expectancy and bmi. For life expectancy and diphtheria these points are also visible in the left corner

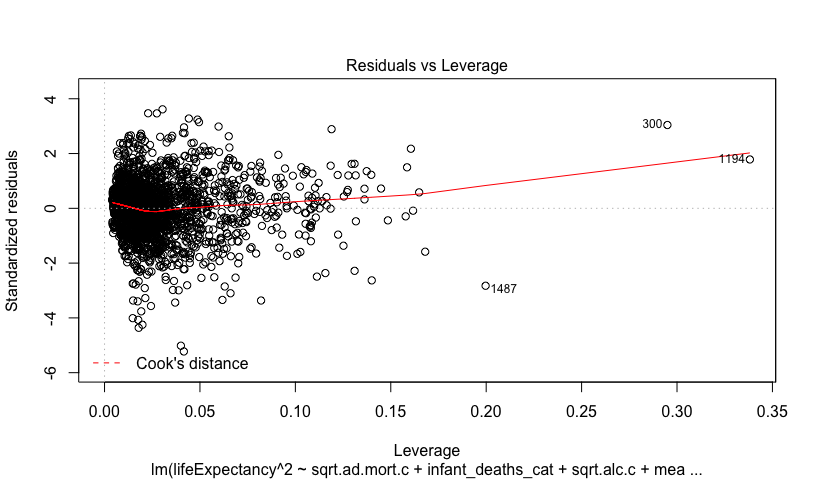
The points are also located more to the middle right side on the relationship between life expectancy and infant deaths. The graph is also very right skewed

influencePlot(aic\_cent)



## StudRes Hat CookD  
## 300 3.047363 0.29495981 0.06142881  
## 977 -5.258427 0.04156382 0.01879871  
## 1194 1.784663 0.33817708 0.02580650  
## 1487 -2.831114 0.19968894 0.03164045  
## 2502 -5.044237 0.04001691 0.01664483

plot\_aic=plot(aic\_cent, which=5)



Both plots are showing leverage value, the graph on the left shows different sizes of circles, notice that the value of 300 has a bigger size compared to the value of 1194, this shows the distance from the Cook’s distance. However, since the value is not at 5, there is no need for further analysis.

## Interpretations

### Interaction Plots

Let us now try to interpret some of the interaction effects that have the highest and lowest p-values:

1. It can be seen that sqrt.ad.mort.c and measles\_cat do not have a high significance (significant at the 0.01 level). Let us explore this:

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

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

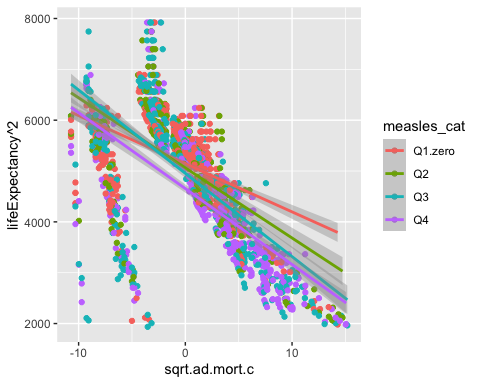
## The following object is masked from 'package:car':  
##   
## recode

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

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

categorize = function (x) {  
 quartiles = summary (x) [c(2, 3, 5)]  
 result = rep ("Q1", length (x))  
 result [which ((quartiles[1] < x) & (x <= quartiles [2]))] = "Q2"  
 result [which ((quartiles[2] < x) & (x <= quartiles [3]))] = "Q3"  
 result [which (quartiles[3] < x)] = "Q4"  
 return (result)  
}  
with(life\_expectancy,  
 qplot(x = sqrt.ad.mort.c, y = lifeExpectancy ^ 2, color = measles\_cat) +  
 geom\_smooth(method = 'lm'))

## `geom\_smooth()` using formula 'y ~ x'

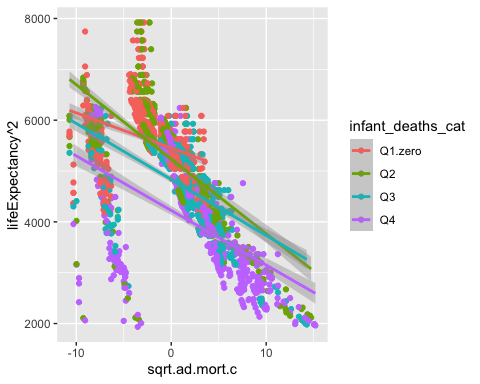


The slopes of sqrt.ad.mort.c and lifeExpectancy squared are not different depending on measles showing that this interaction is not significant.

1. Let us look at the the interaction between sqrt.ad.mort.c and infant\_deaths because this interaction seems significant at the 0.001 level:

with(life\_expectancy,  
 qplot(x = sqrt.ad.mort.c, y = lifeExpectancy ^ 2, color = infant\_deaths\_cat) +  
 geom\_smooth(method = 'lm'))

## `geom\_smooth()` using formula 'y ~ x'

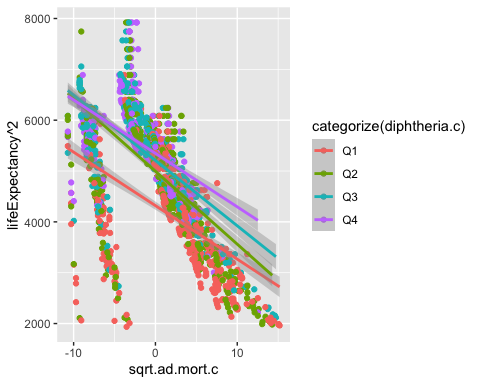


The slopes of sqrt.ad.mort.c and lifeExpectancy squared are different depending on infant deaths since the slopes showing the relationship between sqrt.ad.mort.c and lifeExpectancy squared for number of deaths from zero to quarter 1 and quarter 1 to quarter 2 are different from the second two lines.

1. The interaction between sqrt.ad.mort.c and diphtheria.c has a small p-value. Let us explore this:

with(life\_expectancy,  
 qplot(x = sqrt.ad.mort.c, y = lifeExpectancy ^ 2, color = categorize(diphtheria.c)) +  
 geom\_smooth(method = 'lm'))

## `geom\_smooth()` using formula 'y ~ x'

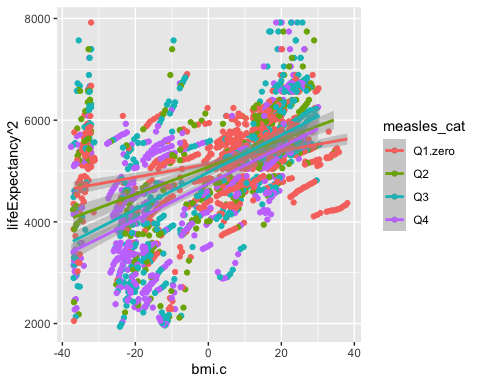


The plot shows that the relationship between sqrt.ad.mort.c and lifeExpectancy^2 is steepest for the second category of diphtheria and broadest for the fourth. This shows that the relationship between the two variables is significant for different values of diphtheria.

1. Let us look at the the interaction between bmi.c and measles\_cat because this interaction seems significant at the 0.001 level:

with(life\_expectancy,  
 qplot(x = bmi.c, y = lifeExpectancy ^ 2, color = measles\_cat) +  
 geom\_smooth(method = 'lm'))

## `geom\_smooth()` using formula 'y ~ x'

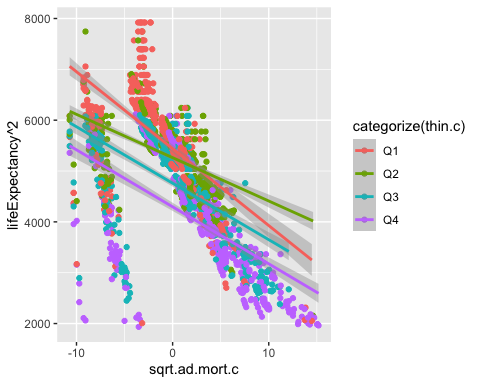


The plot shows that the relationship between sqrt.ad.mort.c and lifeExpectancy^2 is steeper for the third category of measles than for the others. This shows that the relationship between the two is different for different values of measles.

1. Let us look at the the interaction between thin.c and sqrt.ad.mort.c because this interaction has a low p-value and is only significant at the 0.1 significant level:

with(life\_expectancy,  
 qplot(x = sqrt.ad.mort.c, y = lifeExpectancy ^ 2, color = categorize(thin.c)) +  
 geom\_smooth(method = 'lm'))

## `geom\_smooth()` using formula 'y ~ x'



The plot shows that while, the relationship between sqrt.ad.mort.c and lifeExpectancy^2 is steeper for the first category of average thinness than for the others, the relationship between the two is not extremely different for different values of thinness.

It should be noted that all the variables referenced to in this section are centered variables, except for infant\_deaths\_cat and measles\_cat which are categorical variables.

### Confidence Intervals

pc\_aic = predict (aic\_cent,interval = "confidence")  
pc\_aic2 = as.data.frame(pc\_aic)  
sort\_aic = order(as.data.frame(pc\_aic)$fit)  
sqrt(as.data.frame(pc\_aic)[sort\_aic,][c(40,100,500,1000, 1500,2100),])

## fit lwr upr  
## 727 52.95781 49.90043 55.84807  
## 444 56.22562 55.15158 57.27952  
## 2162 64.42436 62.61023 66.18879  
## 2132 70.85774 69.38213 72.30325  
## 469 73.98189 72.32914 75.59852  
## 933 80.31232 79.45934 81.15634

There are only 6 values represented in this, which are lower, medium and bigger.

When the predicted life expectancy is 52.9 years, with 95% confidence the mean of life expectancy is between 49.9 and 55.8 years.

When the predicted life expectancy is 70.9 years, with 95% confidence the mean of life expectancy is between 69.4 and 72.3 years.

When the predicted life expectancy is 80.3 years, with 95% confidence the mean of life expectancy is between 79.5 and 81.2 years.

## Conclusion

After going through all the necessary steps to make the model more reliable and predictable, the final model is the “aic\_cent” which includes multiple variables that are significant to make a prediction. As mentioned, with step-wise regression on the centered interaction effect between multiple variables shows that certain health factors have an influence on life expectancy of a country depending on another health factor(s).

The final model has an adjusted R-squared of 0.71, which means that 71% of the variation on life expectancy squared can be explained by the final model. The residual standard error is 610 life expectancy squared.

Although the final model shows that health factors are statistically significant, but for further improvement, the model does require some transformation due to the high leverage points.