

# Body Fat Prediction

Lien Dao

5/2/2022

```
library(datasets)
library(readr)
url <- "https://docs.google.com/spreadsheets/d/e/2PACX-1vQPprnwKGsFQla0eq25Fnwb9fhSmsGBxSNU-cly8z-63fBB"
bodyfat <- read_csv(url)

#Response variable: BodyFat
#Predictor variables: 14

#libraries
library(ggplot2)
library(gggridges)
library(ggpubr)
library(corrplot)
library(dplyr)
library(tidyverse)
library(plotly)
library(caret)

#EDA
str(bodyfat)

## spec_tbl_df [252 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Density: num [1:252] 1.07 1.09 1.04 1.08 1.03 ...
## $ BodyFat: num [1:252] 12.3 6.1 25.3 10.4 28.7 20.9 19.2 12.4 4.1 11.7 ...
## $ Age     : num [1:252] 23 22 22 26 24 24 26 25 25 23 ...
## $ Weight  : num [1:252] 154 173 154 185 184 ...
## $ Height  : num [1:252] 67.8 72.2 66.2 72.2 71.2 ...
## $ Neck    : num [1:252] 36.2 38.5 34 37.4 34.4 39 36.4 37.8 38.1 42.1 ...
## $ Chest   : num [1:252] 93.1 93.6 95.8 101.8 97.3 ...
## $ Abdomen : num [1:252] 85.2 83 87.9 86.4 100 94.4 90.7 88.5 82.5 88.6 ...
## $ Hip     : num [1:252] 94.5 98.7 99.2 101.2 101.9 ...
## $ Thigh   : num [1:252] 59 58.7 59.6 60.1 63.2 66 58.4 60 62.9 63.1 ...
## $ Knee    : num [1:252] 37.3 37.3 38.9 37.3 42.2 42 38.3 39.4 38.3 41.7 ...
## $ Ankle   : num [1:252] 21.9 23.4 24 22.8 24 25.6 22.9 23.2 23.8 25 ...
## $ Biceps  : num [1:252] 32 30.5 28.8 32.4 32.2 35.7 31.9 30.5 35.9 35.6 ...
## $ Forearm : num [1:252] 27.4 28.9 25.2 29.4 27.7 30.6 27.8 29 31.1 30 ...
## $ Wrist   : num [1:252] 17.1 18.2 16.6 18.2 17.7 18.8 17.7 18.8 18.2 19.2 ...
## - attr(*, "spec")=
##   .. cols(
##     .. Density = col_double(),
##     .. BodyFat = col_double(),
##     .. Age = col_double(),
##     .. Weight = col_double(),
```

```

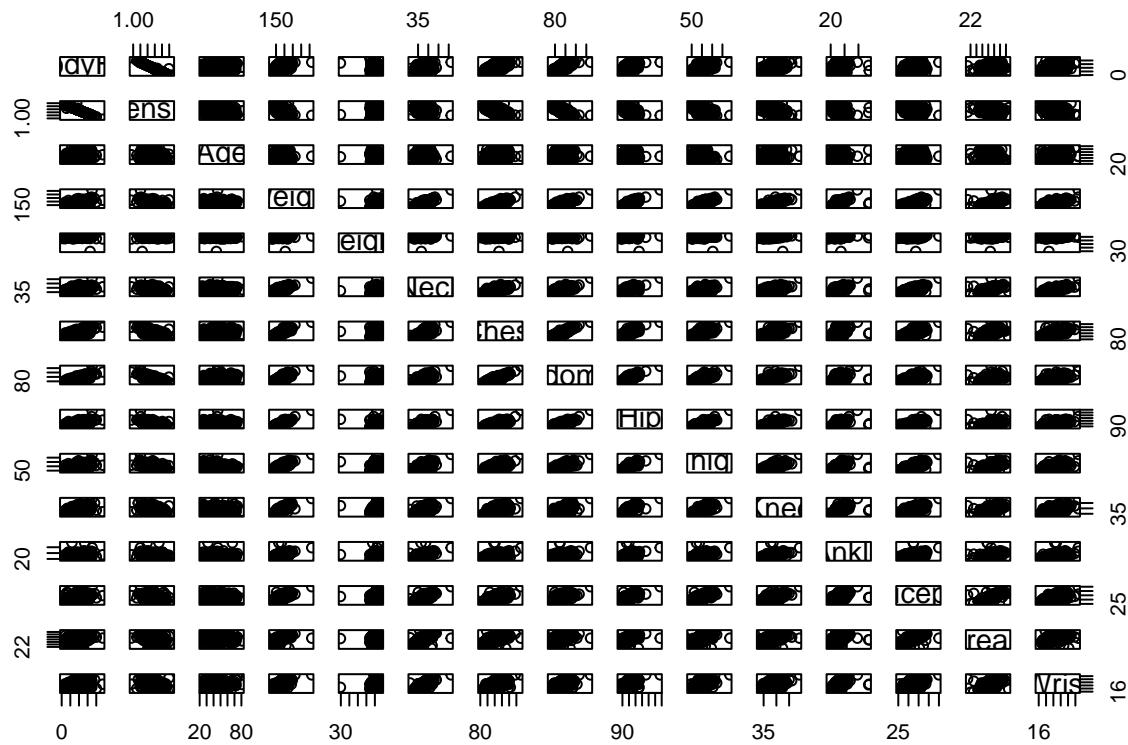
## .. Height = col_double(),
## .. Neck = col_double(),
## .. Chest = col_double(),
## .. Abdomen = col_double(),
## .. Hip = col_double(),
## .. Thigh = col_double(),
## .. Knee = col_double(),
## .. Ankle = col_double(),
## .. Biceps = col_double(),
## .. Forearm = col_double(),
## .. Wrist = col_double()
## ...
## - attr(*, "problems")=<externalptr>
summary(bodyfat)

##      Density      BodyFat       Age       Weight
## Min.  :0.995   Min.   : 0.00   Min.  :22.00   Min.  :118.5
## 1st Qu.:1.041  1st Qu.:12.47  1st Qu.:35.75  1st Qu.:159.0
## Median :1.055  Median  :19.20  Median :43.00  Median :176.5
## Mean   :1.056  Mean    :19.15  Mean   :44.88  Mean   :178.9
## 3rd Qu.:1.070  3rd Qu.:25.30  3rd Qu.:54.00  3rd Qu.:197.0
## Max.   :1.109  Max.    :47.50  Max.   :81.00  Max.   :363.1
##      Height       Neck       Chest      Abdomen
## Min.  :29.50   Min.   :31.10   Min.  : 79.30   Min.  : 69.40
## 1st Qu.:68.25  1st Qu.:36.40  1st Qu.: 94.35  1st Qu.: 84.58
## Median :70.00   Median  :38.00   Median : 99.65  Median : 90.95
## Mean   :70.15   Mean    :37.99   Mean   :100.82  Mean   : 92.56
## 3rd Qu.:72.25  3rd Qu.:39.42  3rd Qu.:105.38 3rd Qu.: 99.33
## Max.   :77.75   Max.    :51.20   Max.   :136.20  Max.   :148.10
##      Hip        Thigh       Knee      Ankle      Biceps
## Min.  : 85.0   Min.   :47.20   Min.  : 33.00   Min.  :19.1   Min.  :24.80
## 1st Qu.: 95.5  1st Qu.:56.00  1st Qu.: 36.98  1st Qu.:22.0   1st Qu.:30.20
## Median : 99.3   Median  :59.00   Median : 38.50  Median :22.8   Median :32.05
## Mean   : 99.9   Mean    :59.41   Mean   : 38.59  Mean   :23.1   Mean   :32.27
## 3rd Qu.:103.5  3rd Qu.:62.35  3rd Qu.: 39.92 3rd Qu.:24.0   3rd Qu.:34.33
## Max.   :147.7   Max.    :87.30   Max.   : 49.10  Max.   :33.9   Max.   :45.00
##      Forearm      Wrist
## Min.  :21.00   Min.   :15.80
## 1st Qu.:27.30  1st Qu.:17.60
## Median :28.70   Median  :18.30
## Mean   :28.66   Mean    :18.23
## 3rd Qu.:30.00  3rd Qu.:18.80
## Max.   :34.90   Max.   :21.40

df <- bodyfat[-c(42, 182),]

pairs(BodyFat~Density+Age+Weight+Height+Neck+Chest+Abdomen+Hip+Thigh
      +Knee+Ankle+Biceps+Forearm+Wrist, data=bodyfat, cex.labels=1.4)

```



```

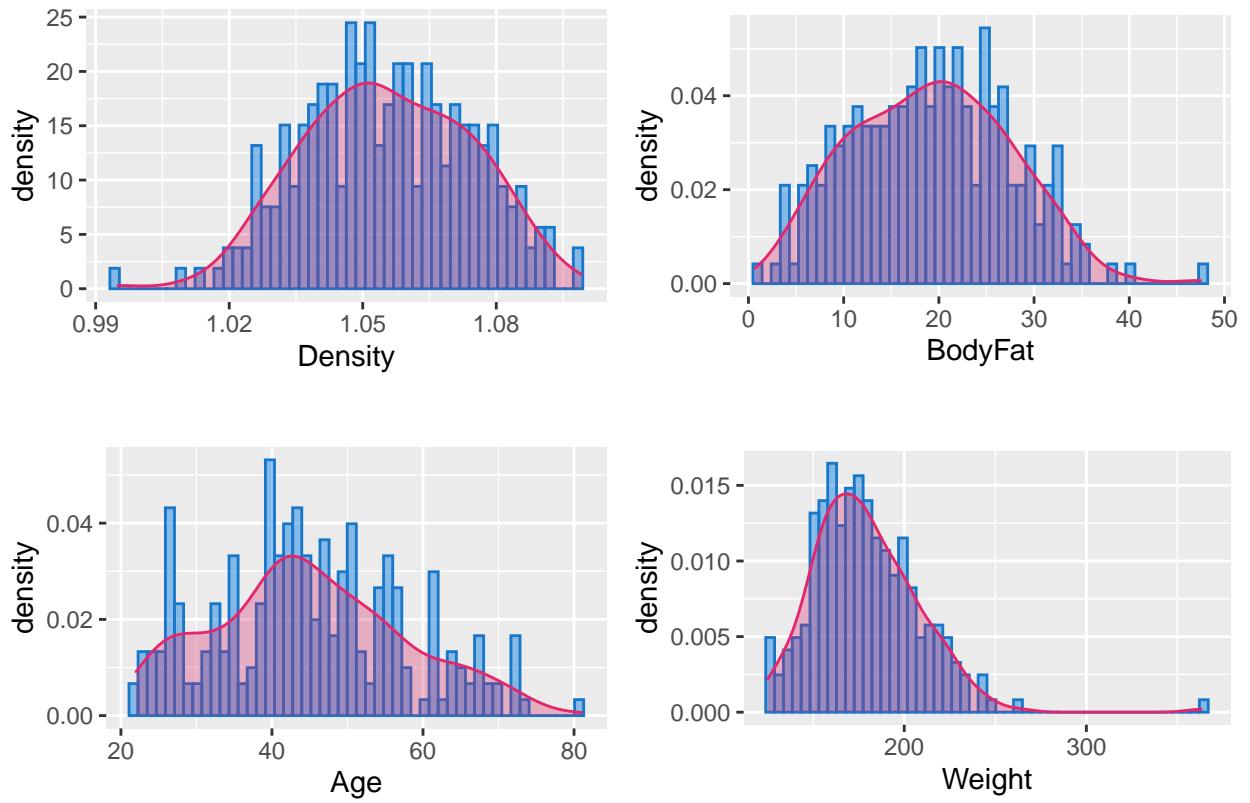
plot_hist_dens <- function(x, na.rm = TRUE, ...) {
  name <- names(x)
  plot_lst <- vector("list", length = length(name))

  for (i in seq_along(name)) {
    p <- ggplot(x,aes_string(y = name[i])) +
      geom_histogram(aes(y = ..density..),bins = 50,color = "#1877C9",
                     fill = "#1E88E5",alpha=0.5) +
      geom_density(color = "#E4276C", fill = "#D81B60", alpha = 0.3) +
      theme(aspect.ratio=9/16)
    plot_lst[[i]] <- p
  }

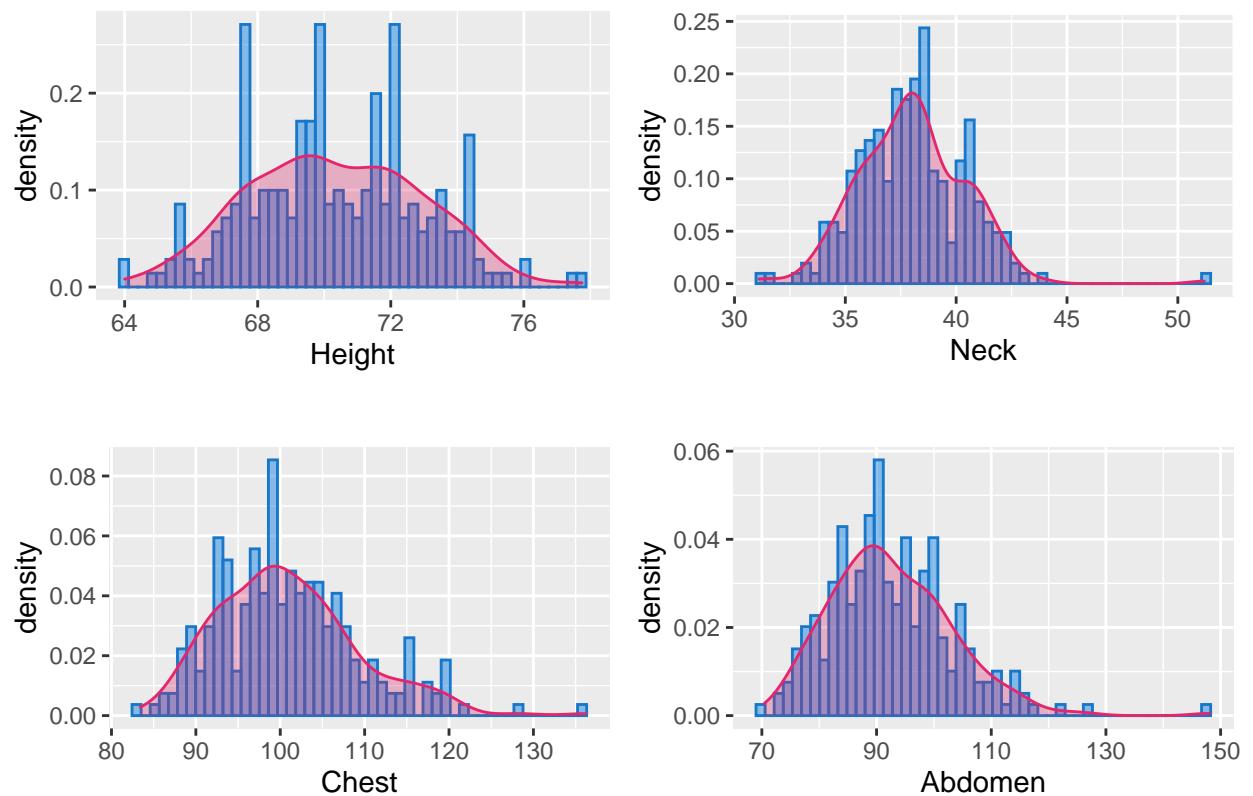
  cowplot::plot_grid(plotlist = plot_lst, nrow = ceiling(length(name)/2))
}

plot_hist_dens(df[,1:4])

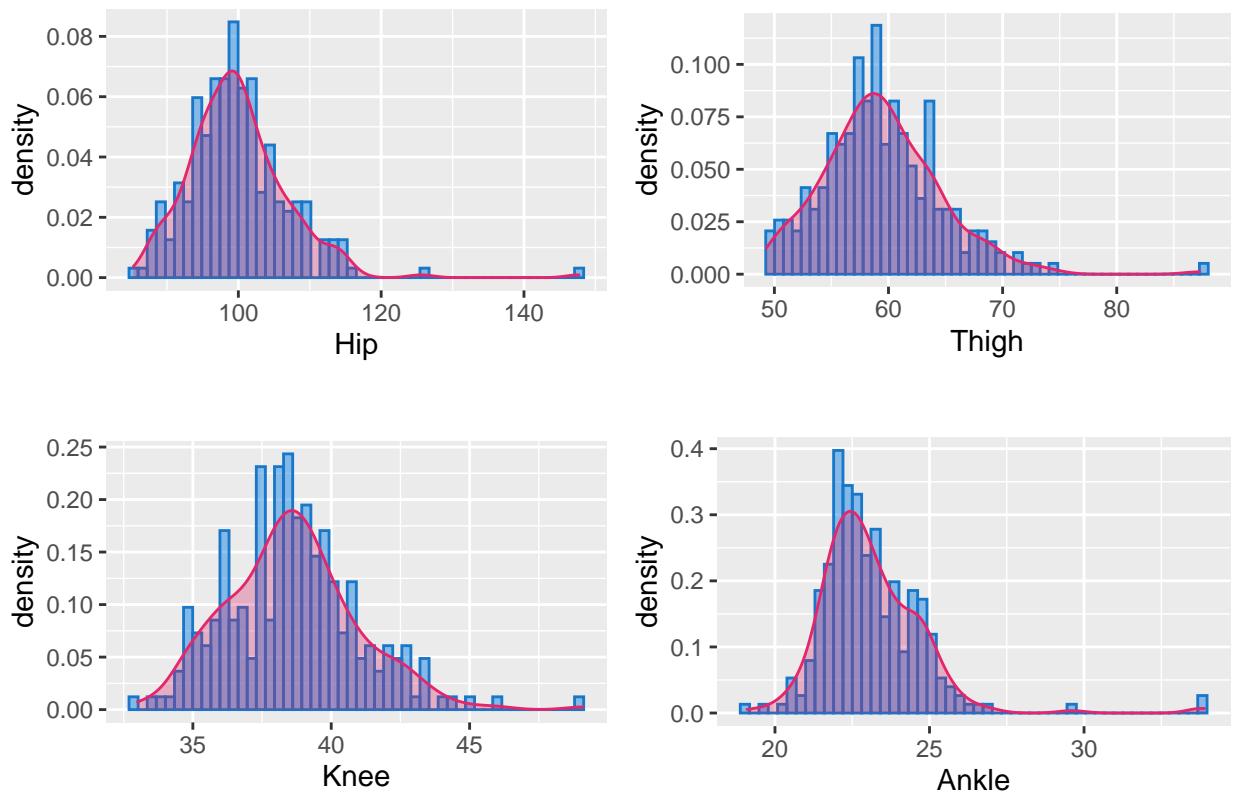
```



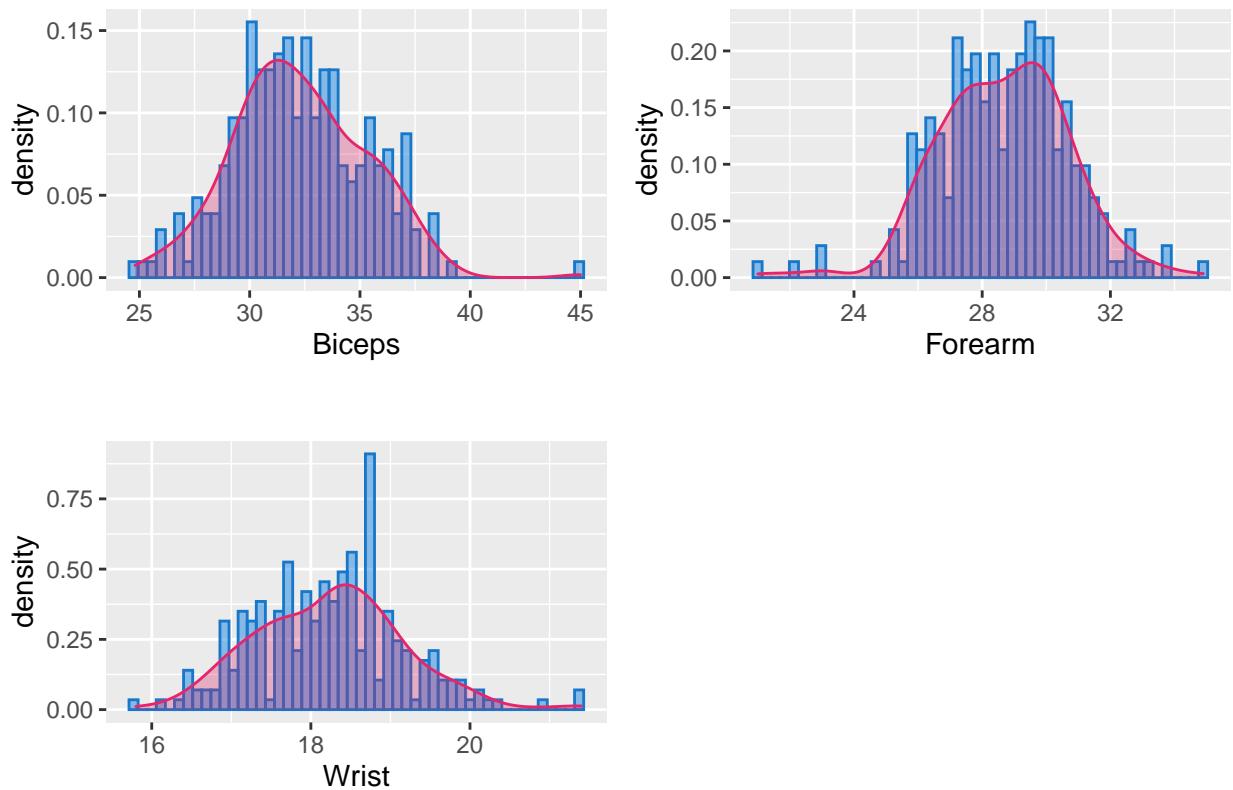
```
plot_hist_dens(df[,5:8])
```



```
plot_hist_dens(df[,9:12])
```



```
plot_hist_dens(df[,13:15])
```



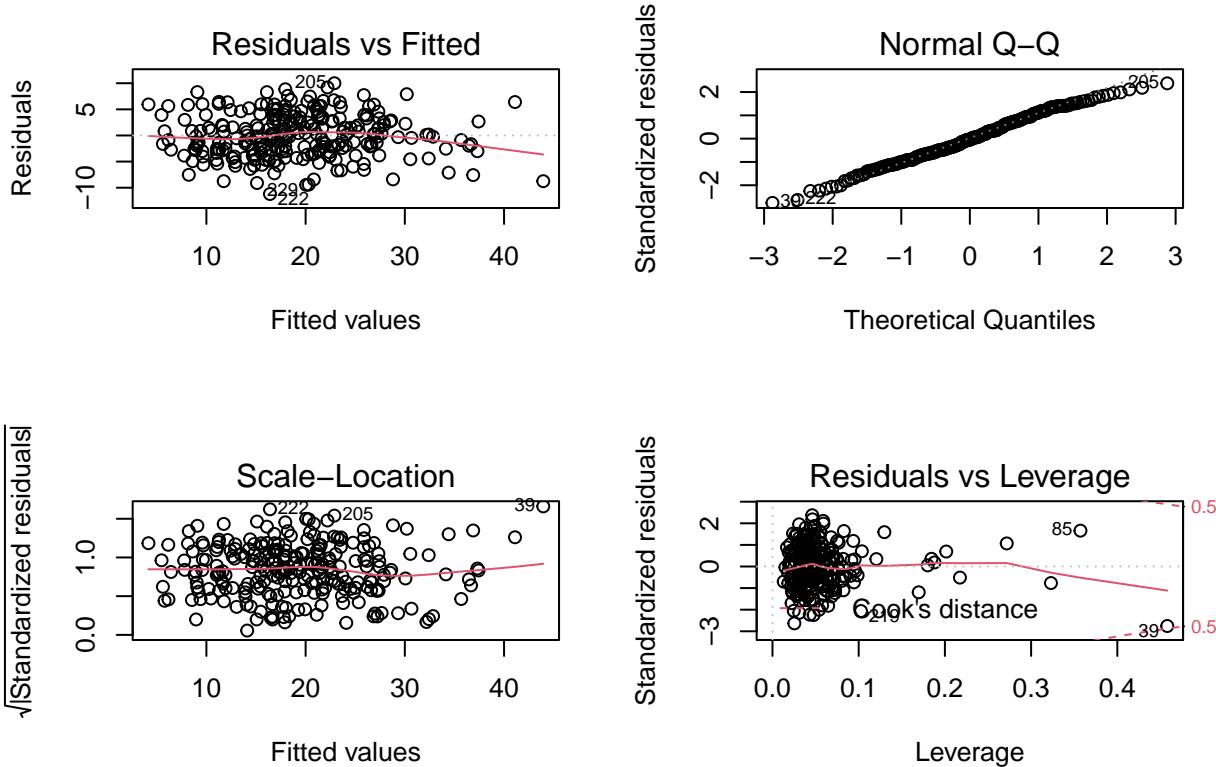
```
m0 <- lm(BodyFat~., data=df[,-1])
summary(m0)
```

```
##
## Call:
## lm(formula = BodyFat ~ ., data = df[, -1])
##
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -11.208 -2.856 -0.165  3.208 10.000 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -17.16255  22.45806 -0.764   0.4455  
## Age          0.06105   0.03242  1.883   0.0610 .  
## Weight       -0.08651   0.06238 -1.387   0.1668  
## Height       -0.06076   0.17950 -0.339   0.7353  
## Neck          -0.46181   0.23634 -1.954   0.0519 .  
## Chest         -0.03449   0.10435 -0.331   0.7413  
## Abdomen       0.95937   0.09043 10.609 <2e-16 ***
## Hip           -0.20924   0.14689 -1.424   0.1556  
## Thigh         0.22246   0.14770  1.506   0.1334  
## Knee          0.00683   0.24818  0.028   0.9781  
## Ankle         0.16661   0.22303  0.747   0.4558  
## Biceps        0.19512   0.17294  1.128   0.2604  
## Forearm       0.43907   0.19989  2.197   0.0290 * 
```

```

## Wrist      -1.63186   0.53617  -3.044   0.0026  **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.313 on 236 degrees of freedom
## Multiple R-squared:  0.7421, Adjusted R-squared:  0.7279
## F-statistic: 52.24 on 13 and 236 DF,  p-value: < 2.2e-16
par(mfrow = c(2, 2))
plot(m0)

```



```

#make this example reproducible
set.seed(666)

#create ID column
df$ids <- 1:nrow(df)

#use 80% of dataset as training set and 20% as test set
train <- df %>% dplyr::sample_frac(0.80)
test  <- dplyr::anti_join(df, train, by = 'ids')

train <- train[,-c(1,16)]
test <- test[,-c(1,16)]

dim(train)

## [1] 200 14

```

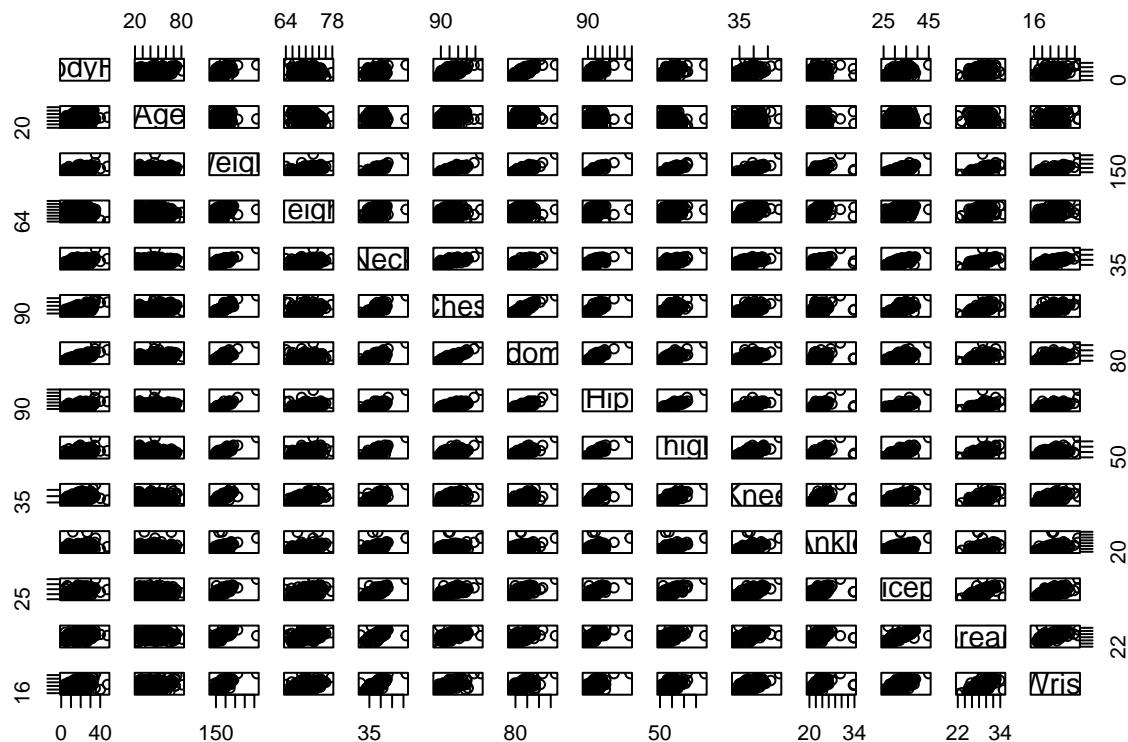
```

dim(test)

## [1] 50 14

pairs(BodyFat~Age+Weight+Height+Neck+Chest+Abdomen+Hip+Thigh
      +Knee+Ankle+Biceps+Forearm+Wrist, data=train, cex.labels=1.4)

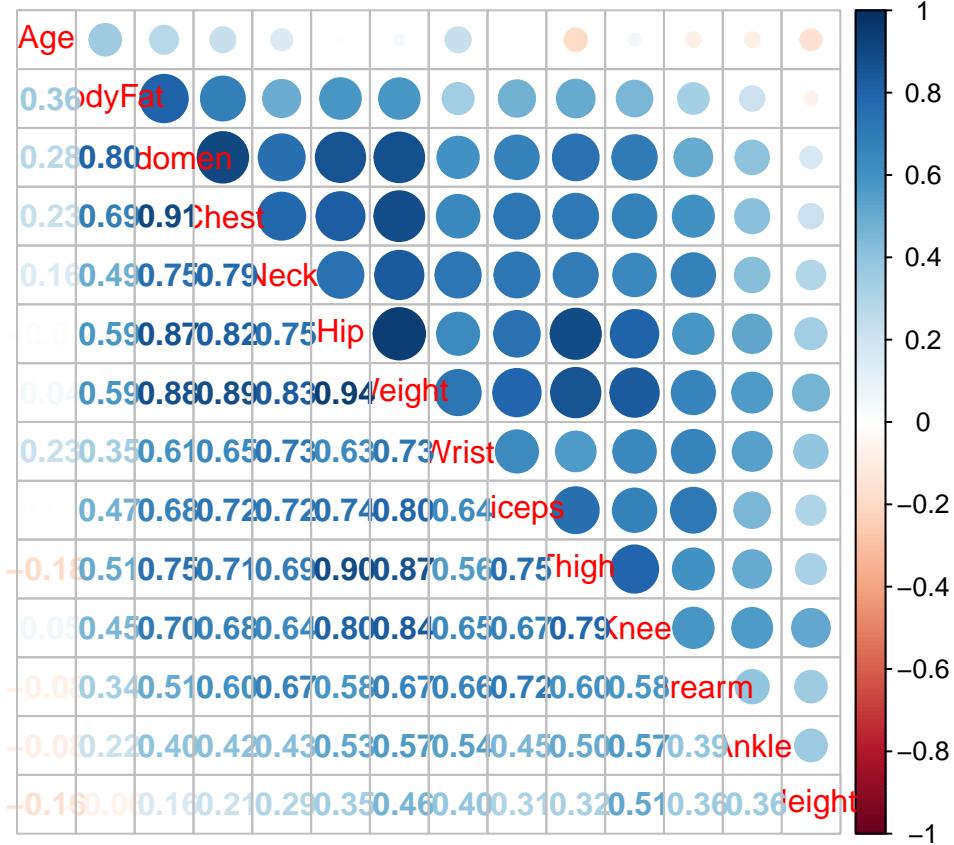
```



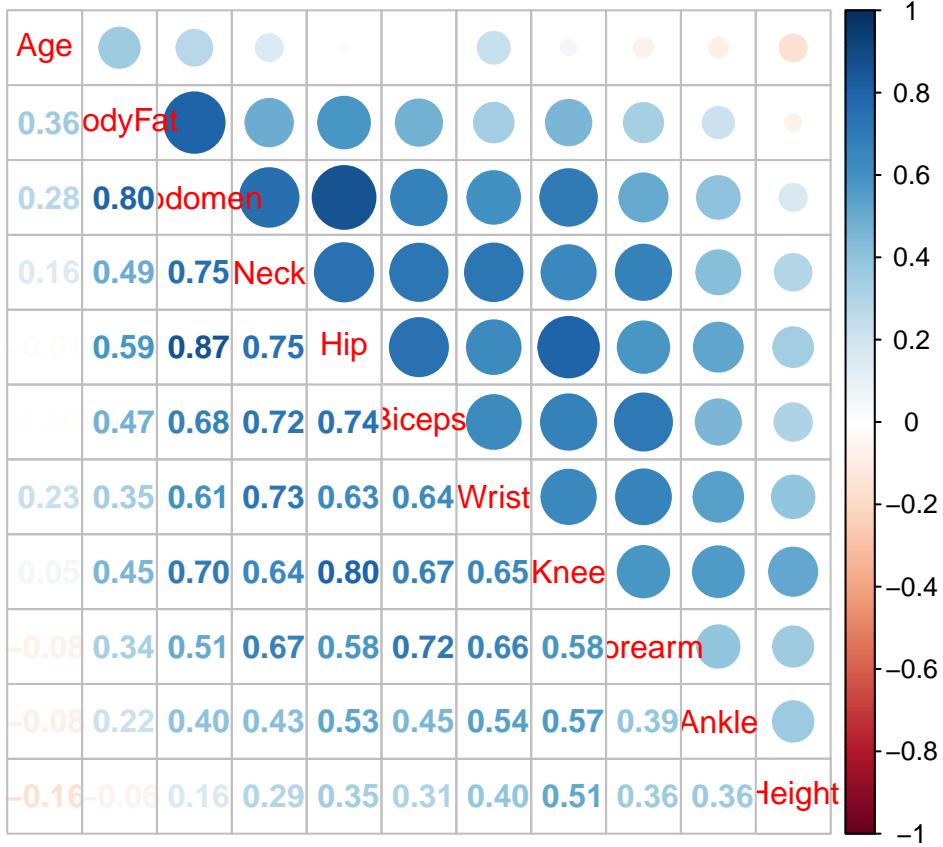
```

corrplot.mixed(cor(train), order = 'AOE')

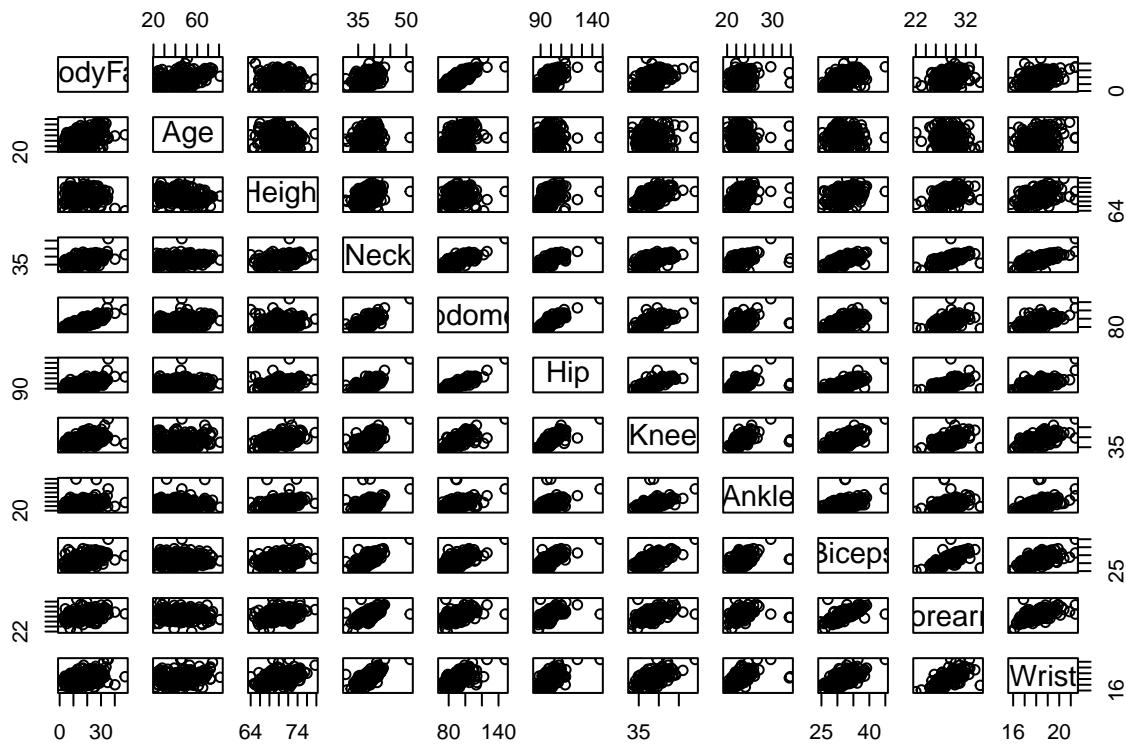
```



```
df_train <- train[,-c(3,6,9)]
corrplot.mixed(cor(df_train), order = 'AOE')
```



```
pairs(BodyFat~., data=df_train, cex.labels=1.4)
```



```

m1 <- lm(BodyFat ~ ., data = df_train)
summary(m1)

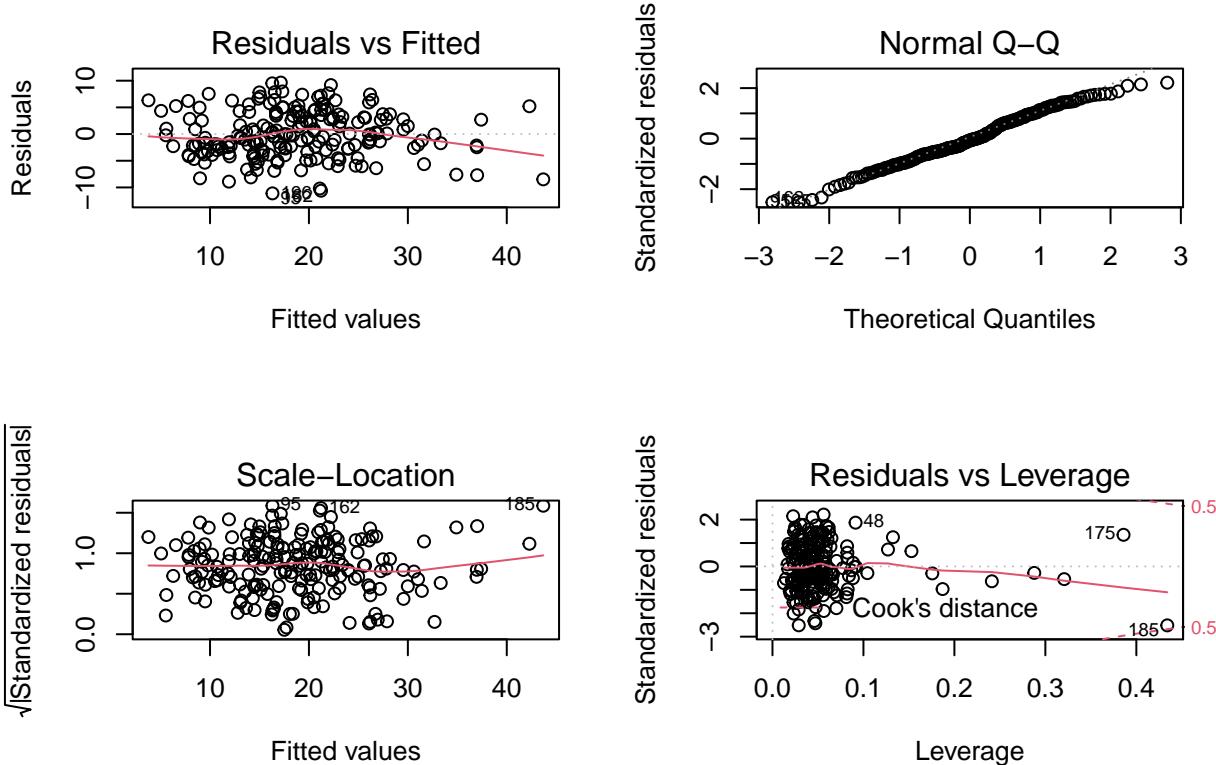
##
## Call:
## lm(formula = BodyFat ~ ., data = df_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -11.1131  -2.7926  -0.2899  3.3935  9.6439 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 12.37194   9.95687   1.243   0.21557  
## Age         0.07015   0.03356   2.090   0.03792 *  
## Height     -0.32333   0.16119  -2.006   0.04630 *  
## Neck        -0.57721   0.25430  -2.270   0.02435 *  
## Abdomen     0.88443   0.07901  11.194 < 2e-16 *** 
## Hip         -0.28683   0.13129  -2.185   0.03014 *  
## Knee        0.00956   0.26269   0.036   0.97101    
## Ankle       0.15489   0.23699   0.654   0.51418    
## Biceps      0.12738   0.18816   0.677   0.49924    
## Forearm     0.61789   0.27713   2.230   0.02695 *  
## Wrist       -1.69720   0.60898  -2.787   0.00586 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

## 
## Residual standard error: 4.483 on 189 degrees of freedom
## Multiple R-squared:  0.7328, Adjusted R-squared:  0.7186 
## F-statistic: 51.83 on 10 and 189 DF,  p-value: < 2.2e-16
par(mfrow = c(2, 2))
plot(m1)

```



```

N <- 252
cutoff <- 4/(N-2)
cutoff

## [1] 0.016

cooksD <- cooks.distance(m1)
influential <- cooksD[(cooksD > cutoff)]
influential

##      38       48       54       75       94       95      106
## 0.01840487 0.03193728 0.01718092 0.02182240 0.01972390 0.01710774 0.01645766
##      108      154      162      163      175      179      185
## 0.02161278 0.01677926 0.02651845 0.02657077 0.10428284 0.01954974 0.44170200
##      196
## 0.02387093

library(alr3)

## Loading required package: car
## Loading required package: carData

```

```

## 
## Attaching package: 'car'

## The following object is masked from 'package:purrr':
## 
##      some

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

summary(powerTransform(cbind(BodyFat, Age, Height, Neck, Abdomen, Hip, Knee,
                             Ankle, Biceps, Forearm, Wrist) ~ 1, data=df_train))

## bcPower Transformations to Multinormality
##          Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## BodyFat    0.9359      1.00     0.7409    1.1310
## Age        0.8079      1.00     0.4037    1.2121
## Height     0.3890      1.00    -2.8279    3.6059
## Neck       -0.1356     0.00    -1.1141    0.8430
## Abdomen    -0.4383     0.00    -0.9943    0.1176
## Hip        -2.3180     -2.00   -3.1712   -1.4647
## Knee       -2.0975     -1.00   -3.3983   -0.7967
## Ankle      -5.5699     -5.57   -6.7839   -4.3560
## Biceps     -0.3688     0.00   -1.2751    0.5375
## Forearm    -0.2328     0.00   -1.3581    0.8925
## Wrist      -1.7464     -1.00  -3.4004   -0.0925
## 
## Likelihood ratio test that transformation parameters are equal to 0
## (all log transformations)
##                               LRT df      pval
## LR test, lambda = (0 0 0 0 0 0 0 0 0 0) 285.6713 11 < 2.22e-16
## 
## Likelihood ratio test that no transformations are needed
##                               LRT df      pval
## LR test, lambda = (1 1 1 1 1 1 1 1 1 1) 245.2517 11 < 2.22e-16

df_train$logNeck <- log(df_train$Neck)
df_train$logAbdomen <- log(df_train$Abdomen)
df_train$newHip <- 1/(df_train$Hip^2)
df_train$newKnee <- 1/(df_train$Knee)
df_train$logBiceps <- log(df_train$Biceps)
df_train$logForearm <- log(df_train$Forearm)
df_train$newWrist <- 1/(df_train$Wrist)

new_dftrain <- df_train[,-c(4,5,6,7,9,10,11)]

m2<- lm(BodyFat ~ ., data = new_dftrain)
summary(m2)

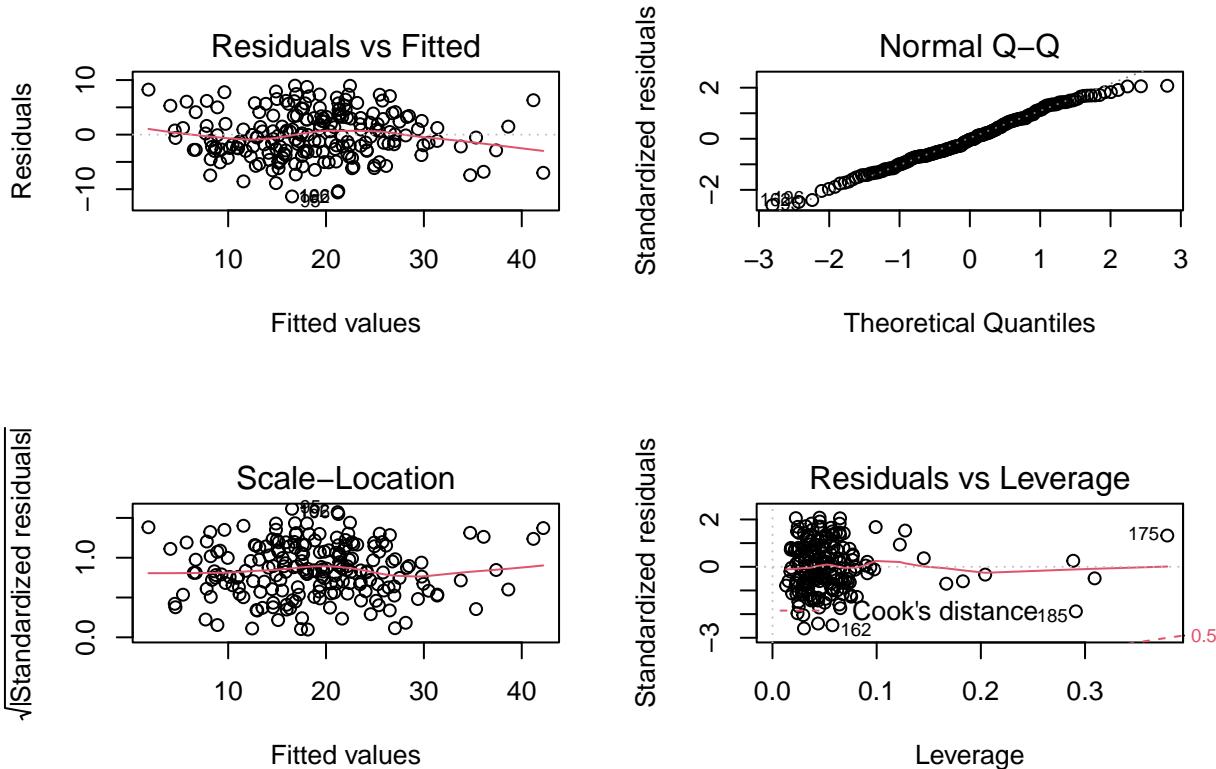
## 
## Call:
## lm(formula = BodyFat ~ ., data = new_dftrain)
## 
## Residuals:
##      Min      1Q  Median      3Q      Max
## -11.329  -2.899  -0.175   3.247   8.973

```

```

## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -3.186e+02  5.748e+01 -5.544 9.86e-08 ***
## Age          8.811e-02  3.266e-02  2.698 0.00761 **  
## Height       -4.119e-01  1.601e-01 -2.573 0.01084 *   
## Ankle         1.229e-01  2.321e-01  0.530 0.59694    
## logNeck      -2.097e+01  9.330e+00 -2.248 0.02573 *   
## logAbdomen    7.398e+01  7.294e+00 10.143 < 2e-16 ***
## newHip        2.186e+04  7.009e+04  0.312 0.75553    
## newKnee       2.010e+02  4.020e+02  0.500 0.61758    
## logBiceps     3.362e+00  5.947e+00  0.565 0.57252    
## logForearm    1.516e+01  8.034e+00  1.886 0.06077 .  
## newWrist      5.731e+02  2.014e+02  2.845 0.00493 **  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 4.415 on 189 degrees of freedom
## Multiple R-squared:  0.7408, Adjusted R-squared:  0.7271 
## F-statistic: 54.02 on 10 and 189 DF,  p-value: < 2.2e-16
par(mfrow = c(2, 2))
plot(m2)

```

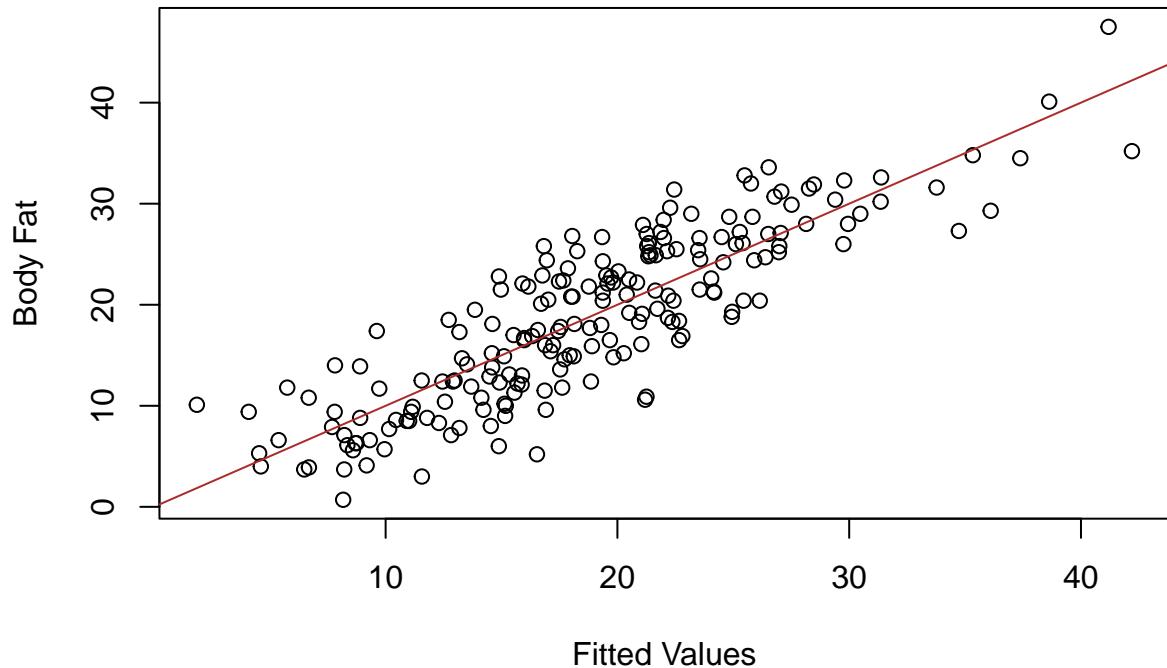


```

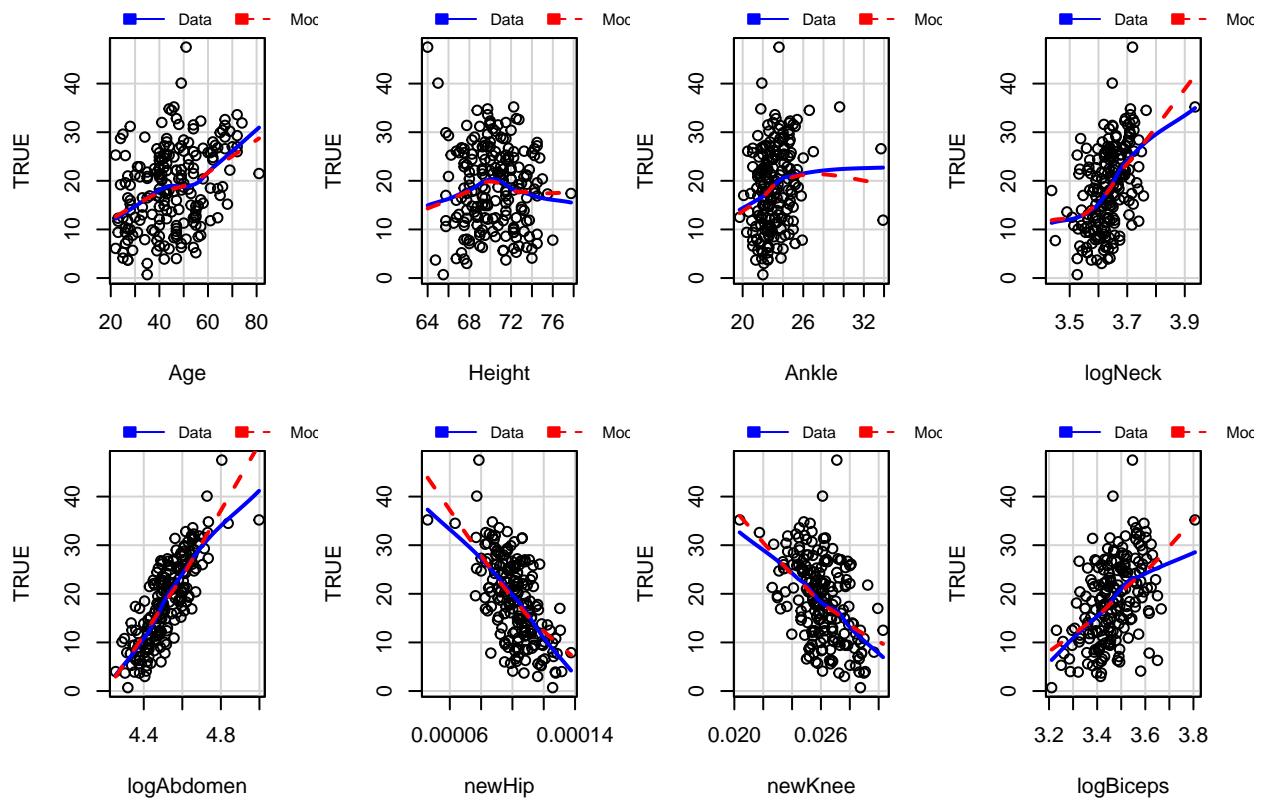
#A plot of BodyFat against fitted values with a straight line added
par(mfrow=c(1,1))
plot(m2$fitted.values,new_dftrain$BodyFat,xlab="Fitted Values",ylab="Body Fat")

```

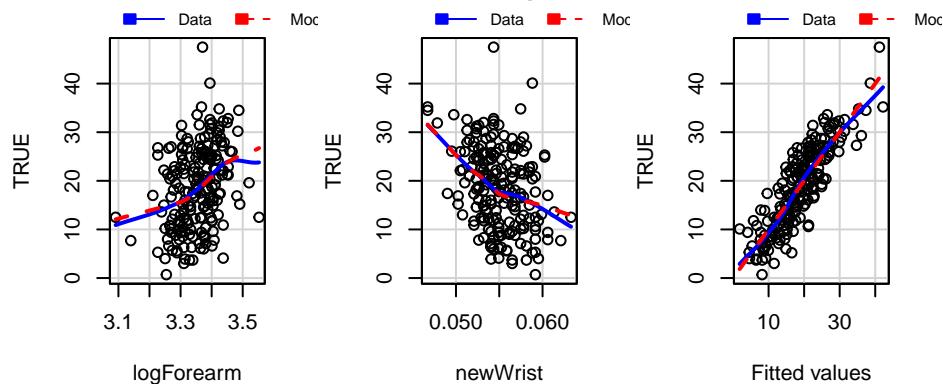
```
abline(lsfit(m2$fitted.values,new_dftrain$BodyFat),col="#aa2c2c")
```



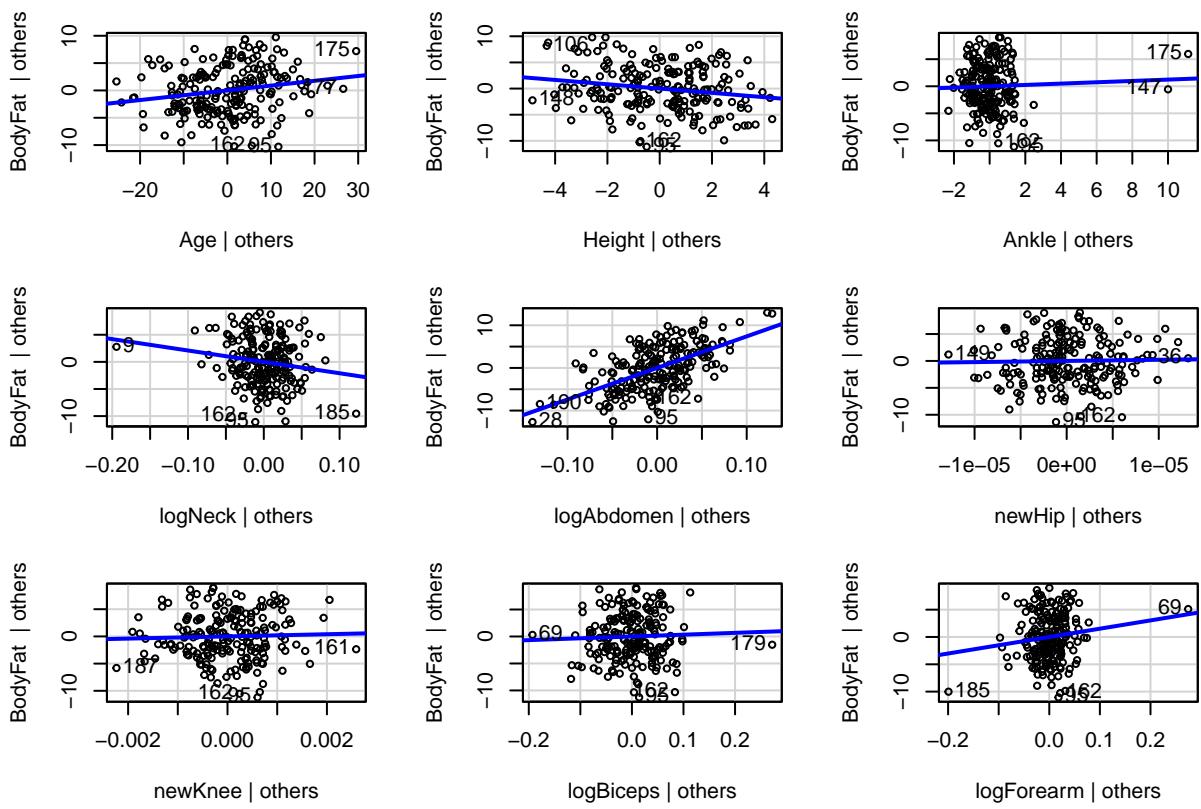
```
library(alr3)
mmps(m2,layout=c(2,4))
```



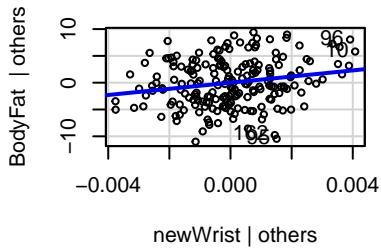
### Marginal Model Plots



```
library(car)
avPlots(m2)
```



## Added-Variable Plots



```

library(car)
vif(m2)

##          Age      Height      Ankle    logNeck logAbdomen     newHip     newKnee
##  1.803144  1.666269  1.697590  3.514741  6.880452  8.722456  4.192326
##  logBiceps logForearm  newWrist
##  3.161898  2.972972  3.327019

# A number of these variance inflation factors exceed 5, the cut-off often used, and so
# the associated regression coefficients are poorly estimated due to multicollinearity.

attach(new_dftrain)
#vars <- cbind(Age,Height,Neck,Ankle,Biceps,abdomen_hip,logWeight,
#               # newChest,newKnee,newForearm,logWrist,logThigh)
vars <- cbind(Age,Height,Ankle,newKnee,logNeck,logAbdomen,newHip,logBiceps,
              logForearm,newWrist)
library(leaps)
all_subsets <- regsubsets(as.matrix(vars),BodyFat,nvmax=10)
rs <- summary(all_subsets)
rs

## Subset selection object
## 10 Variables (and intercept)
##          Forced in Forced out
##  Age           FALSE      FALSE
##  Height        FALSE      FALSE
##  Ankle         FALSE      FALSE
##  newKnee       FALSE      FALSE

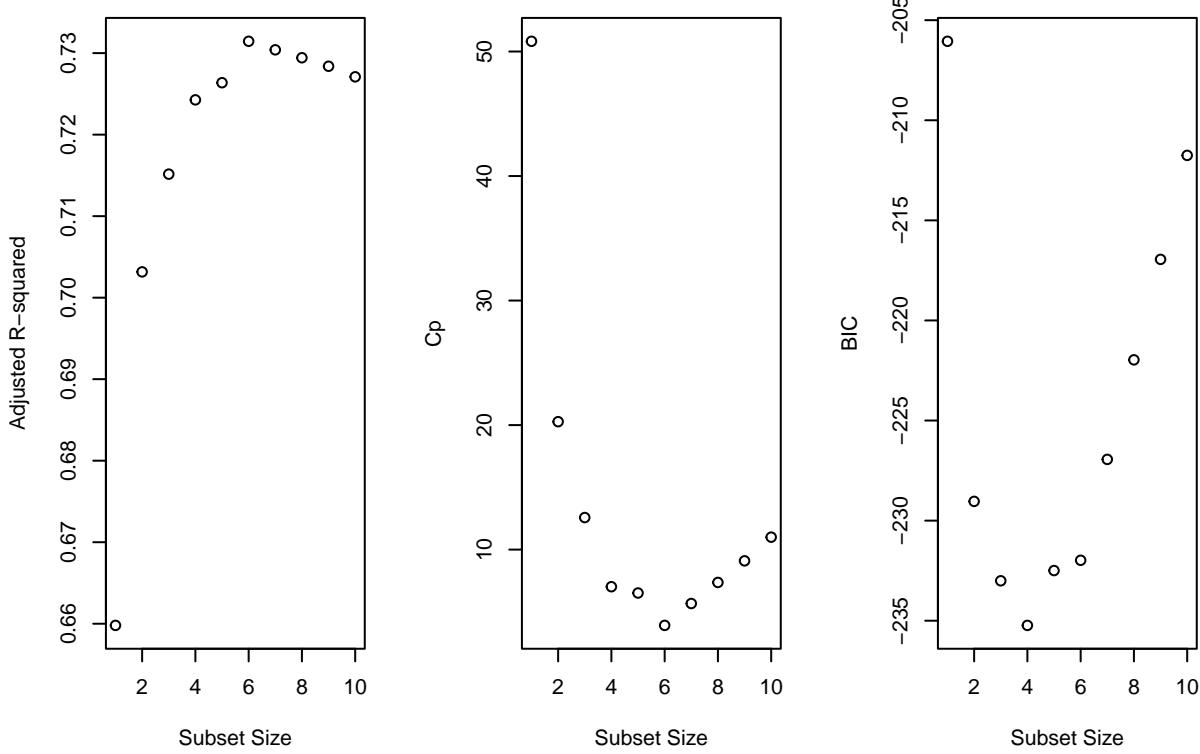
```

```

## logNeck      FALSE      FALSE
## logAbdomen   FALSE      FALSE
## newHip       FALSE      FALSE
## logBiceps    FALSE      FALSE
## logForearm   FALSE      FALSE
## newWrist     FALSE      FALSE
## 1 subsets of each size up to 10
## Selection Algorithm: exhaustive
##          Age Height Ankle newKnee logNeck logAbdomen newHip logBiceps
## 1  ( 1 ) " " " " " " " " *" " " " "
## 2  ( 1 ) " " "*" " " " " " *" " " " "
## 3  ( 1 ) " " "*" " " " " " *" " *" " " "
## 4  ( 1 ) "*" "*" " " " " " " *" " *" " " "
## 5  ( 1 ) "*" "*" " " " " " *" " *" " " " "
## 6  ( 1 ) "*" "*" " " " " " *" " *" " " " "
## 7  ( 1 ) "*" "*" " " " " *" " *" " *" " " "
## 8  ( 1 ) "*" "*" " " " " *" " *" " *" " *"
## 9  ( 1 ) "*" "*" " *" " *" " *" " *" " " " "
## 10 ( 1 ) "*" "*" " *" " *" " *" " *" " *" " *"
##          logForearm newWrist
## 1  ( 1 ) " " " "
## 2  ( 1 ) " " " "
## 3  ( 1 ) " " " "
## 4  ( 1 ) " " "*" "
## 5  ( 1 ) " " "*" "
## 6  ( 1 ) "*" "*" "
## 7  ( 1 ) "*" "*" "
## 8  ( 1 ) "*" "*" "
## 9  ( 1 ) "*" "*" "
## 10 ( 1 ) "*" "*" "
ls(rs)

## [1] "adjr2"   "bic"     "cp"      "obj"      "outmat"  "rsq"      "rss"      "which"
par(mfrow=c(1,3))
plot(1:10,rs$adjr2,xlab="Subset Size",ylab="Adjusted R-squared")
plot(1:10,rs$cp,xlab="Subset Size",ylab="Cp")
plot(1:10,rs$bic,xlab="Subset Size",ylab="BIC")

```



```
library(car)
par(mfrow=c(1,3))
subsets(all_subsets,statistic=c("adjr2"),legend = T)
```

```
##          Abbreviation
## Age           Ag
## Height        H
## Ankle         An
## newKnee       nK
## logNeck       lN
## logAbdomen    lA
## newHip        nH
## logBiceps     lB
## logForearm    lF
## newWrist      nW
```

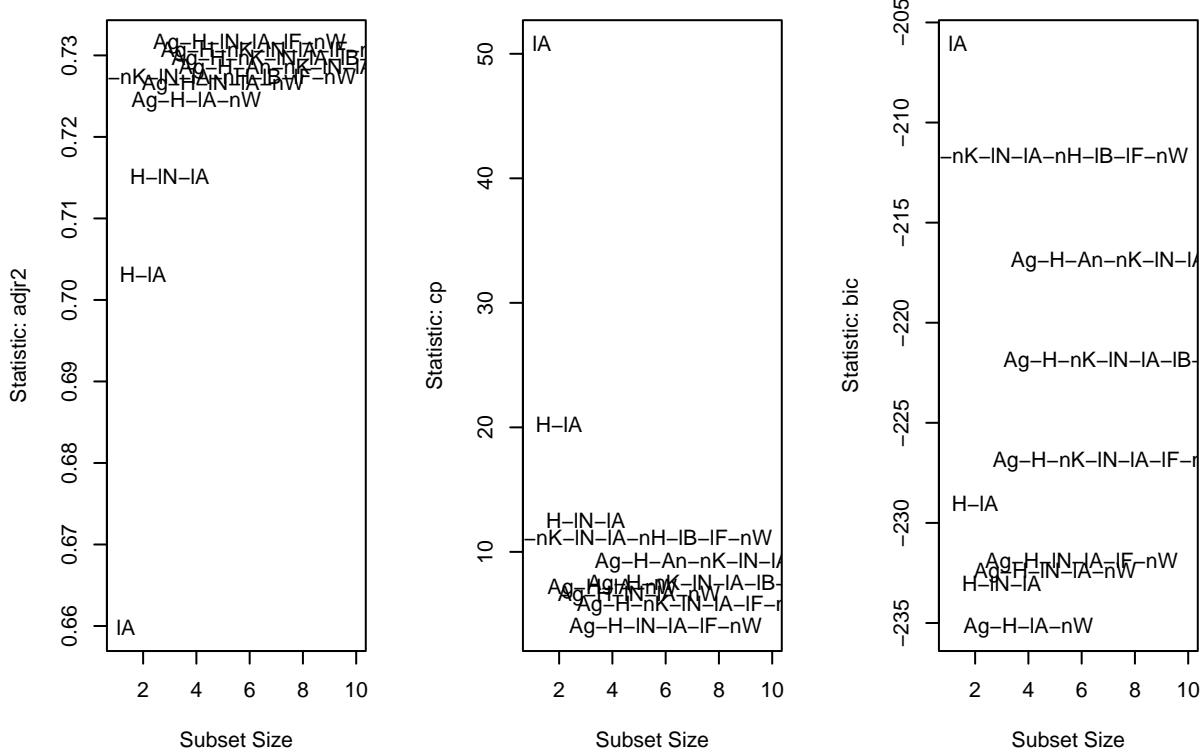
```
subsets(all_subsets,statistic=c("cp"),legend = T)
```

```
##          Abbreviation
## Age           Ag
## Height        H
## Ankle         An
## newKnee       nK
## logNeck       lN
## logAbdomen    lA
## newHip        nH
## logBiceps     lB
```

```

## logForearm          lF
## newWrist            nW
subsets(all_subsets,statistic=c("bic"),legend = T)

```



```

##           Abbreviation
## Age             Ag
## Height          H
## Ankle           An
## newKnee         nK
## logNeck         lN
## logAbdomen      lA
## newHip          nH
## logBiceps       lB
## logForearm      lF
## newWrist         nW
data.frame(
  Adj.R2 = which.max(rs$adjr2),
  CP = which.min(rs$cp),
  BIC = which.min(rs$bic))

```

```

##   Adj.R2 CP BIC
## 1     6   6   4

```

\*Backward elimination based on AIC

```

N <- length(m2$residuals)
backAIC <- step(m2,direction="backward", data=new_dftrain)

```

```

## Start: AIC=604.68
## BodyFat ~ Age + Height + Ankle + logNeck + logAbdomen + newHip +
##      newKnee + logBiceps + logForearm + newWrist
##
##          Df Sum of Sq   RSS   AIC
## - newHip     1     1.89 3685.7 602.78
## - newKnee    1     4.87 3688.7 602.94
## - Ankle      1     5.47 3689.3 602.97
## - logBiceps   1     6.23 3690.0 603.01
## <none>           3683.8 604.68
## - logForearm  1    69.36 3753.2 606.41
## - logNeck     1    98.50 3782.3 607.95
## - Height      1   129.05 3812.9 609.56
## - Age         1   141.88 3825.7 610.24
## - newWrist    1   157.79 3841.6 611.07
## - logAbdomen   1  2005.20 5689.0 689.59
##
## Step: AIC=602.78
## BodyFat ~ Age + Height + Ankle + logNeck + logAbdomen + newKnee +
##      logBiceps + logForearm + newWrist
##
##          Df Sum of Sq   RSS   AIC
## - Ankle      1     5.2 3690.9 601.06
## - logBiceps   1     5.5 3691.2 601.08
## - newKnee    1     8.9 3694.6 601.26
## <none>           3685.7 602.78
## - logForearm  1    68.9 3754.6 604.49
## - logNeck     1    99.6 3785.3 606.11
## - Height      1   135.1 3820.8 607.98
## - newWrist    1   160.5 3846.2 609.30
## - Age         1   192.6 3878.3 610.97
## - logAbdomen   1  3658.9 7344.6 738.68
##
## Step: AIC=601.06
## BodyFat ~ Age + Height + logNeck + logAbdomen + newKnee + logBiceps +
##      logForearm + newWrist
##
##          Df Sum of Sq   RSS   AIC
## - logBiceps   1     6.0 3696.9 599.39
## - newKnee    1     6.3 3697.2 599.40
## <none>           3690.9 601.06
## - logForearm  1    65.6 3756.4 602.58
## - logNeck     1   100.6 3791.4 604.44
## - Height      1   134.5 3825.4 606.22
## - newWrist    1   158.4 3849.3 607.46
## - Age         1   188.1 3878.9 609.00
## - logAbdomen   1  3659.4 7350.3 736.83
##
## Step: AIC=599.39
## BodyFat ~ Age + Height + logNeck + logAbdomen + newKnee + logForearm +
##      newWrist
##
##          Df Sum of Sq   RSS   AIC
## - newKnee    1     4.8 3701.6 597.64

```

```

## <none>           3696.9 599.39
## - logForearm   1     93.5 3790.4 602.38
## - logNeck      1     95.4 3792.3 602.48
## - Height       1    137.9 3834.8 604.71
## - newWrist     1    155.0 3851.9 605.60
## - Age          1    182.3 3879.1 607.01
## - logAbdomen   1   3948.8 7645.7 742.72
##
## Step: AIC=597.64
## BodyFat ~ Age + Height + logNeck + logAbdomen + logForearm +
##           newWrist
##
##             Df Sum of Sq   RSS   AIC
## <none>           3701.6 597.64
## - logForearm   1     89.6 3791.3 600.43
## - logNeck      1     92.6 3794.3 600.59
## - newWrist     1    173.2 3874.9 604.79
## - Age          1    192.1 3893.7 605.76
## - Height       1    202.4 3904.0 606.29
## - logAbdomen   1   5369.4 9071.0 774.90

*Backward elimination based on BIC

backBIC <- step(m2,direction="backward", data=new_dftrain, k=log(N))

## Start: AIC=640.96
## BodyFat ~ Age + Height + Ankle + logNeck + logAbdomen + newHip +
##           newKnee + logBiceps + logForearm + newWrist
##
##             Df Sum of Sq   RSS   AIC
## - newHip        1     1.89 3685.7 635.76
## - newKnee       1     4.87 3688.7 635.92
## - Ankle         1     5.47 3689.3 635.96
## - logBiceps     1     6.23 3690.0 636.00
## - logForearm    1    69.36 3753.2 639.39
## - logNeck       1    98.50 3782.3 640.94
## <none>           3683.8 640.96
## - Height        1   129.05 3812.9 642.55
## - Age           1   141.88 3825.7 643.22
## - newWrist      1   157.79 3841.6 644.05
## - logAbdomen    1  2005.20 5689.0 722.58
##
## Step: AIC=635.76
## BodyFat ~ Age + Height + Ankle + logNeck + logAbdomen + newKnee +
##           logBiceps + logForearm + newWrist
##
##             Df Sum of Sq   RSS   AIC
## - Ankle         1     5.2 3690.9 630.74
## - logBiceps     1     5.5 3691.2 630.76
## - newKnee       1     8.9 3694.6 630.95
## - logForearm    1    68.9 3754.6 634.17
## <none>           3685.7 635.76
## - logNeck       1    99.6 3785.3 635.80
## - Height        1   135.1 3820.8 637.66
## - newWrist      1   160.5 3846.2 638.99

```

```

## - Age      1    192.6 3878.3 640.65
## - logAbdomen 1    3658.9 7344.6 768.36
##
## Step: AIC=630.74
## BodyFat ~ Age + Height + logNeck + logAbdomen + newKnee + logBiceps +
##           logForearm + newWrist
##
##          Df Sum of Sq   RSS   AIC
## - logBiceps  1     6.0 3696.9 625.77
## - newKnee    1     6.3 3697.2 625.79
## - logForearm 1    65.6 3756.4 628.97
## <none>          3690.9 630.74
## - logNeck    1   100.6 3791.4 630.82
## - Height     1   134.5 3825.4 632.61
## - newWrist   1   158.4 3849.3 633.85
## - Age        1   188.1 3878.9 635.39
## - logAbdomen 1   3659.4 7350.3 763.22
##
## Step: AIC=625.77
## BodyFat ~ Age + Height + logNeck + logAbdomen + newKnee + logForearm +
##           newWrist
##
##          Df Sum of Sq   RSS   AIC
## - newKnee    1     4.8 3701.6 620.73
## - logForearm 1    93.5 3790.4 625.47
## - logNeck    1    95.4 3792.3 625.57
## <none>          3696.9 625.77
## - Height     1   137.9 3834.8 627.80
## - newWrist   1   155.0 3851.9 628.69
## - Age        1   182.3 3879.1 630.10
## - logAbdomen 1   3948.8 7645.7 765.80
##
## Step: AIC=620.73
## BodyFat ~ Age + Height + logNeck + logAbdomen + logForearm +
##           newWrist
##
##          Df Sum of Sq   RSS   AIC
## - logForearm 1    89.6 3791.3 620.22
## - logNeck    1    92.6 3794.3 620.37
## <none>          3701.6 620.73
## - newWrist   1   173.2 3874.9 624.58
## - Age        1   192.1 3893.7 625.55
## - Height     1   202.4 3904.0 626.08
## - logAbdomen 1   5369.4 9071.0 794.69
##
## Step: AIC=620.22
## BodyFat ~ Age + Height + logNeck + logAbdomen + newWrist
##
##          Df Sum of Sq   RSS   AIC
## - logNeck    1    48.8 3840.1 617.48
## <none>          3791.3 620.22
## - newWrist   1   106.3 3897.5 620.45
## - Age        1   128.8 3920.1 621.60
## - Height     1   191.6 3982.9 624.78

```

```

## - logAbdomen 1    5495.7 9287.0 794.10
##
## Step: AIC=617.48
## BodyFat ~ Age + Height + logAbdomen + newWrist
##
##          Df Sum of Sq      RSS      AIC
## <none>            3840.1 617.48
## - Age       1     149.8 3989.9 619.83
## - Height    1     198.9 4039.0 622.28
## - newWrist   1     240.4 4080.5 624.33
## - logAbdomen 1    7261.7 11101.8 824.50

```

\*Forward selection based on AIC

```

mint <- lm(BodyFat~1,data=new_dftrain)
forwardAIC <- step(mint,scope=list(lower=~1,
                                     upper=~Age+Height+Ankle+newKnee+logNeck
                                     +logAbdomen+newHip+logBiceps
                                     +logForearm+newWrist),
                     direction="forward", data=new_dftrain)

```

```

## Start: AIC=854.71
## BodyFat ~ 1
##
##          Df Sum of Sq      RSS      AIC
## + logAbdomen 1     9401.7 4810.8 640.06
## + newHip     1     5309.7 8902.9 763.16
## + logNeck    1     3444.8 10767.7 801.20
## + logBiceps   1     3258.8 10953.8 804.62
## + newKnee    1     2956.2 11256.3 810.07
## + Age        1     1800.9 12411.6 829.61
## + logForearm 1     1637.9 12574.6 832.22
## + newWrist   1     1636.2 12576.3 832.25
## + Ankle      1     657.8 13554.8 847.24
## <none>           14212.5 854.71
## + Height     1      55.3 14157.2 855.93
##
## Step: AIC=640.06
## BodyFat ~ logAbdomen
##
```

```

##          Df Sum of Sq      RSS      AIC
## + Height    1     634.43 4176.4 613.78
## + newWrist  1     494.36 4316.5 620.38
## + newHip    1     409.98 4400.9 624.25
## + logNeck   1     384.88 4426.0 625.39
## + newKnee   1     372.94 4437.9 625.92
## + Age       1     230.54 4580.3 632.24
## + Ankle     1     186.25 4624.6 634.17
## + logForearm 1     152.74 4658.1 635.61
## + logBiceps  1     124.01 4686.8 636.84
## <none>           4810.8 640.06
##
## Step: AIC=613.78
## BodyFat ~ logAbdomen + Height
##
```

```

##               Df Sum of Sq   RSS   AIC
## + logNeck     1  189.065 3987.4 606.51
## + newWrist    1  186.526 3989.9 606.64
## + Age         1   95.874 4080.5 611.13
## + newHip      1   87.810 4088.6 611.53
## <none>          4176.4 613.78
## + newKnee     1   36.951 4139.5 614.00
## + Ankle        1   34.647 4141.8 614.11
## + logBiceps    1   23.890 4152.5 614.63
## + logForearm   1   21.169 4155.3 614.76
##
## Step:  AIC=606.51
## BodyFat ~ logAbdomen + Height + logNeck
##
##               Df Sum of Sq   RSS   AIC
## + Age         1   89.812 3897.5 603.96
## + newWrist    1   67.233 3920.1 605.11
## + newHip      1   51.710 3935.6 605.90
## <none>          3987.4 606.51
## + newKnee     1   22.159 3965.2 607.40
## + Ankle        1   16.976 3970.4 607.66
## + logForearm   1    5.185 3982.2 608.25
## + logBiceps    1    0.265 3987.1 608.50
##
## Step:  AIC=603.96
## BodyFat ~ logAbdomen + Height + logNeck + Age
##
##               Df Sum of Sq   RSS   AIC
## + newWrist    1  106.267 3791.3 600.43
## <none>          3897.5 603.96
## + logForearm   1   22.678 3874.9 604.79
## + newKnee     1   13.317 3884.2 605.27
## + newHip      1    9.085 3888.5 605.49
## + logBiceps    1    7.915 3889.6 605.55
## + Ankle        1    6.968 3890.6 605.60
##
## Step:  AIC=600.43
## BodyFat ~ logAbdomen + Height + logNeck + Age + newWrist
##
##               Df Sum of Sq   RSS   AIC
## + logForearm   1   89.636 3701.6 597.64
## <none>          3791.3 600.43
## + logBiceps    1   30.885 3760.4 600.79
## + Ankle        1    1.189 3790.1 602.37
## + newKnee     1    0.892 3790.4 602.38
## + newHip      1    0.530 3790.7 602.40
##
## Step:  AIC=597.64
## BodyFat ~ logAbdomen + Height + logNeck + Age + newWrist + logForearm
##
##               Df Sum of Sq   RSS   AIC
## <none>          3701.6 597.64
## + newKnee     1   4.7584 3696.9 599.39
## + logBiceps   1   4.4555 3697.2 599.40

```

```

## + newHip      1    3.2198 3698.4 599.47
## + Ankle      1    3.0809 3698.6 599.48

*Forward selection based on BIC

forwardBIC <- step(mint, scope=list(lower=~1,
                                     upper=~Age+Height+Ankle+newKnee+logNeck
                                     +logAbdomen+newHip+logBiceps
                                     +logForearm+newWrist),
                     direction="forward", data=new_dftrain, k=log(N))

## Start: AIC=858.01
## BodyFat ~ 1
##
##          Df Sum of Sq     RSS     AIC
## + logAbdomen 1    9401.7 4810.8 646.66
## + newHip     1    5309.7 8902.9 769.76
## + logNeck    1    3444.8 10767.7 807.80
## + logBiceps   1    3258.8 10953.8 811.22
## + newKnee    1    2956.2 11256.3 816.67
## + Age        1    1800.9 12411.6 836.21
## + logForearm 1    1637.9 12574.6 838.82
## + newWrist   1    1636.2 12576.3 838.85
## + Ankle      1    657.8 13554.8 853.83
## <none>           14212.5 858.01
## + Height     1    55.3 14157.2 862.53
##
## Step: AIC=646.66
## BodyFat ~ logAbdomen
##
##          Df Sum of Sq     RSS     AIC
## + Height     1    634.43 4176.4 623.67
## + newWrist   1    494.36 4316.5 630.27
## + newHip     1    409.98 4400.9 634.14
## + logNeck    1    384.88 4426.0 635.28
## + newKnee    1    372.94 4437.9 635.82
## + Age        1    230.54 4580.3 642.14
## + Ankle      1    186.25 4624.6 644.06
## + logForearm 1    152.74 4658.1 645.50
## <none>           4810.8 646.66
## + logBiceps   1    124.01 4686.8 646.73
##
## Step: AIC=623.67
## BodyFat ~ logAbdomen + Height
##
##          Df Sum of Sq     RSS     AIC
## + logNeck    1    189.065 3987.4 619.71
## + newWrist   1    186.526 3989.9 619.83
## <none>           4176.4 623.67
## + Age        1    95.874 4080.5 624.33
## + newHip     1    87.810 4088.6 624.72
## + newKnee    1    36.951 4139.5 627.19
## + Ankle      1    34.647 4141.8 627.31
## + logBiceps   1    23.890 4152.5 627.82
## + logForearm 1    21.169 4155.3 627.96

```

```

##  

## Step: AIC=619.71  

## BodyFat ~ logAbdomen + Height + logNeck  

##  

##          Df Sum of Sq    RSS    AIC  

## <none>            3987.4 619.71  

## + Age      1   89.812 3897.5 620.45  

## + newWrist 1   67.233 3920.1 621.60  

## + newHip   1   51.710 3935.6 622.39  

## + newKnee  1   22.159 3965.2 623.89  

## + Ankle    1   16.976 3970.4 624.15  

## + logForearm 1   5.185 3982.2 624.74  

## + logBiceps 1   0.265 3987.1 624.99  

detach(new_dftrain)  

df_test <- test[,-c(3,6,9)]  

df_test$logNeck <- log(df_test$Neck)  

df_test$logAbdomen <- log(df_test$Abdomen)  

df_test$newHip <- 1/(df_test$Hip^2)  

df_test$newKnee <- 1/(df_test$Knee)  

df_test$logBiceps <- log(df_test$Biceps)  

df_test$logForearm <- log(df_test$Forearm)  

df_test$newWrist <- 1/(df_test$Wrist)  

new_dftest <- df_test[,-c(4,5,6,7,9,10,11)]  

m2.R2 <- lm(BodyFat~Age+Height+logNeck+logAbdomen+logForearm+newWrist, data=new_dftest)  

summary(m2.R2)  

##  

## Call:  

## lm(formula = BodyFat ~ Age + Height + logNeck + logAbdomen +  

##     logForearm + newWrist, data = new_dftest)  

##  

## Residuals:  

##      Min       1Q   Median       3Q      Max  

## -5.9293 -2.0985 -0.2405  1.1916  8.2307  

##  

## Coefficients:  

##              Estimate Std. Error t value Pr(>|t|)  

## (Intercept) -3.264e+02  7.269e+01 -4.491 5.26e-05 ***  

## Age         -1.324e-03  6.173e-02 -0.021  0.9830  

## Height      -1.107e-01  2.458e-01 -0.450  0.6548  

## logNeck     -2.097e+01  1.515e+01 -1.384  0.1735  

## logAbdomen   7.876e+01  6.800e+00 11.582 8.33e-15 ***  

## logForearm   6.174e+00  6.063e+00  1.018  0.3142  

## newWrist    9.805e+02  3.753e+02  2.612  0.0123 *  

## ---  

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

##  

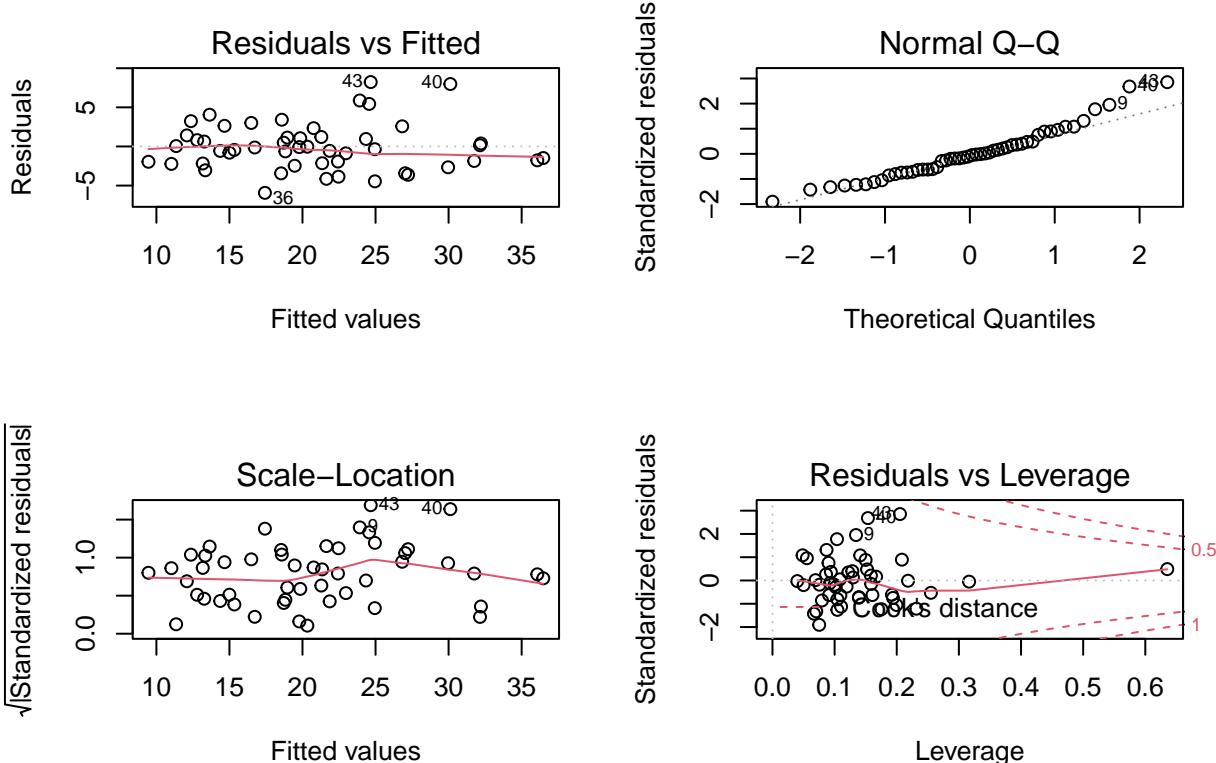
## Residual standard error: 3.235 on 43 degrees of freedom  

## Multiple R-squared:  0.8305, Adjusted R-squared:  0.8068  

## F-statistic: 35.1 on 6 and 43 DF,  p-value: 4.991e-15

```

```
par(mfrow = c(2, 2))
plot(m2.R2)
```



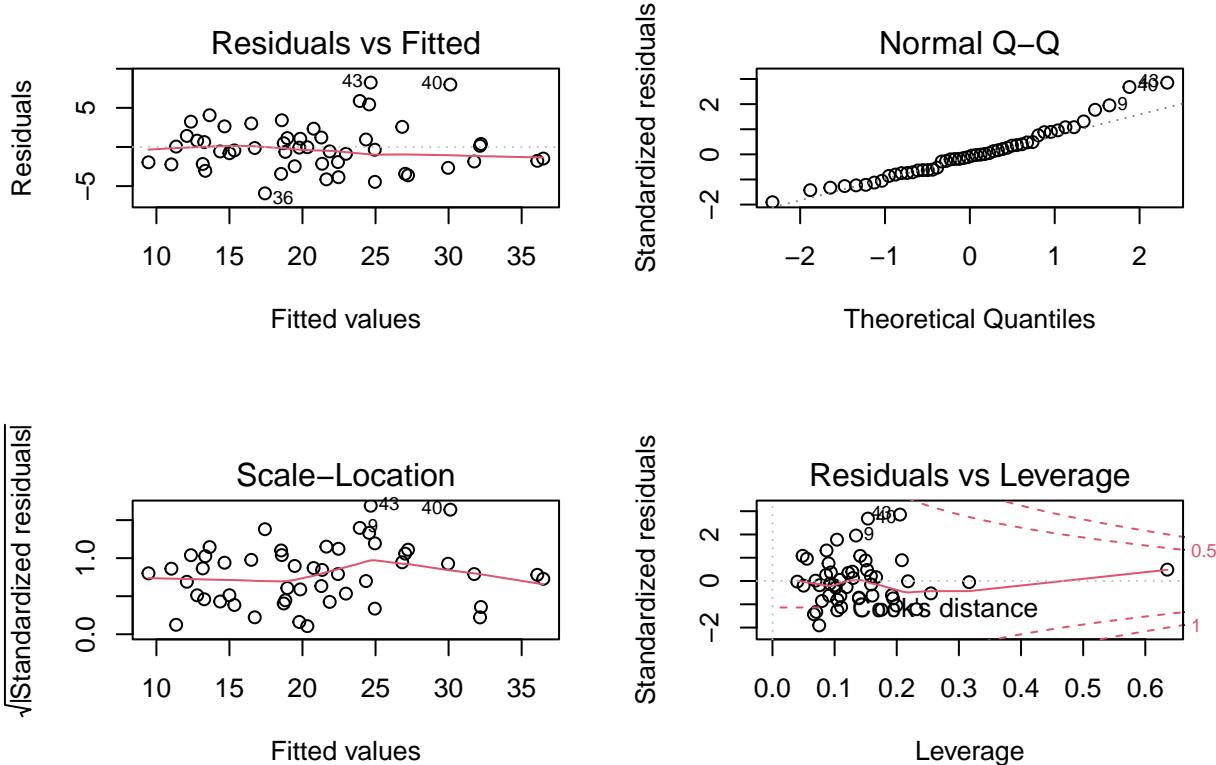
```
m2.AIC.6 <- lm(BodyFat ~ logForearm + logNeck + newWrist + Age + Height + logAbdomen, data=new_dftest)
summary(m2.AIC.6)
```

```
##
## Call:
## lm(formula = BodyFat ~ logForearm + logNeck + newWrist + Age +
##     Height + logAbdomen, data = new_dftest)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -5.9293 -2.0985 -0.2405  1.1916  8.2307 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -3.264e+02  7.269e+01 -4.491 5.26e-05 ***
## logForearm   6.174e+00  6.063e+00  1.018  0.3142    
## logNeck     -2.097e+01  1.515e+01 -1.384  0.1735    
## newWrist    9.805e+02  3.753e+02  2.612  0.0123 *  
## Age        -1.324e-03  6.173e-02 -0.021  0.9830    
## Height     -1.107e-01  2.458e-01 -0.450  0.6548    
## logAbdomen  7.876e+01  6.800e+00 11.582 8.33e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

## 
## Residual standard error: 3.235 on 43 degrees of freedom
## Multiple R-squared:  0.8305, Adjusted R-squared:  0.8068
## F-statistic: 35.1 on 6 and 43 DF,  p-value: 4.991e-15
par(mfrow = c(2, 2))
plot(m2.AIC.6)

```



```

m2.BIC.4 <- lm(BodyFat~Age + Height + logAbdomen + newWrist,data=new_dftest)
summary(m2.BIC.4)

```

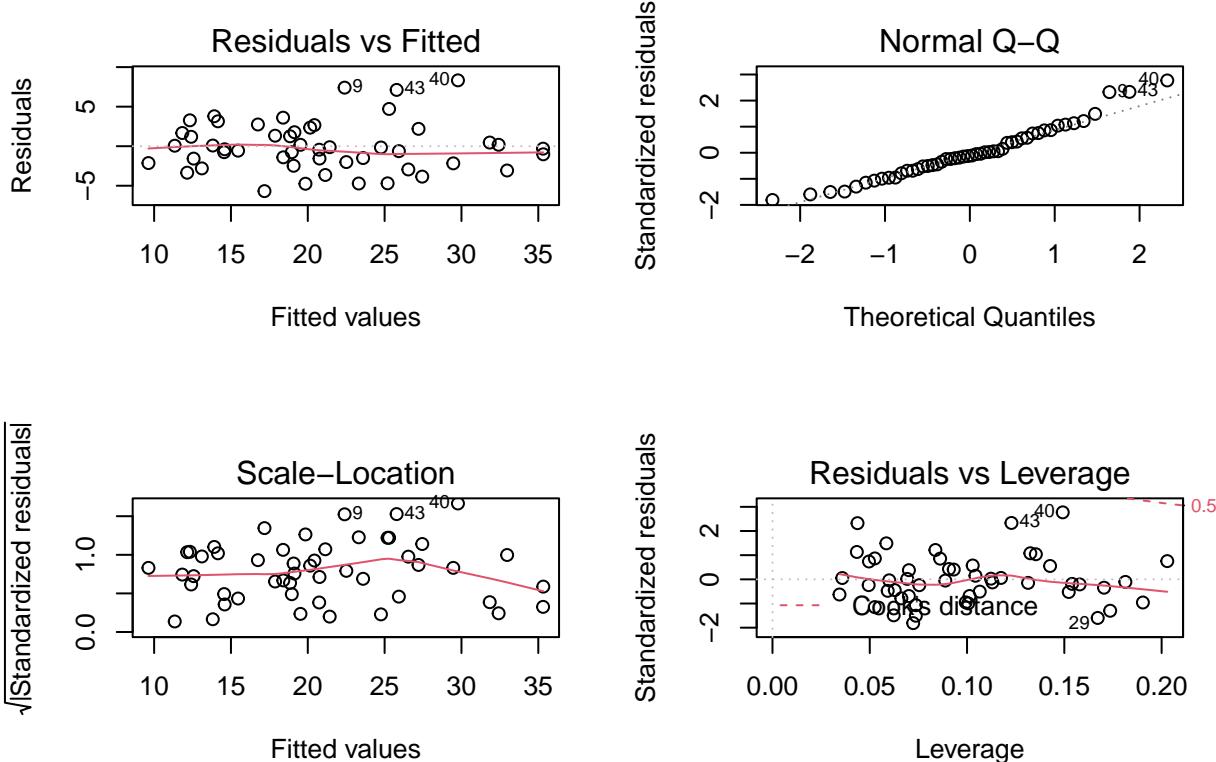
```

## 
## Call:
## lm(formula = BodyFat ~ Age + Height + logAbdomen + newWrist,
##      data = new_dftest)
## 
## Residuals:
##      Min    1Q   Median    3Q    Max 
## -5.6813 -2.1005 -0.3509  1.7435  8.3262 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -391.04008   47.65841 -8.205 1.74e-10 ***
## Age          0.02246    0.05816   0.386   0.701    
## Height       -0.06708   0.24500  -0.274   0.786    
## logAbdomen    76.00060   5.71954  13.288 < 2e-16 ***
## newWrist     1297.77229  295.81639   4.387 6.85e-05 ***
## 
```

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.256 on 45 degrees of freedom
## Multiple R-squared: 0.8202, Adjusted R-squared: 0.8043
## F-statistic: 51.34 on 4 and 45 DF, p-value: 3.305e-16
par(mfrow = c(2, 2))
plot(m2.BIC.4)

```

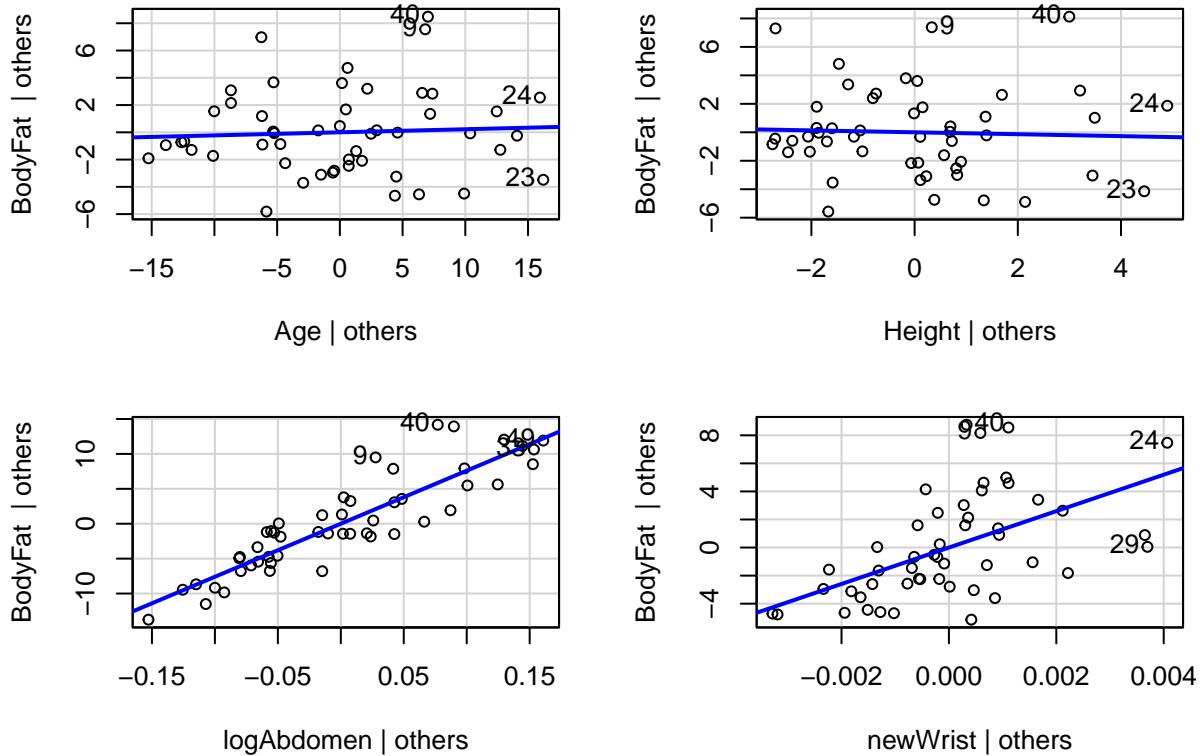


```

library(car)
avPlots(m2.BIC.4)

```

## Added-Variable Plots



Final model

```
m2.BIC.4.final <- lm(BodyFat ~ logAbdomen + newWrist, data=new_dftest)
summary(m2.BIC.4.final)
```

```
##
## Call:
## lm(formula = BodyFat ~ logAbdomen + newWrist, data = new_dftest)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -5.6236 -2.1651 -0.1035  1.7830  7.9769 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -392.629    35.533 -11.050 1.16e-14 ***
## logAbdomen    75.541     5.574  13.553 < 2e-16 ***
## newWrist     1296.969   240.921   5.383 2.28e-06 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 3.214 on 47 degrees of freedom
## Multiple R-squared:  0.8171, Adjusted R-squared:  0.8093 
## F-statistic: 105 on 2 and 47 DF,  p-value: < 2.2e-16
par(mfrow = c(2, 2))
plot(m2.BIC.4.final)
```

