Jared Castaneda
Khoa Do
Kurt Prutsman
CPSC 375

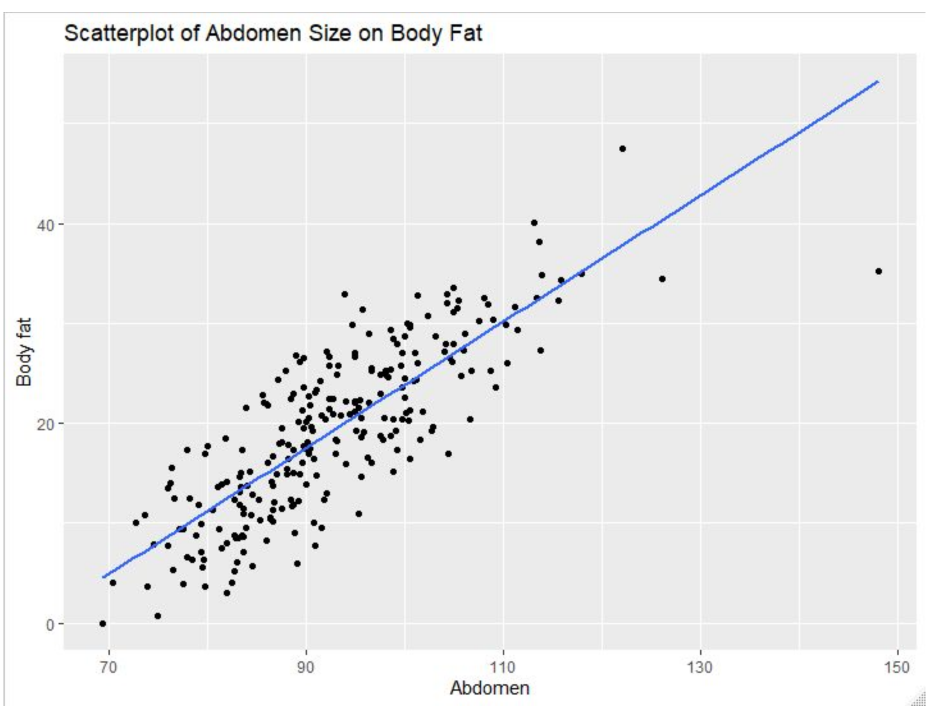
Project 1

We began the project by researching methods to determine body fat other than body density. One method we found[1] that is also accurate is measuring skinfolds and from there putting them into a special equation. For men, the skinfold areas are the abdomen, chest, and thighs. While we didn't have the skinfold measurements, we did have information on the abdomen, chest and thigh. It seemed reasonable to begin the project primarily looking at those variables modeled to body fat.

Next, we plotted all the variables though linear modeling hoping to find useful variables other than abdomen, chest and thigh measurement that might not be as obvious. This helped us visualize the useful variables. We also cleaned the data by omitting all the 'NA'.

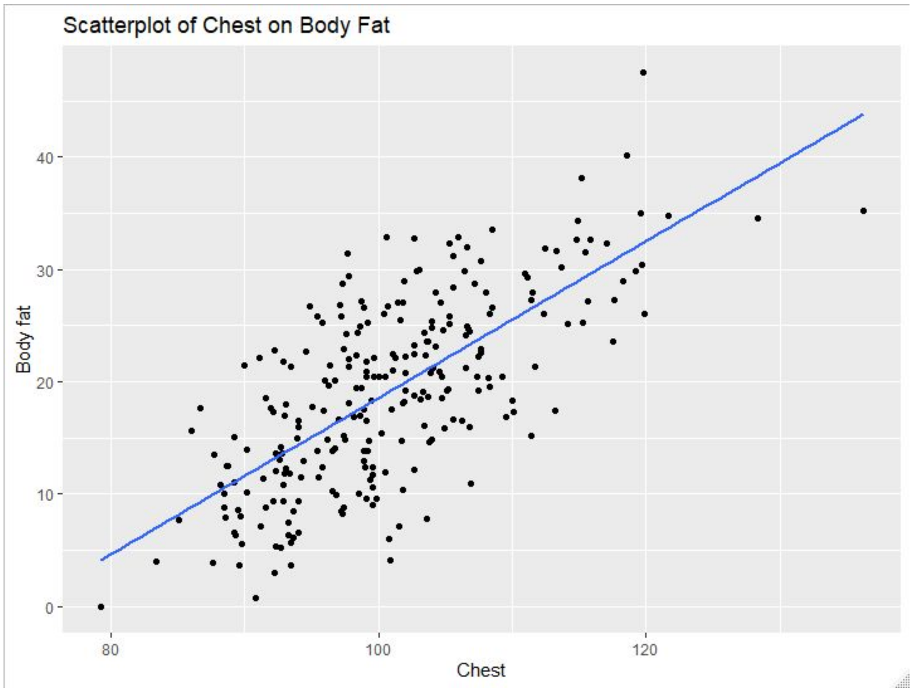
```
newdata <- na.omit(data)
```

```
ggplot(data, aes(x=Abdomen, y=bodyfat)) + geom_point() + labs(x = "Abdomen", y = "Body fat", title="Scatterplot of Abdomen Size on Body Fat") + geom_smooth(method = 'lm', se = FALSE)
```

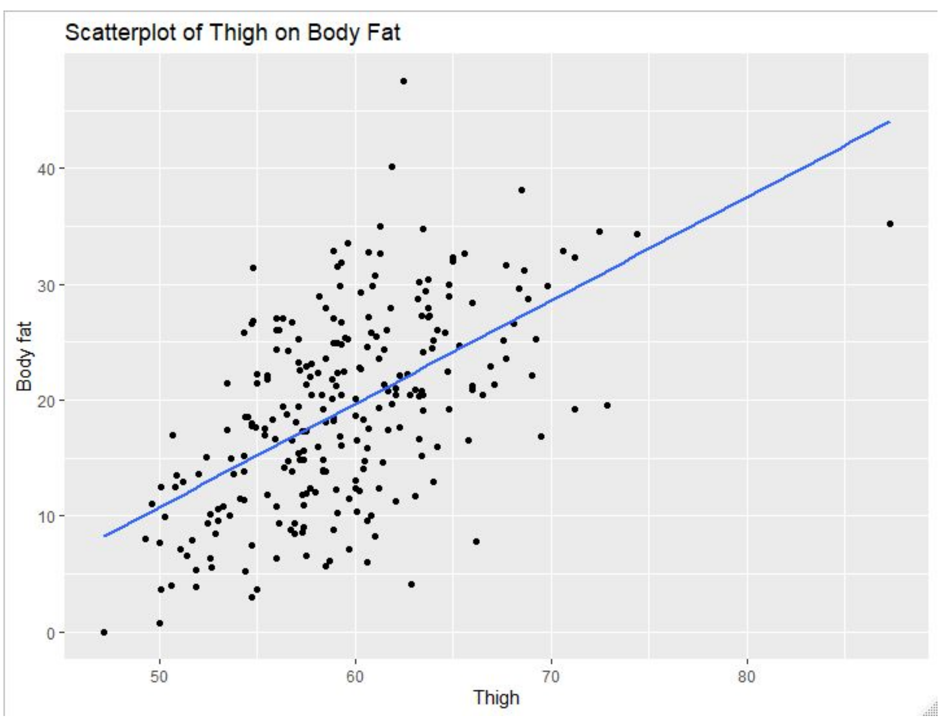


[1] Brusco, J. (2016, February 25). Equations for skinfold measurements to determine body fat percentage. Retrieved from https://www.livestrong.com

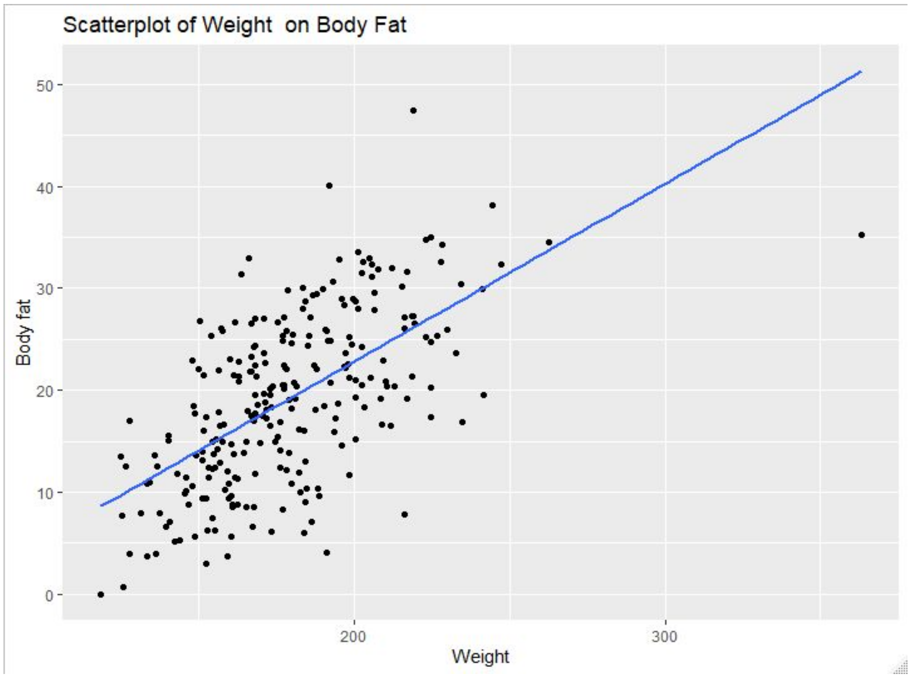
ggplot(data, aes(x=Chest, y=bodyfat)) + geom_point() + labs(x = "Chest", y = "Body fat", title="Scatterplot of Chest on Body Fat") + geom_smooth(method = 'lm', se = FALSE)



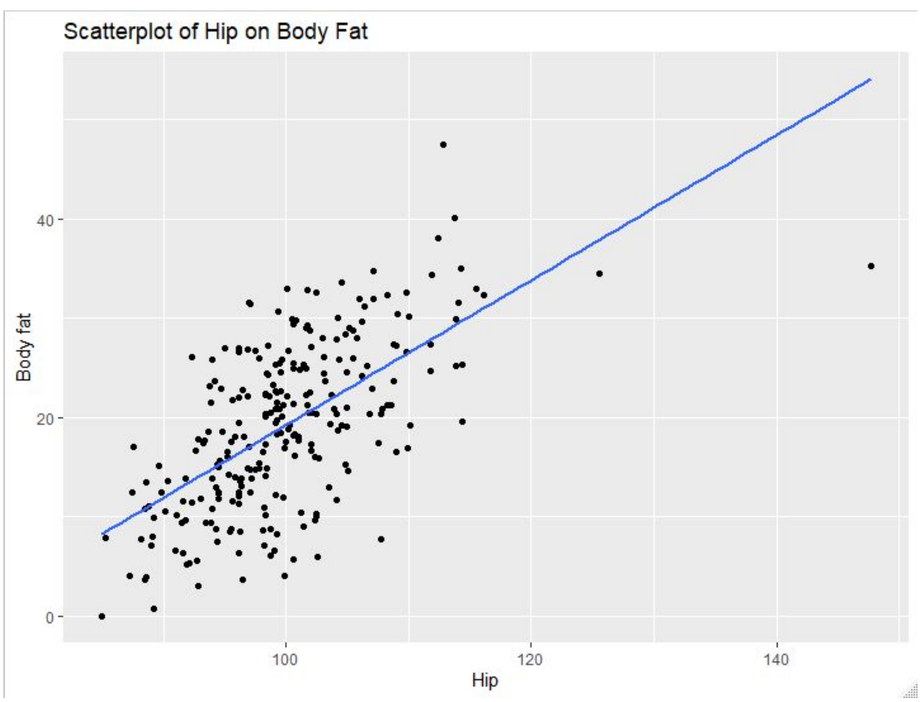
ggplot(data, aes(x=Thigh, y=bodyfat)) + geom_point() + labs(x = "Thigh Size", y = "Body fat", title="Scatterplot of Thigh Size on Body Fat") + geom_smooth(method = 'lm', se = FALSE)



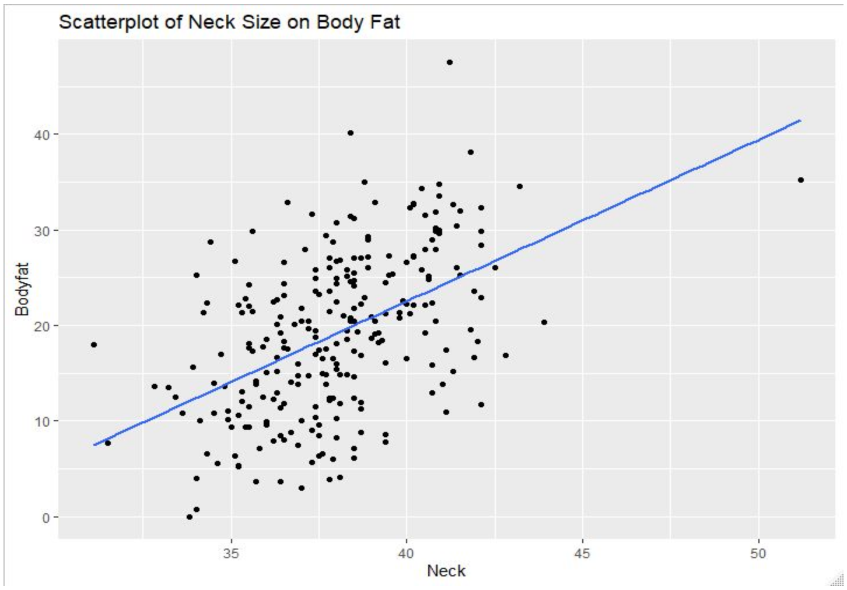
```
ggplot(data, aes(x=Weight, y=bodyfat)) + geom_point() + labs(x = "Weight", y = "Body fat",
title="Scatterplot of Weight  on Body Fat") + geom_smooth(method = 'lm', se = FALSE)
```




```
ggplot(data, aes(x=Hip, y=bodyfat)) + geom_point() + labs(x = "Hip", y = "Body fat",
title="Scatterplot of Hip on Body Fat") + geom_smooth(method = 'lm', se = FALSE)
```

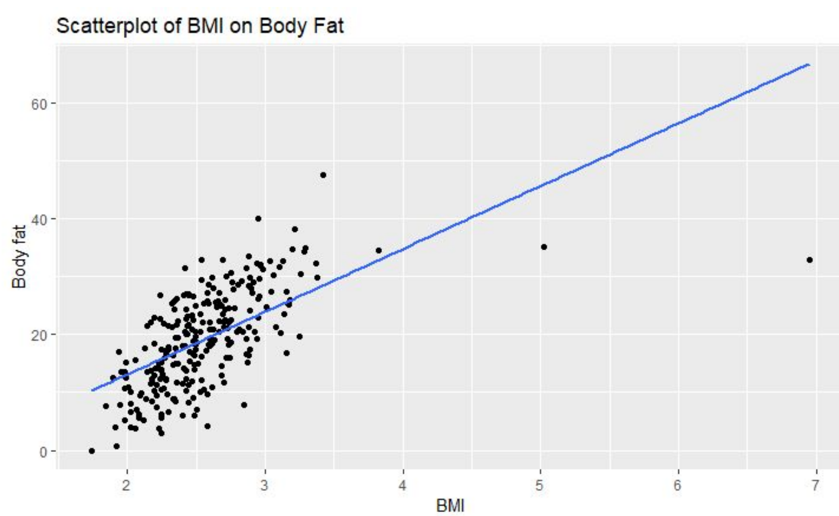


```
ggplot(data, aes(x=Neck, y=bodyfat)) + geom_point() + labs(x = "Neck", y = "Bodyfat", title="Scatterplot of Neck Size on Body Fat") + geom_smooth(method = 'lm', se = FALSE)
```



After plotting every variable in relation to body fat[2], we figured we could try creating new variables. Since we had information on the height and weight, a new BMI variable seemed to make sense.

```
data <- data %>% mutate(BMI=Weight/Height)
> ggplot(data, aes(x=BMI, y=bodyfat)) + geom_point() + labs(x = "BMI", y = "Body fat", title="Scatterplot of BMI on Body Fat") + geom_smooth(method = 'lm', se = FALSE)
```



[2] For the sake of being concise, not all linear models of all the variables are included in this report. However, they were completed and available upon request.

Once we were done with basic exploratory data analysis, we began trying different combinations of variables for modeling body fat. Going back the skinfold method of measuring body fat, we started off with those key areas of the body(chest, abdomen and thigh). Our evaluation metric was the Adjusted R-Squared.

Chest

```
> newdataChest <- lm(data=newdata, bodyfat~Chest)
> summary(newdataChest)
Multiple R-squared:  0.4937,  Adjusted R-squared:  0.4917
```

Abdomen

```
> newdataAbdomen <- lm(data=newdata, bodyfat~Abdomen)
> summary(newdataAbdomen)
Multiple R-squared:  0.6617,  Adjusted R-squared:  0.6603
```

Thigh

```
> newdataThigh <- lm(data=newdata, bodyfat~Thigh)
> summary(newdataThigh)
Multiple R-squared:  0.3132,  Adjusted R-squared:  0.3104
```

Out of the three variables, abdomen had the highest Adjust R-squared score. That made us believe that would be one of the more important variables to determine body fat. We then began different combinations of those three key variables.

Chest and Thigh

```
> newdataThighandChest <- lm(data=newdata, bodyfat~Thigh+Chest)
> summary(newdataThighandChest)
Multiple R-squared:  0.4984,  Adjusted R-squared:  0.4943
```

Thigh and Abdomen

```
> newdataThighandAbdomen <- lm(data=newdata, bodyfat~Thigh+Abdomen)
> summary(newdataThighandAbdomen)
Multiple R-squared:  0.6716,  Adjusted R-squared:  0.669
```

Chest And Abdomen

```
> newdataChestandAbdomen <- lm(data=newdata, bodyfat~Chest+Abdomen)
> summary(newdataChestandAbdomen)
Multiple R-squared:  0.6728,  Adjusted R-squared:  0.6702
```

The chest and abdomen combination gave us our highest Adjusted R-squared value. Next we began looking at models of all three preferred variables along with logarithm and square transformations.

Chest And Abdomen And Thigh

```
> newdataChestandAbdomenandThigh <- lm(data=newdata, bodyfat~Chest+Abdomen+Thigh)
> summary(newdataChestandAbdomenandThigh)
Multiple R-squared:  0.6807,  Adjusted R-squared:  0.6768
```

Chest And Abdomen And Thigh Log

```
> newdataChestandAbdomenandThighLOG<- lm(data=newdata,
bodyfat~log(Chest)+log(Abdomen)+log(Thigh))
> summary(newdataChestandAbdomenandThighLOG)
Multiple R-squared:  0.6936,  Adjusted R-squared:  0.6899
```

Chest And Abdomen And Thigh Square

```
> newdataChestandAbdomenandThighSQUARE <- lm(data=newdata,
bodyfat~Chest^2+Abdomen^2+Thigh^2)
> summary(newdataChestandAbdomenandThighSQUARE)
Multiple R-squared:  0.6807,  Adjusted R-squared:  0.6768
```

Next, we went back to a combination of two preferred variables with log and square transformations.

Abdomen And Thigh Log

```
> newdataAbdomenandThighLOG <- lm(data=newdata, bodyfat~log(Abdomen)+log(Thigh))
> summary(newdataAbdomenandThighLOG)
Multiple R-squared:  0.6841,  Adjusted R-squared:  0.6816
```

Abdomen And Thigh Square

```
> newdataAbdomenandThighSQUARE <- lm(data=newdata, bodyfat~Abdomen^2+Thigh^2)
> summary(newdataAbdomenandThighSQUARE)
Multiple R-squared:  0.6716,  Adjusted R-squared:  0.669
```

Our next step was introducing a new variable and continue transforming it since we weren't improving the Adjusted R-squared score much with only combinations of our 3 preferred variables. Taking a cue from Columbian pop artist Shakira and her famous song "Hips Don't Lie", we explored if our Adjusted R-squared score would improve with the Hip variable.

Hip And Thigh

```
> newdataHipandThigh <- lm(data=newdata, bodyfat~Hip+Thigh)
> summary(newdataHipandThigh)
Multiple R-squared:  0.3909,  Adjusted R-squared:  0.386
```

Hip And Thigh Log

```
> newdataHipandThighLOG <- lm(data=newdata, bodyfat~log(Hip)+log(Thigh))
> summary(newdataHipandThighLOG)
Multiple R-squared:  0.4052,  Adjusted R-squared:  0.4004
```

We quickly abandoned and Hip variable since it was lowering our Adjusted R-squared significantly. Next we explored how introducing the Weight variable might change improve our evaluation metric. We looked at combining it with Abdomen since originally it gave us our highest Adjusted R-squared score.

Abdomen And Weight

```
> newdataAbdomenandWeight <- lm(data=newdata, bodyfat~Abdomen+Weight)
> summary(newdataAbdomenandWeight)
Multiple R-squared:  0.7188,  Adjusted R-squared:  0.7165
```

Abdomen And Weight Log

```
> newdataAbdomenandWeightLog <- lm(data=newdata, bodyfat~log(Abdomen)+log(Weight))
> summary(newdataAbdomenandWeightLog)
Multiple R-squared:   0.72,     Adjusted R-squared:  0.7177
```

At this point we determined Weight would also likely be a key variable. Our previous best combination was Chest and Abdomen and Thigh Log and that gave us an Adjusted R-square score of 0.6899. It seemed logical to combine Weight with that combination for a new high score.

Chest And Abdomen And Thigh And Weight Log

```
> newdataChestandAbdomenandThighandWeightLOG <- lm(data=newdata,
bodyfat~log(Chest)+log(Abdomen)+log(Thigh) + log(Weight))
> summary(newdataChestandAbdomenandThighandWeightLOG)
Multiple R-squared:  0.7229,  Adjusted R-squared:  0.7184
```

At this point we reached our highest Adjusted R-square score of 0.7184. We wanted to add yet another variable. We decided to create a new variable in BMI with Height/Weight and compare it to other variables we saw success with.

BMI

```
> newdataBMI <- lm(data=newdata, bodyfat~BMI)
> summary(newdataBMI)
```

Multiple R-squared: 0.4915, **Adjusted R-squared: 0.4894**

BMI Log

```
> newdataBMILOG <- lm(data=newdata, bodyfat~log(BMI))
> summary(newdataBMILOG)
```

Multiple R-squared: 0.4603, **Adjusted R-squared: 0.4581**

BMI And Abdomen

```
> newdataBMIandAbdomen <- lm(data=newdata, bodyfat~BMI+Abdomen)
> summary(newdataBMIandAbdomen)
```

Multiple R-squared: 0.6647, **Adjusted R-squared: 0.662**

BMI And Abdomen Log

```
> newdataBMIandAbdomenLOG <- lm(data=newdata, bodyfat~log(BMI)+log(Abdomen))
> summary(newdataBMIandAbdomenLOG)
```

Multiple R-squared: 0.6813, **Adjusted R-squared: 0.6787**

BMI And Chest

```
> newdataBMIandChest <- lm(data=newdata, bodyfat~BMI+Chest)
> summary(newdataBMIandChest)
```

Multiple R-squared: 0.5201, **Adjusted R-squared: 0.5163**

BMI And Thigh

```
> newdataBMIandThigh <- lm(data=newdata, bodyfat~BMI+Thigh)
> summary(newdataBMIandThigh)
```

Multiple R-squared: 0.4977, **Adjusted R-squared: 0.4936**

At this point we realized BMI was not yielding any high scores so we stopped exploring it. We moved on with exploring what kind of improvements to the Adjusted R-squared score we would get with the variable Height, along with other variables we saw success with.

Height

```
> newdataHeight <- lm(data=newdata, bodyfat~Height)
> summary(newdataHeight)
```

Multiple R-squared: 0.008009, **Adjusted R-squared: 0.004041**

Height And Abdomen

```
> newdataHeightandAbdomen <- lm(data=newdata, bodyfat~Height+Abdomen)
> summary(newdataHeightandAbdomen)
```

Multiple R-squared: 0.6878, **Adjusted R-squared: 0.6853**

We found it odd that Height alone was not a very accurate predictor of Body Fat (which intuitively makes sense) and yet when combined with other previous key variables made such a drastic improvement. Since it had such a phenomenal property, we decided to keep combining Height. We also found it odd that BMI (a formula of weight to height) was not an accurate predictor of Body Fat while Height and Weight individually were.

Our previous best combination was Chest And Abdomen And Thigh And Weight Log. The score was 0.7184. We decided to add Height to it.

Height And Chest And Abdomen And Thigh And Weight Log

```
> newdataHeightChestandAbdomenandThighandWeightLOG <- lm(data=newdata,
bodyfat~log(Chest)+log(Abdomen)+log(Thigh) + log(Weight) + log(Height))
> summary(newdataHeightChestandAbdomenandThighandWeightLOG)
```

Multiple R-squared: 0.7258, **Adjusted R-squared: 0.7214**

Height And Chest And Abdomen And Thigh And Weight

```
> newdataHeightChestandAbdomenandThighandWeight <- lm(data=newdata, bodyfat~Chest +
Abdomen+Thigh + Weight + Height)
> summary(newdataHeightChestandAbdomenandThighandWeight)
```

Multiple R-squared: 0.7244, **Adjusted R-squared: 0.7188**

0.7214 was a new high Adjusted R-squared score so we kept the Height and Weight variables and moved on to exploring what kind of improvements could be made by including Age. Intuitively, we didn't think Age would matter that much. However, compared to other variables such as Wrist or Ankle, Age seemed like a more likely predictor. We combined Age with our previous highest Adjusted R-Squared score.

Height And Chest And Abdomen And Thigh And Weight And Age Log

```
> newdataHeightandAbdomenandThighandWeightandAgeLOG <- lm(data=newdata,
bodyfat~log(Height)+log(Abdomen)+log(Thigh) + log(Weight) + log(Age))
> summary(newdataHeightandAbdomenandThighandWeightandAgeLOG)
```

Multiple R-squared: 0.7262, **Adjusted R-squared: 0.7206**

The Age variable didn’t make much difference but we decided not to remove it. At this point we were thinking larger body parts are more indicative of body fat, so we re-introduced the Hips variable. We combined it with our previous best combination.

Height And Chest And Abdomen And Thigh And Weight And Age And Hip Log

```
> newdataHeightandAbdomenandThighandWeightandAgeandHipsLOG <- lm(data=newdata, bodyfat~log(Height)+log(Abdomen)+log(Thigh) + log(Weight) + log(Age) + log(Hip))
> summary(newdataHeightandAbdomenandThighandWeightandAgeandHipsLOG)
Multiple R-squared:  0.7277,  Adjusted R-squared:  0.721
```

Height And Chest And Abdomen And Thigh And Weight And Age Log + Hip

```
> newdataHeightandAbdomenandThighandWeightandAgeandHipsLOG1 <- lm(data=newdata, bodyfat~log(Height)+log(Abdomen)+log(Thigh) + log(Weight) + log(Age) + Hip)
> summary(newdataHeightandAbdomenandThighandWeightandAgeandHipsLOG1)
Multiple R-squared:  0.7298,  Adjusted R-squared:  0.7232
```

There was a slightly significant bump (0.0018) in Adjusted R Squared score by adding Hip (without the logarithm). At this point we reached a new high score of 0.7232.   We decided to try adding Age without a logarithm to see what would happen.

Height And Chest And Abdomen And Thigh And Weight Log + Age and Hip

```
> newdataHeightandAbdomenandThighandWeightandAgeandHipsLOG2 <- lm(data=newdata, bodyfat~log(Height)+log(Abdomen)+log(Thigh) + log(Weight) + Age + Hip)
> summary(newdataHeightandAbdomenandThighandWeightandAgeandHipsLOG2)
Multiple R-squared:  0.7298,  Adjusted R-squared:  0.7231
```

We decided to settle on this function since it gave us the best Adjusted R-squared score of 0.7232.

```
temp <- lm(data=newdata, bodyfat~log(Height)+log(Abdomen)+log(Thigh) + log(Weight) +
log(Age) + Hip)
> summary(temp)
Multiple R-squared:  0.7298,  Adjusted R-squared:  0.7232
```

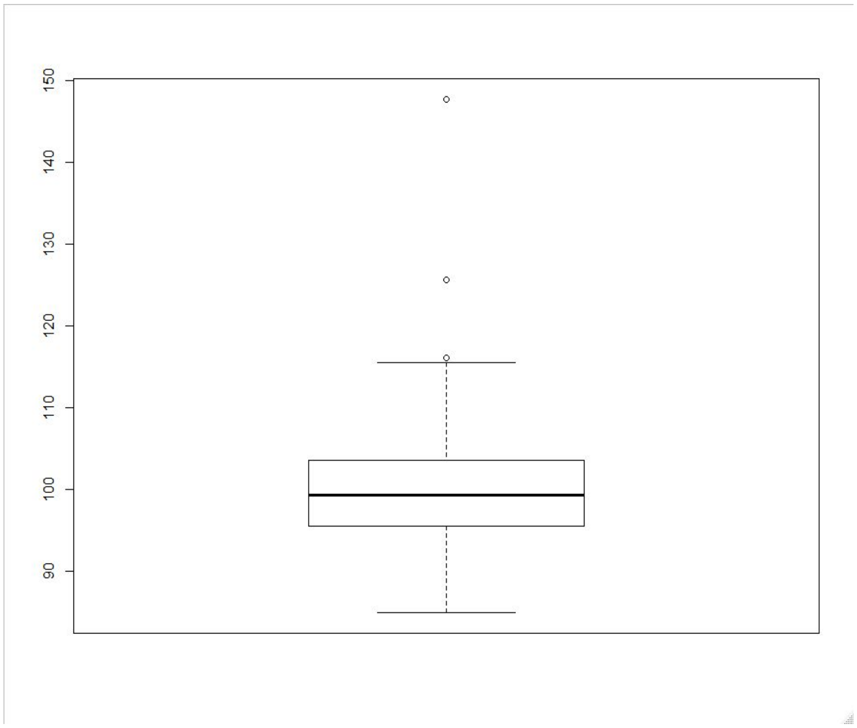
```
Call:
lm(formula = bodyfat ~ log(Height) + log(Abdomen) + log(Thigh) +
   log(Weight) + log(Age) + Hip, data = newdata)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2752  -3.1141  -0.1787   2.8749  10.1915

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -276.9745    33.0531  -8.380 4.17e-15 ***
log(Height)    -9.3382     4.8817  -1.913  0.05693 .
log(Abdomen)   87.8393     6.7465  13.020  < 2e-16 ***
log(Thigh)     14.3104     7.9009   1.811  0.07133 .
log(Weight)   -19.2166     6.3019  -3.049  0.00255 **
log(Age)        0.4452     1.2406   0.359  0.72003
Hip            -0.2201     0.1217  -1.809  0.07175 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.403 on 245 degrees of freedom
Multiple R-squared:  0.7298,  Adjusted R-squared:  0.7232
F-statistic: 110.3 on 6 and 245 DF,  p-value: < 2.2e-16
```

Now we started removing outliers. We used boxplot() to get easier access to quartiles, maximums and minimum whiskers. I found the easiest way to do this is by creating boxplots for every variable. The boxplot() that you make has access to $stats and $out. $stats showed you the minimum, quartile 1, median, quartile 3, and the maximum. Anything outside the minimum or maximum is an outlier, or what's inside $out.

What I did after was store the maximum ($stats[5]) into its respective maximum value and a minimum if the boxplot showed anything below the minimum ($stats[1]). These values will come in handy later when it comes to filtering. Anything outside this range is an outlier.

Hip:

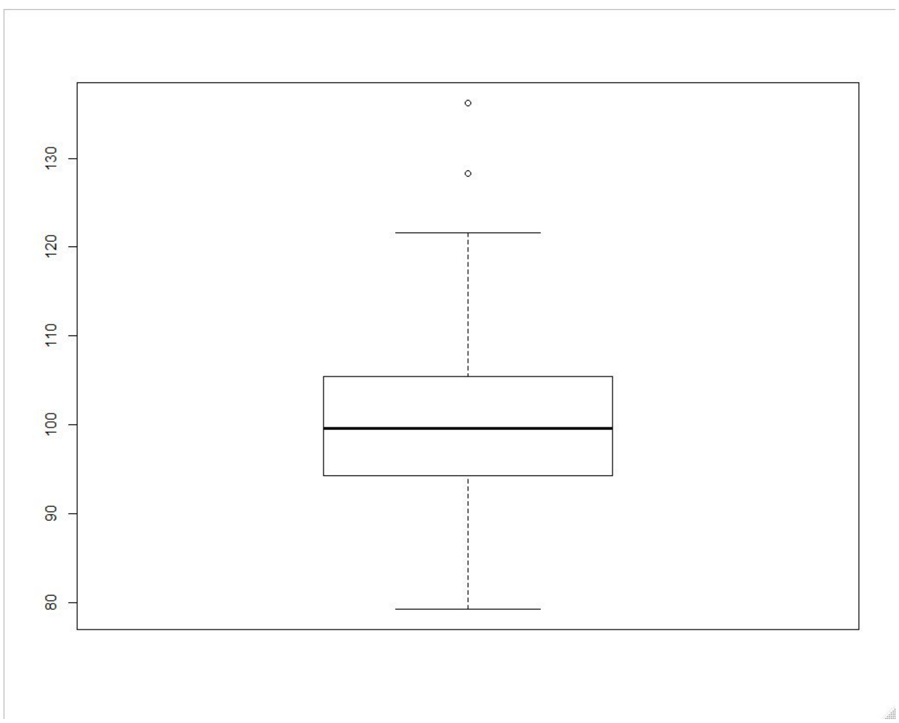
```
> bpHip <- boxplot(data$Hip)
> boxplot.stats(data$Hip)
$stats
[1]  85.00  95.50  99.30 103.55 115.50

$n
[1] 252

$conf
[1]  98.49878 100.10122

$out
[1] 116.1 147.7 125.6


> bpHip$stats #Vector with important points
       [,1]
[1,]  85.00 #Lower whisker
[2,]  95.50 #lower hinge
[3,]  99.30 #median
[4,] 103.55 #upper hinge
[5,] 115.50 #upper whisker

#Get value of largest whisker. Anything above this is an outlier.
> bpHipMax <- bpHip$stats[5]
```



Chest

> bpChest <- boxplot(data$Chest)

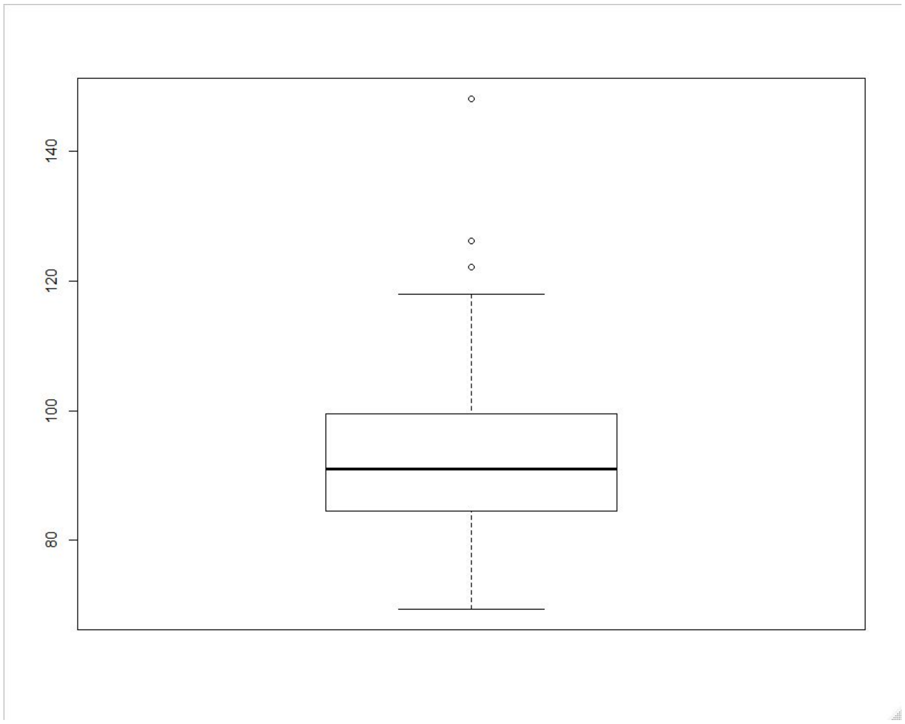
> bpChestMax <- bpChest$stats[5]



Abdomen

> bpAbdomen <- boxplot(data$Abdomen)

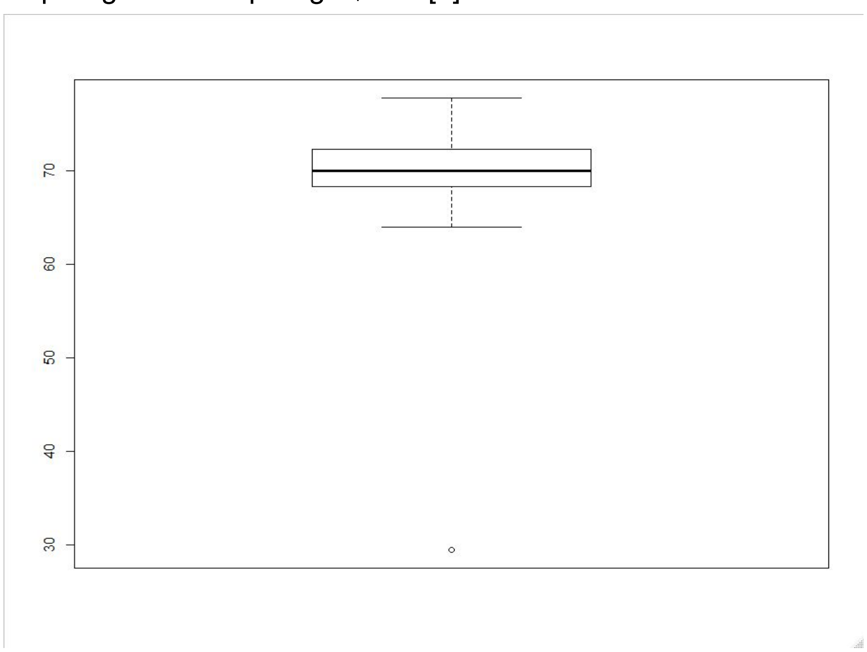
> bpAbdomenMax <- bpAbdomen$stats[5]



Height (Notice only one below the minimum)

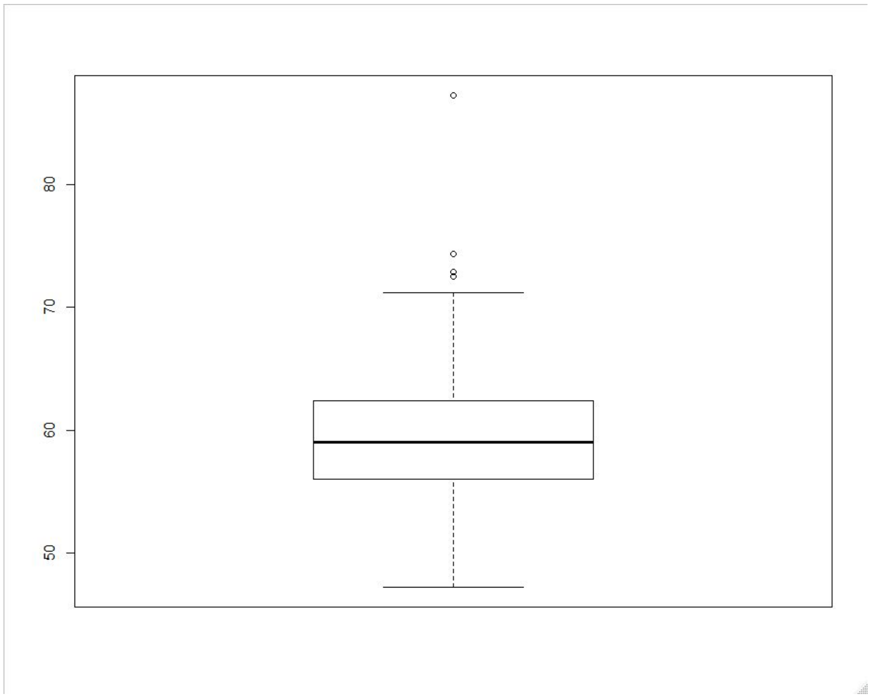
> bpHeight <- boxplot(data$Height)

> bpHeightMin <- bpHeight$stats[1]



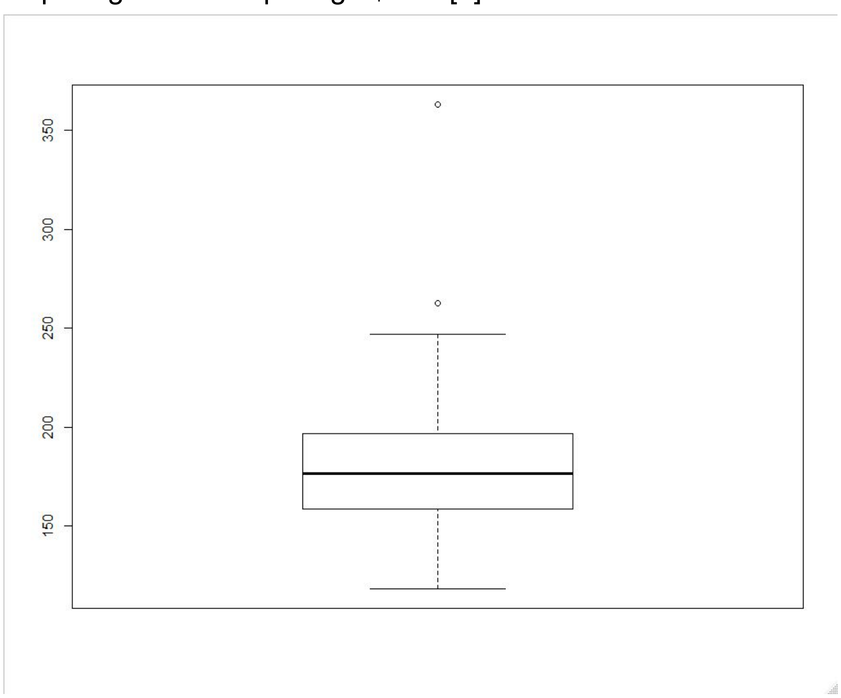
Thigh

```
> bpThigh <- boxplot(data$Thigh)
> bpThighMax <- bpThigh$stats[5]
```



Weight

```
> bpWeight <- boxplot(data$Weight)
> bpWeightMax <- bpWeight$stats[5]
```



And so on.

```
> bpNeck <- boxplot(data$Neck)
> bpNeckMax <- bpNeck$stats[5]
> bpNeckMin <- bpNeck$stats[1]
> bpKnee <- boxplot(data$Knee)
> bpKneeMax <- bpKnee$stats[5]
> bpAnkle <- boxplot(data$Ankle)
> bpAnkleMax <- bpAnkle$stats[5]
> bpBiceps <- boxplot(data$Biceps)
> bpBicepsMax <- bpBiceps$stats[5]
> bpForearm <- boxplot(data$Forearm)
> bpForearmMax <- bpForearm$stats[5]
> bpForearmMin <- bpForearm$stats[1]
```

Once we had all those values, it was time to filter the data.

```
filterdata <- data %>% filter(Hip <= bpHipMax & Chest <= bpChestMax & Abdomen <= bpAbdomenMax & Height > bpHeightMin & Thigh <= bpThighMax & Weight <= bpWeightMax & Ankle <= bpAnkleMax & Biceps <= bpBicepsMax & Forearm >= bpForearmMin & Forearm <= bpForearmMax & Neck >= bpNeckMin & Neck <= bpNeckMax & Knee <= bpKneeMax & Wrist <= bpWristMax & Wrist => bpWristMin)
```

This new data has 233 objects compared to the original 252.

filterdata	233 obs. of 15 variables

We can now try some of our other linear models again.

```
> temp <- lm(data=filterdata, bodyfat~log(Height)+log(Abdomen)+log(Thigh) + log(Weight) + log(Age) + Hip)
> summary(temp)=
```

Multiple R-squared: 0.6992, Adjusted R-squared: 0.6913

One problem we encountered is why our adjusted r-squared for this command is lower than the unfiltered data set.

```
> temp2 <- lm(data=filterdata, bodyfat~log(Height)+log(Abdomen)+log(Thigh) + log(Weight) + Age + Hip)
> summary(temp2)
```

Multiple R-squared: 0.6991, Adjusted R-squared: 0.6911

Exploratory Data Analysis: 3rd Approach

For the third approach, we decided to do get R to do the hard work of finding the best combination of variables and functions. This is the high level overview with the details in the R code:

- Get the attribute that we want to predict (bodyfat)
- Get the list of all the possible independent variables, such as weight, abdomen, and thigh measurement
- Get the list of possible math functions, such as log, tan, squared
- **Generate all possible combinations of these variables and applied functions up to user defined variable K**
- Use the generated combo list and generate lm models and compile them into summaries
- scan this explore list and selected the equations with the best Adjusted R Squared value
- Generate the equation for the best set

How to Run R Code

proj1.R can be run on the command line or in R Studio

Here are some variables and it's effect on execution:

```
# determines the minimum number of combinations of variables generated
minK = 5

# determines the maximum number of combinations generated
k <- 5

# list of math functions that can be applied to a variable as part of a combination
funcNameList <- c("log", "exp", "sqrt", "log10", "floor", "ceiling", "sin", "cos", "tan")
postscriptOpNameList <- c("^2", "^3")

# the best of 3 equations are produced
maxItemCount <- 3

# set this flag to generate the combinations
# set it to FALSE if you already have the combinations generated to save time
# the combinations are stored in files
doGenerateCombo <- TRUE

# maximum count of combinations stored in a file
chunkSize <- 3000

# the directory where the combo files are stored
workDir <- "C:/temp/cpsc375proj1/work/"
```

```
# the filespec of the combo files
filename <- "comboList"
ext <- ".csv"
```

Execution of R Code

We ran through Paul's script. We tested with k = 2 and k = 3. In this case, k represents the number of variables being tested.

When k = 2:

Results from best fit list:

```
[1] "max ~ bodyfat Weight 0.37255111373424"
[1] "max ~ bodyfat Neck 0.237643088129366"
[1] "max ~ bodyfat Chest 0.491650041989838"
[1] "max ~ bodyfat Abdomen 0.660318770251814"
[1] "max ~ bodyfat Hip 0.388439692002688"
[1] "max ~ bodyfat Thigh 0.310413232223974"
[1] "max ~ bodyfat log(Weight) 0.392820011398408"
[1] "max ~ bodyfat log(Chest) 0.495209083100418"
[1] "max ~ bodyfat log(Abdomen) 0.674057641566212"
[1] "max ~ bodyfat log(Hip) 0.402761495034679"
[1] "max ~ bodyfat Chest^2 0.491650041989838"
[1] "max ~ bodyfat Abdomen^2 0.660318770251814"
[1] "max ~ bodyfat Age + Chest 0.518792392327339"
[1] "max ~ bodyfat Age + Abdomen 0.670476701636566"
[1] "max ~ bodyfat Age + log(Chest) 0.522126896819114"
[1] "max ~ bodyfat Age + log(Abdomen) 0.682715559529513"
[1] "max ~ bodyfat Age + Abdomen^2 0.670476701636566"
[1] "max ~ bodyfat Weight + Abdomen 0.716539455076649"
[1] "max ~ bodyfat Weight + log(Abdomen) 0.72562296592071"
[1] "max ~ bodyfat Weight + Abdomen^2 0.716539455076649"
[1] "max ~ bodyfat Height + Abdomen 0.685262389974487"
[1] "max ~ bodyfat Height + log(Abdomen) 0.701859180798446"
[1] "max ~ bodyfat Neck + log(Abdomen) 0.709387941823287"
[1] "max ~ bodyfat Abdomen + Weight^2 0.716539455076649"
[1] "max ~ bodyfat Wrist + log(Abdomen) 0.717469750714956"
[1] "max ~ bodyfat log(Weight) + log(Abdomen) 0.71774626046345"
[1] "max ~ bodyfat log(Abdomen) + log(Wrist) 0.716917296661175"
[1] "max ~ bodyfat log(Abdomen) + Weight^2 0.72562296592071"
[1] "max ~ bodyfat log(Abdomen) + Wrist^2 0.717469750714956"
```

Highest adjusted r-squared:

```
$adj.r.squared
[1] 0.725623
```

$adj.r.squared
[1] 0.725623

$adj.r.squared
[1] 0.7174698

When k = 3:
Results from best fit list:
[1] "max ~ bodyfat Weight 0.37255111373424"
[1] "max ~ bodyfat Neck 0.237643088129366"
[1] "max ~ bodyfat Chest 0.491650041989838"
[1] "max ~ bodyfat Abdomen 0.660318770251814"
[1] "max ~ bodyfat Hip 0.388439692002688"
[1] "max ~ bodyfat Thigh 0.310413232223974"
[1] "max ~ bodyfat log(Weight) 0.392820011398408"
[1] "max ~ bodyfat log(Chest) 0.495209083100418"
[1] "max ~ bodyfat log(Abdomen) 0.674057641566212"
[1] "max ~ bodyfat log(Hip) 0.402761495034679"
[1] "max ~ bodyfat Chest^2 0.491650041989838"
[1] "max ~ bodyfat Abdomen^2 0.660318770251814"
[1] "max ~ bodyfat Age + Chest 0.518792392327339"
[1] "max ~ bodyfat Age + Abdomen 0.670476701636566"
[1] "max ~ bodyfat Age + log(Chest) 0.522126896819114"
[1] "max ~ bodyfat Age + log(Abdomen) 0.682715559529513"
[1] "max ~ bodyfat Age + Abdomen^2 0.670476701636566"
[1] "max ~ bodyfat Weight + Abdomen 0.716539455076649"
[1] "max ~ bodyfat Weight + log(Abdomen) 0.72562296592071"
[1] "max ~ bodyfat Weight + Abdomen^2 0.716539455076649"
[1] "max ~ bodyfat Height + Abdomen 0.685262389974487"
[1] "max ~ bodyfat Height + log(Abdomen) 0.701859180798446"
[1] "max ~ bodyfat Neck + log(Abdomen) 0.709387941823287"
[1] "max ~ bodyfat Abdomen + Weight^2 0.716539455076649"
[1] "max ~ bodyfat Wrist + log(Abdomen) 0.717469750714956"
[1] "max ~ bodyfat log(Weight) + log(Abdomen) 0.71774626046345"
[1] "max ~ bodyfat log(Abdomen) + log(Wrist) 0.716917296661175"
[1] "max ~ bodyfat log(Abdomen) + Weight^2 0.72562296592071"
[1] "max ~ bodyfat log(Abdomen) + Wrist^2 0.717469750714956"
[1] "max ~ bodyfat Age + Weight + log(Abdomen) 0.724636985647084"
[1] "max ~ bodyfat Age + Wrist + log(Abdomen) 0.730459566009974"
[1] "max ~ bodyfat Age + log(Abdomen) + log(Wrist) 0.729907970595516"
[1] "max ~ bodyfat Age + log(Abdomen) + Wrist^2 0.730459566009974"
[1] "max ~ bodyfat Weight + Height + log(Abdomen) 0.729584837544779"

[1] "max ~ bodyfat Weight + Neck + log(Abdomen) 0.730816026646246"
[1] "max ~ bodyfat Weight + Abdomen + log(Weight) 0.730222886370719"
[1] "max ~ bodyfat Weight + Wrist + log(Abdomen) 0.737437420033583"
[1] "max ~ bodyfat Weight + log(Neck) + log(Abdomen) 0.730758091646047"
[1] "max ~ bodyfat Weight + log(Abdomen) + log(Wrist) 0.737341875258134"
[1] "max ~ bodyfat Weight + log(Abdomen) + sin(Wrist) 0.734145689147994"
[1] "max ~ bodyfat Weight + log(Abdomen) + Wrist^2 0.737437420033583"
[1] "max ~ bodyfat Wrist + log(Abdomen) + Weight^2 0.737437420033583"

[1] "**bodyfat = -346.030186086169 + 208.728057539987 * log10(Abdomen)+-0.0992285090208676 * ceiling(Weight)+-1.47439740274781 * Wrist"**
$adj.r.squared
[1] 0.7377635

Highest adjusted r-squared:
$adj.r.squared
[1] 0.7377635

$adj.r.squared
[1] 0.7374374

$adj.r.squared
[1] 0.7374374

$adj.r.squared
[1] 0.7374374

There was an increase with the values when k = 3.