628 Body Fat Data Project

Group 3 Members:

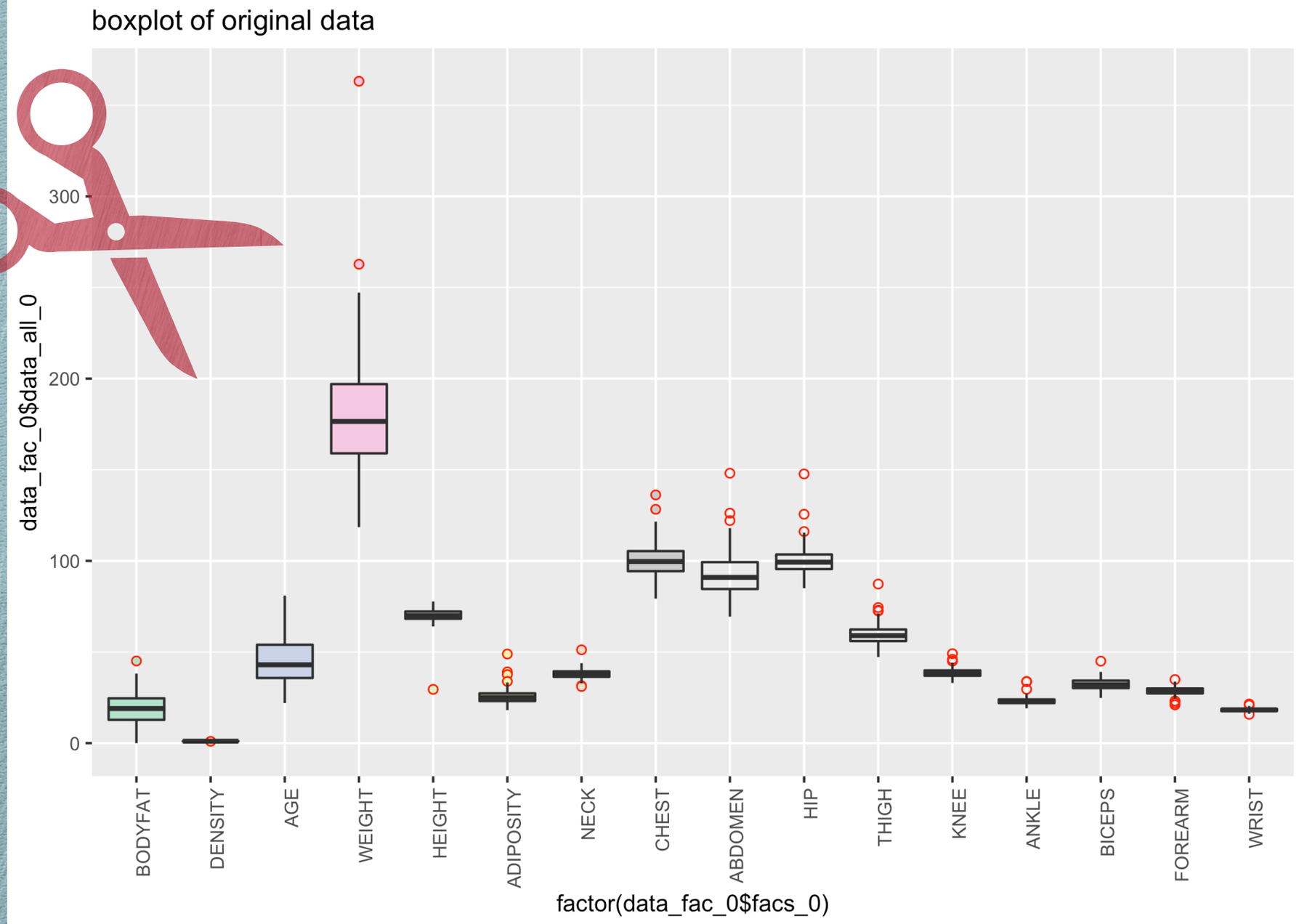
Lingfeng Zhu lzhu88@wisc.edu

Ruochen Yin ryin26@wisc.edu

Jiahan Li jli936@wisc.edu

Chong Wei cwei48@wisc.edu

OUTLINE



Data Cleaning

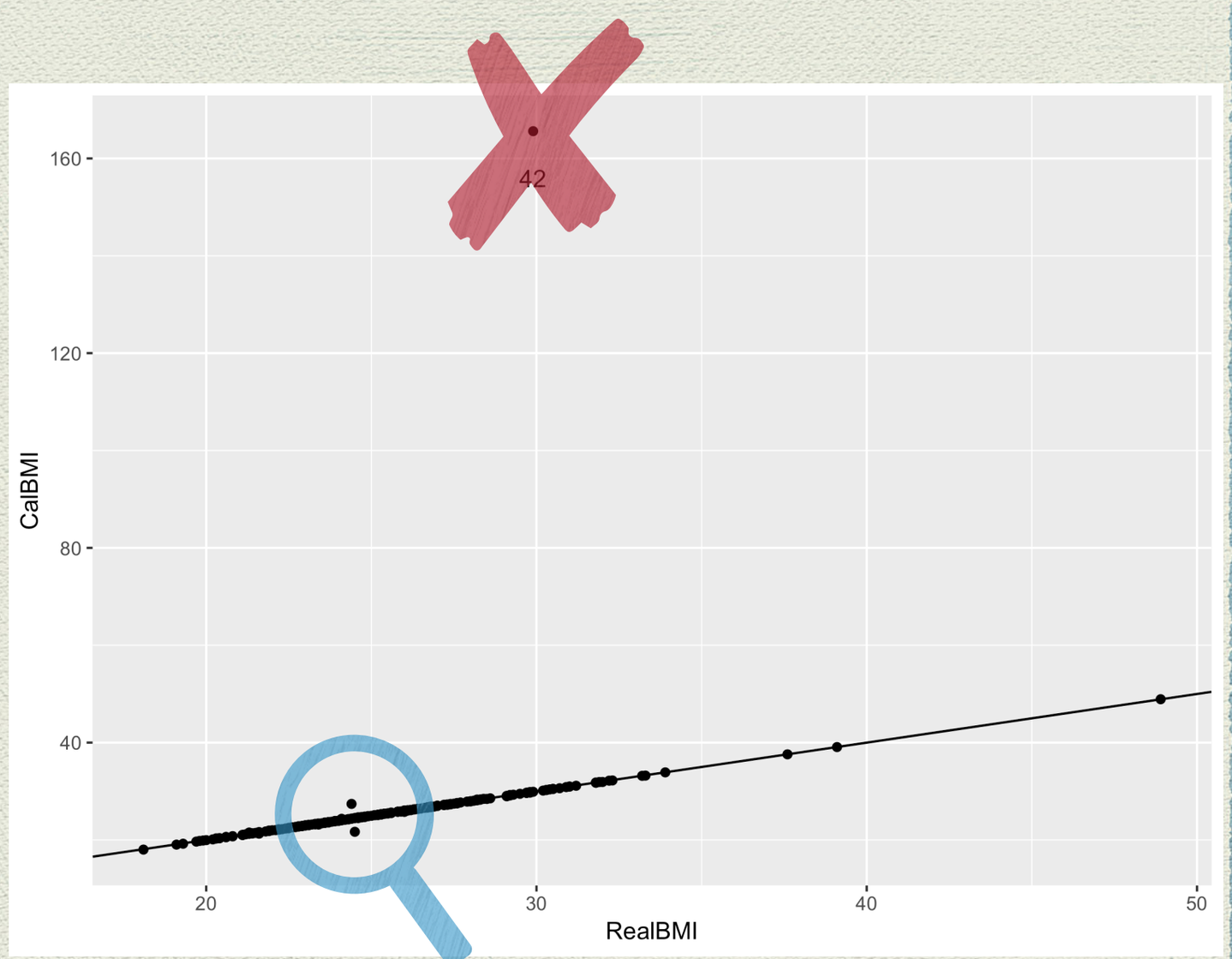
boxplot of original data

Check Quantile

- Calculating the rage of weight, height, density, bodyfat between quantile 0.01, 0.99 :
- Weight_q = (125.505, 245.720)
- Height_q = (64.3825, 76.0000)
- Density_q = (1.016040, 1.095393)
- Bodyfat_q = (4.355, 35.582)

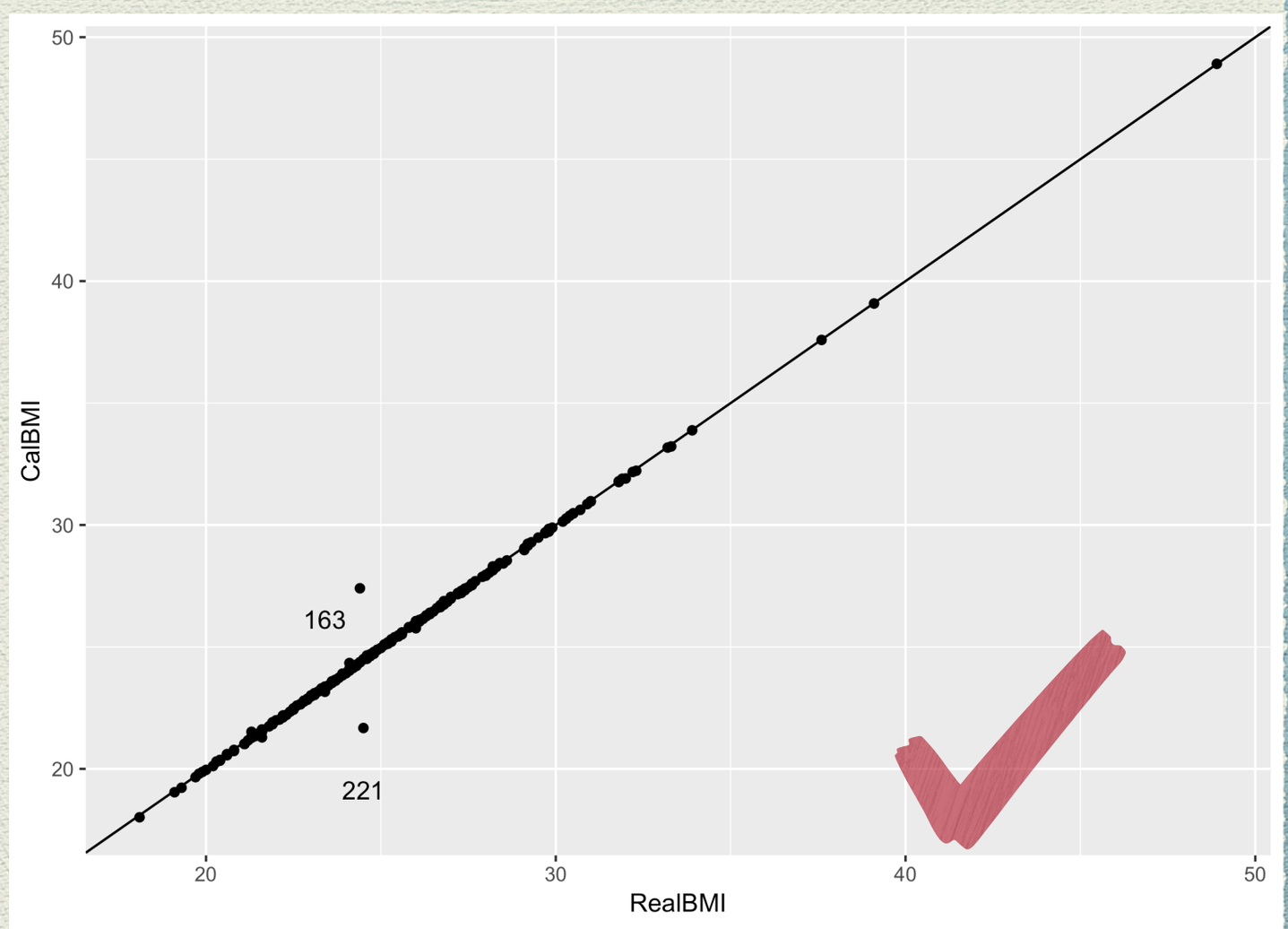
Check outliers —- Bodyfat

- Relationship between bodyfat and density:
- $$BODYFAT = \frac{495}{DENSITY} - 450$$



Check outliers —- Bodyfat

- Relationship between bodyfat and density:
- $$BODYFAT = \frac{495}{DENSITY} - 450$$



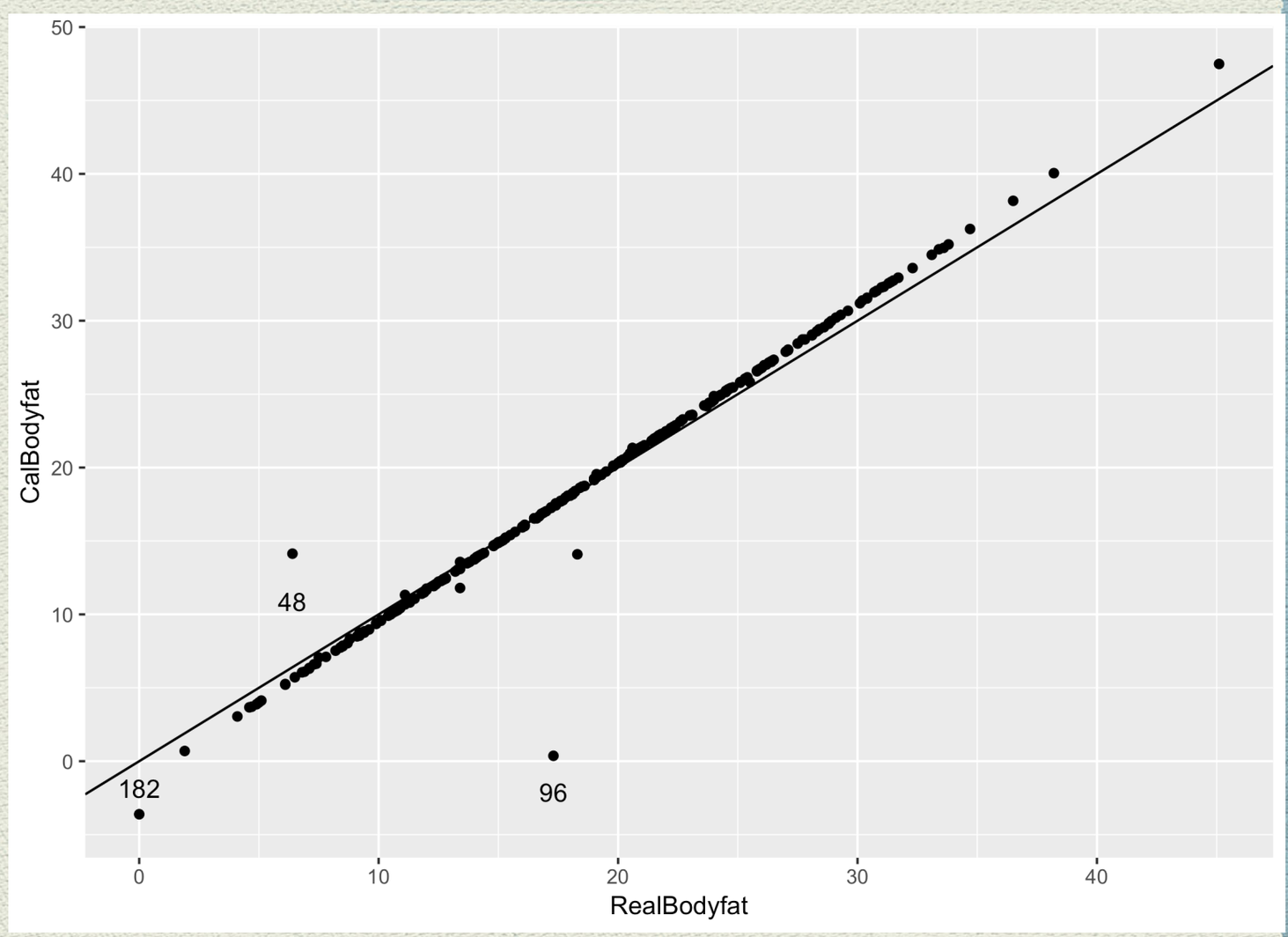
Check outliers

- From formula of BMI, we have point 42's height 29.5 is abnormal, so we change it according to the formula and let it become 69.43.
- Although point 163, 221 are not perfectly matched the calculated BMI, but their height and weight is in the range, so we won't change them.

Check outliers —- BMI

- Relationship among adiposity, weight and height:

- $$BMI = \frac{703 * WEIGHT}{HEIGHT^2}$$



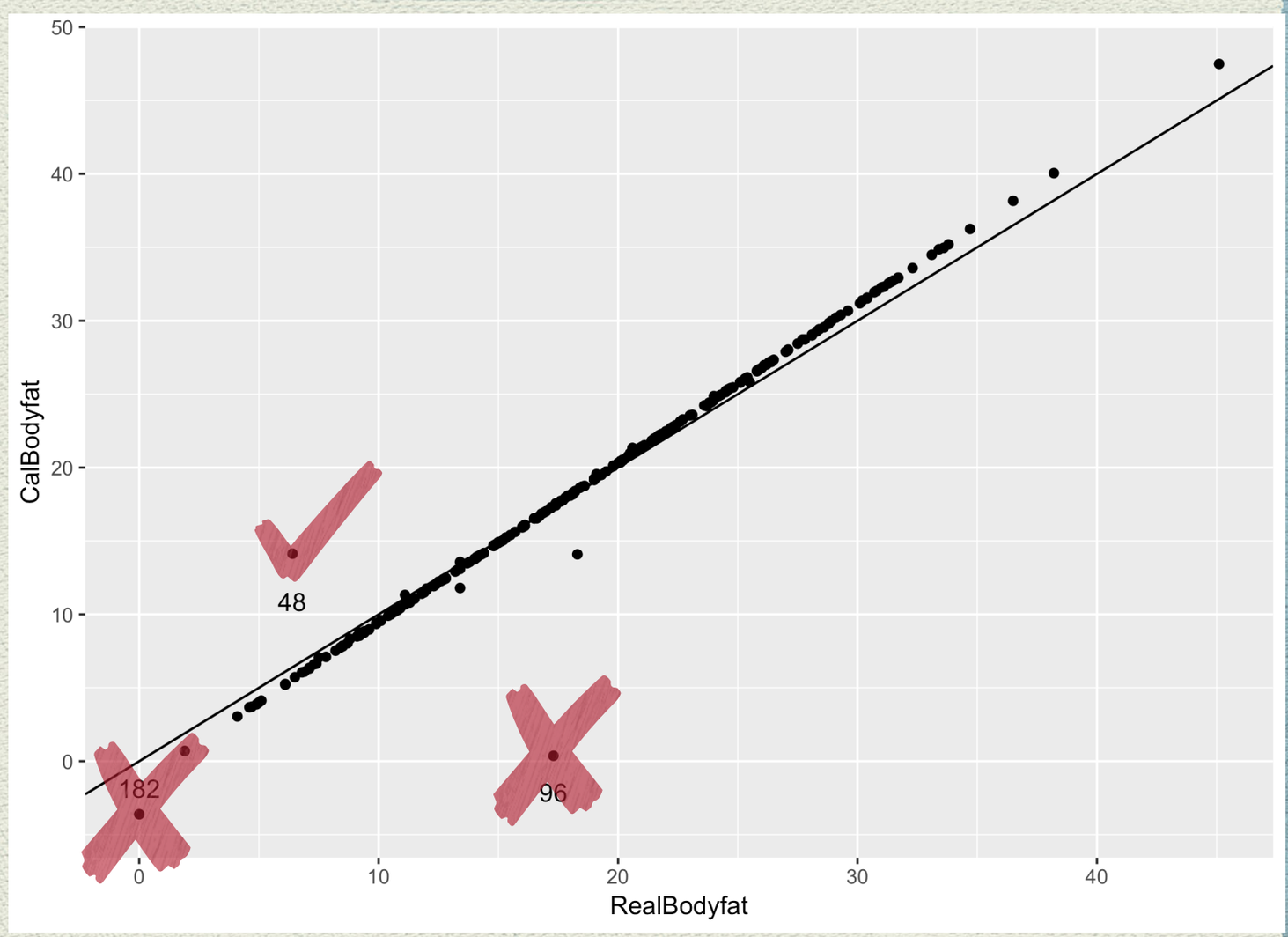
Check outliers

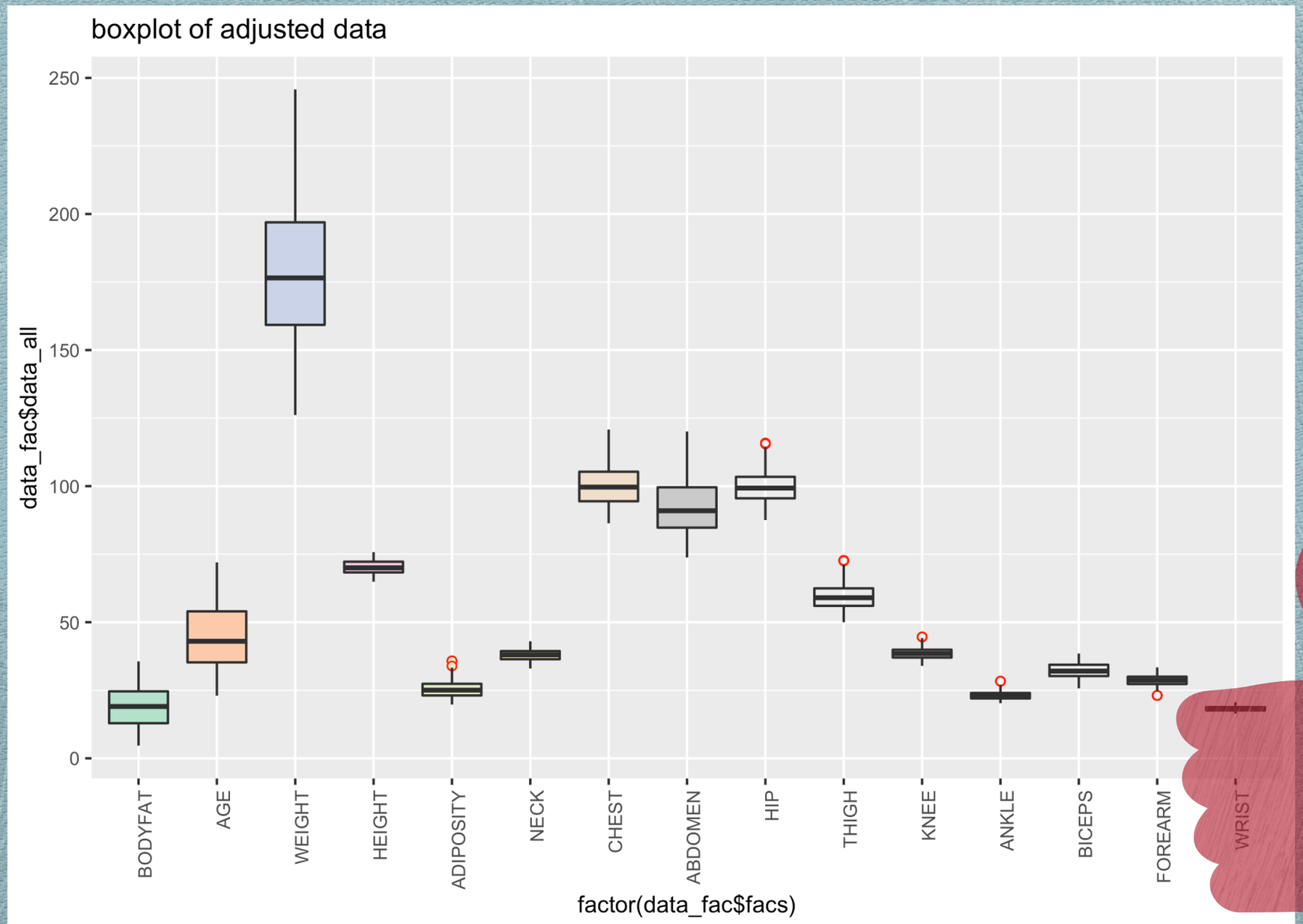
- From formula of Bodyfat, point 182 has the real bodyfat lower than zero, which is impossible. Besides, the density is also close to 0 which means we can’t adjust it, so we just delete 182.
- Point 48, 96 is away from the calculated bodyfat, too. Point 48 has normal density and bodyfat, so we will keep it. However, point 96 has density away from the range, so we will delete point 96.
- In conclusion, we delete 2 points:96,182

Check outliers —- BIM

- Relationship among adiposity, weight and height:

- $$BIM = \frac{703 * WEIGHT}{HEIGHT^2}$$





Data Cleaning

boxplot of adjusted data

AIC

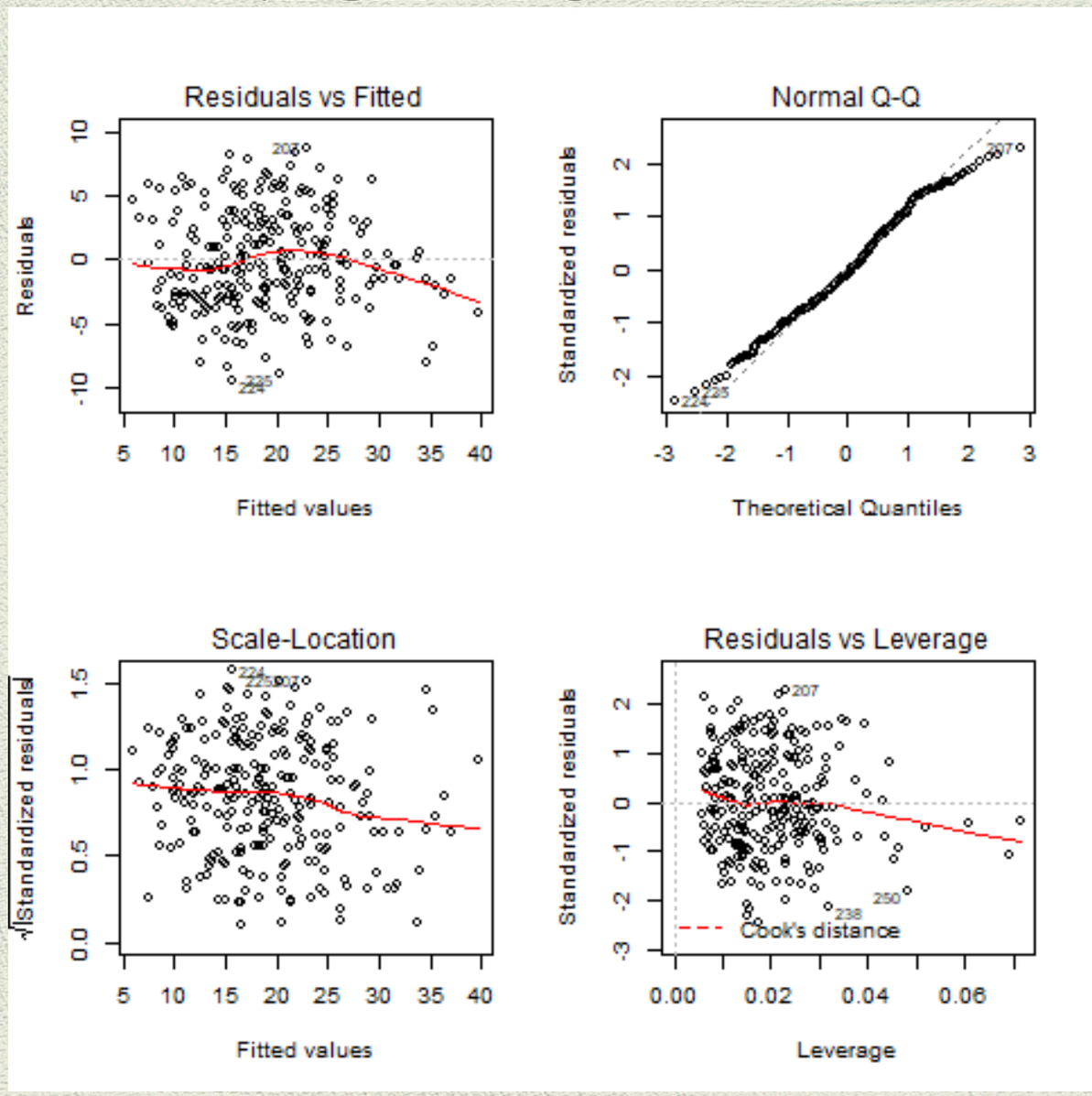
BIC

LASSO

Feature Selection

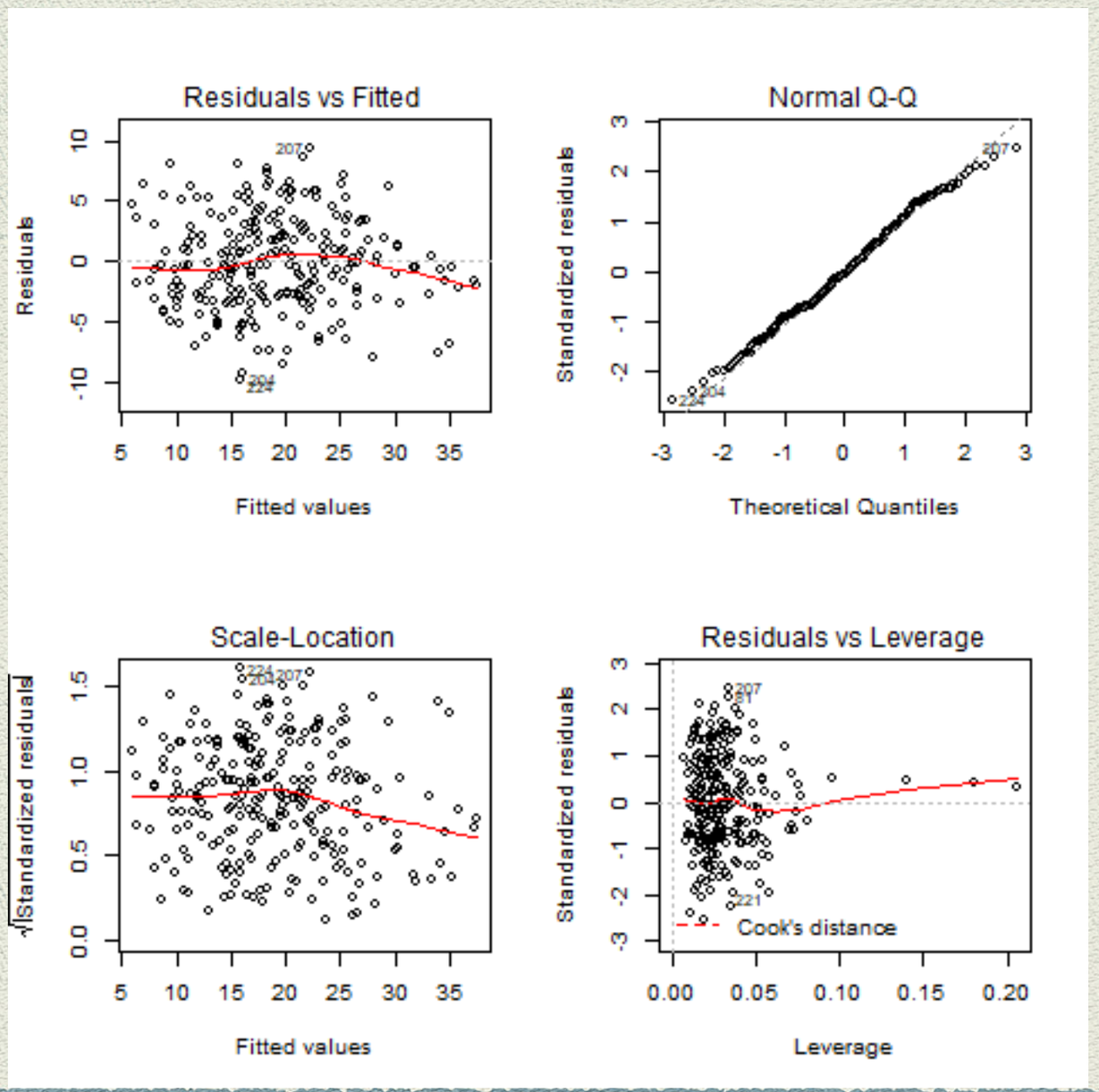
AIC

BODYFAT ~ ABDOMEN + WEIGHT + WRIST + AGE + THIGH + NECK + FOREARM



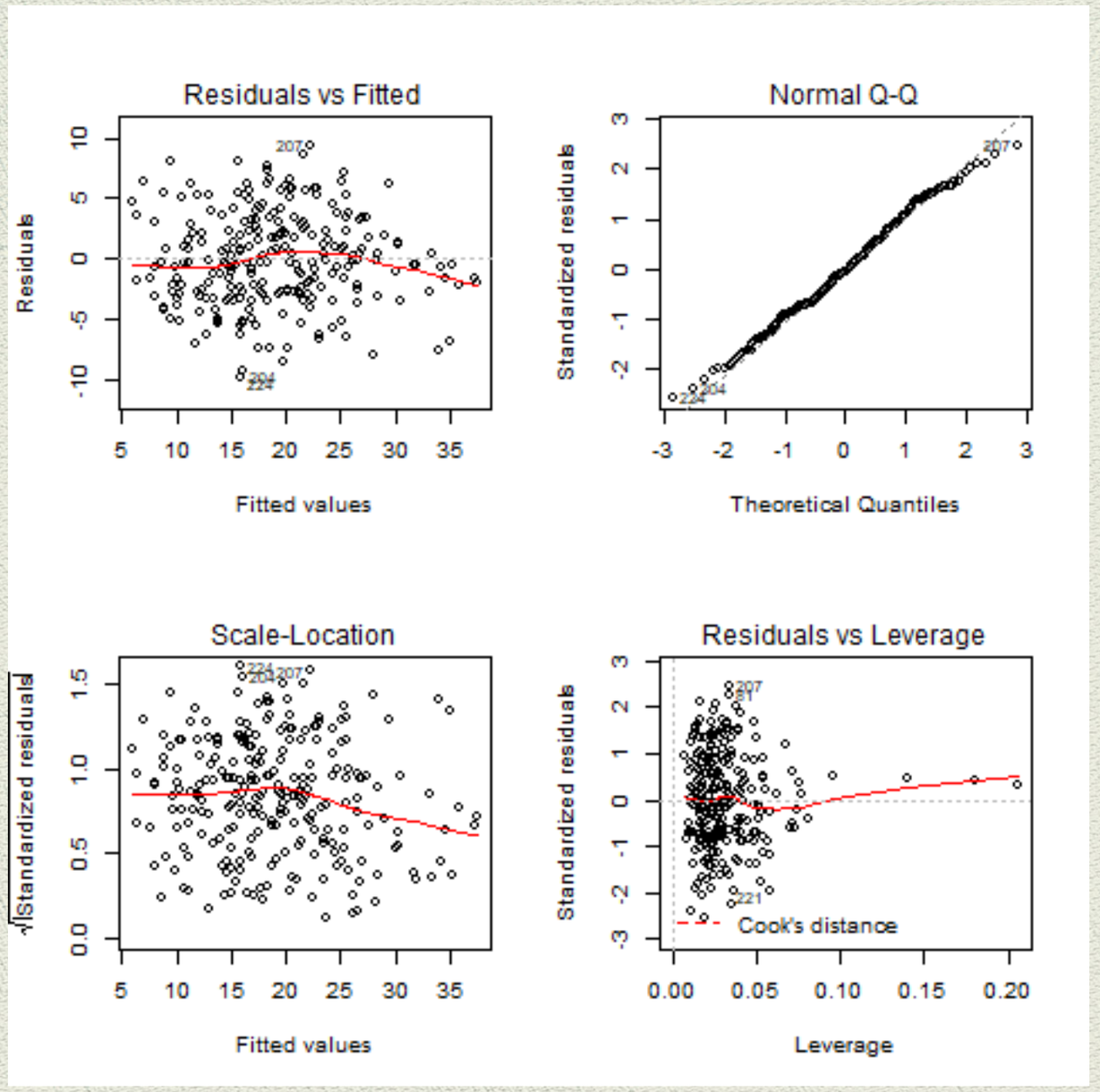
BIC

BODYFAT ~ AGE + ABDOMEN + WRIST + HEIGHT



LASSO

BODYFAT ~ AGE + HEIGHT + ABDOMEN + WRIST



Comparison

AIC

BODYFAT ~ ABDOMEN + WEIGHT + WRIST + AGE + THIGH + NECK + FOREARM

R-square : 0.7418 Cv-mean: 15.4747

BIC

BODYFAT ~ AGE + HEIGHT + ABDOMEN + WRIST

R-square : 0.7336 Cv-mean: 15.6591

LASSO

BODYFAT ~ AGE + HEIGHT + ABDOMEN + WRIST

R-square : 0.7336 Cv-mean: 15.6591

Model 1: Bodyfat ~ Weight + Abdomen

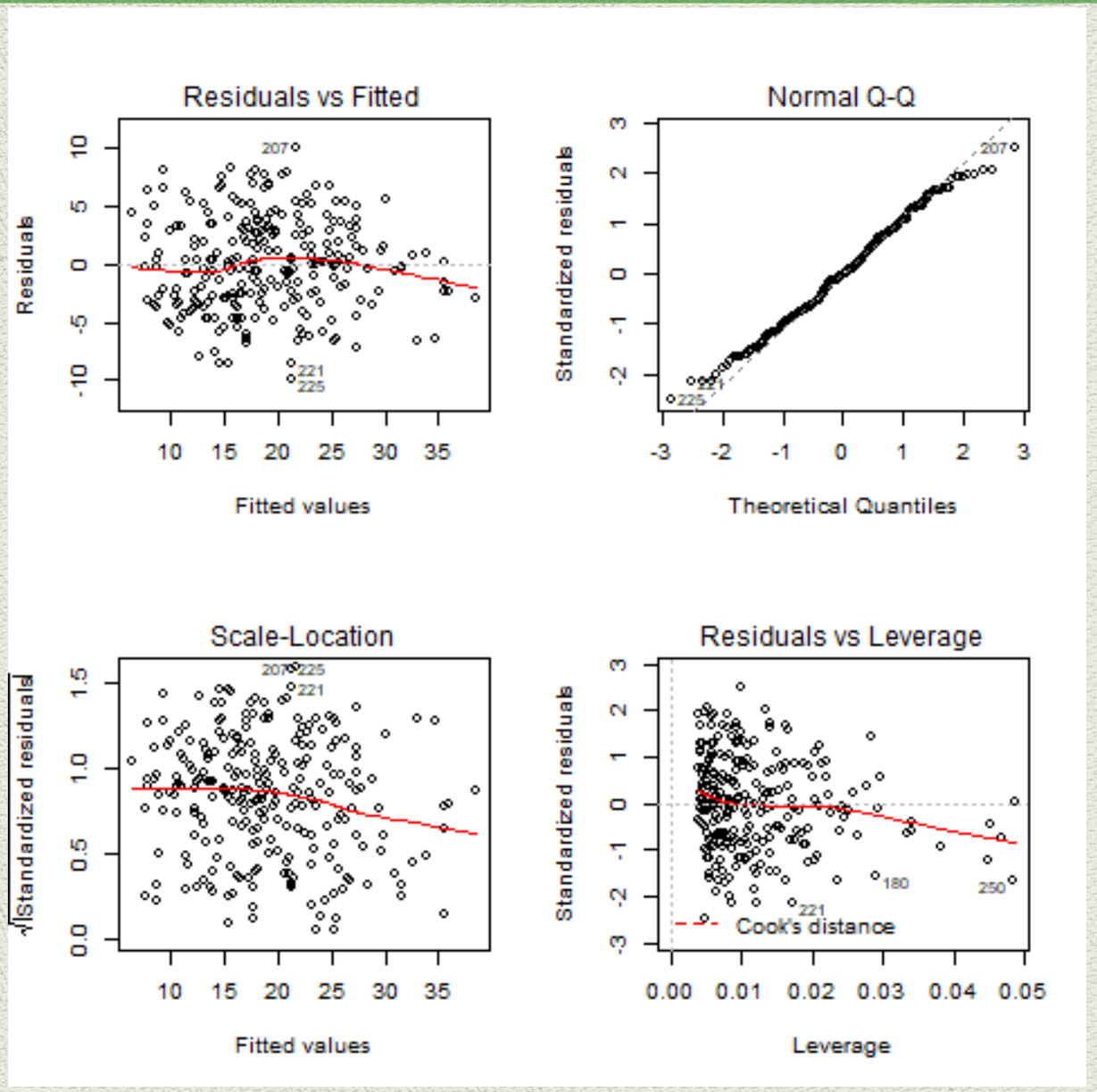
Model 2: Bodyfat ~ Weight * Abdomen

Model 3: Bodyfat ~ Abdomen

Trying other models

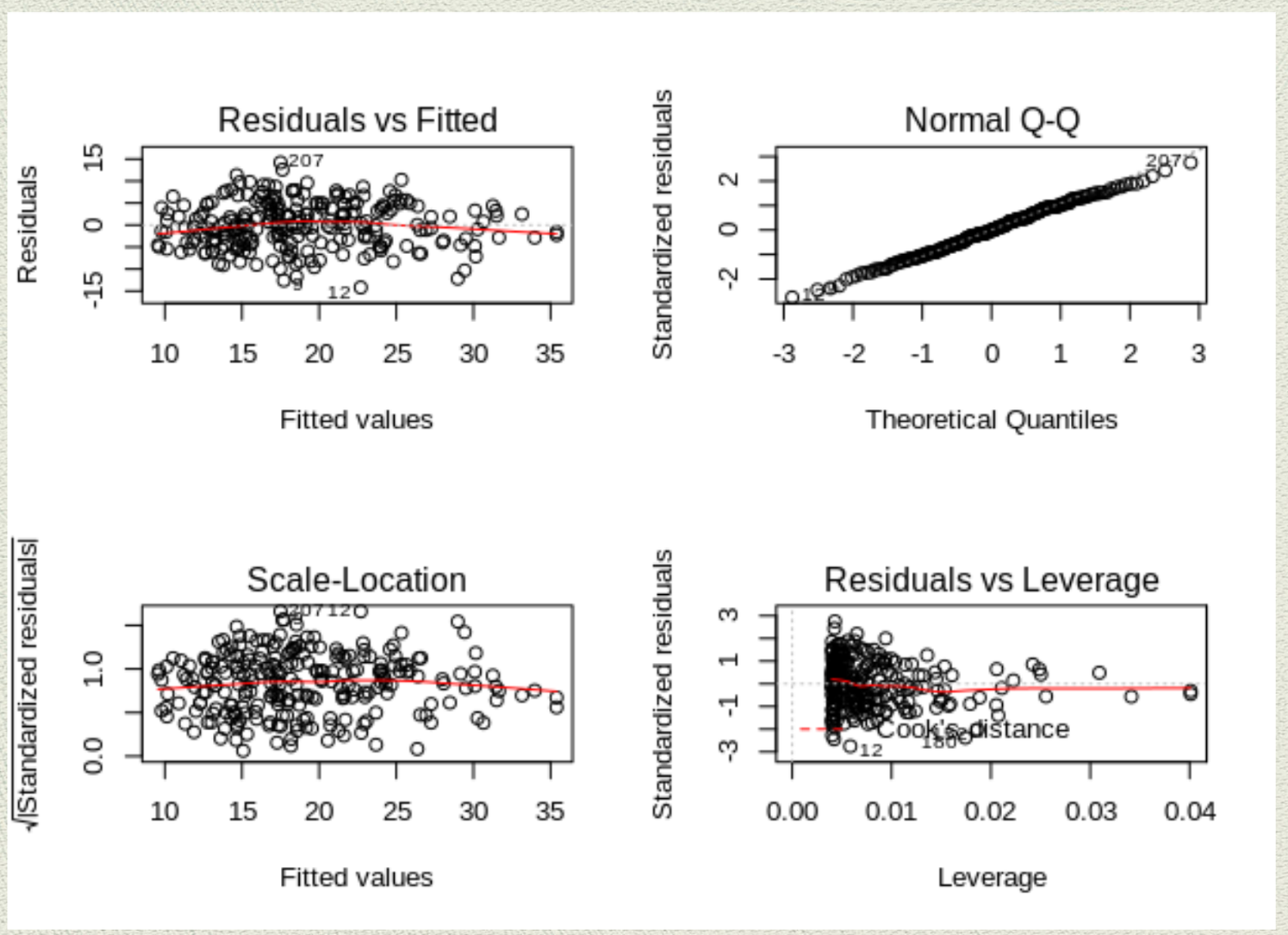
Model 1: Bodyfat ~ Weight + Abdomen

R-square : 0.7181 Cv-mean: 16.2636



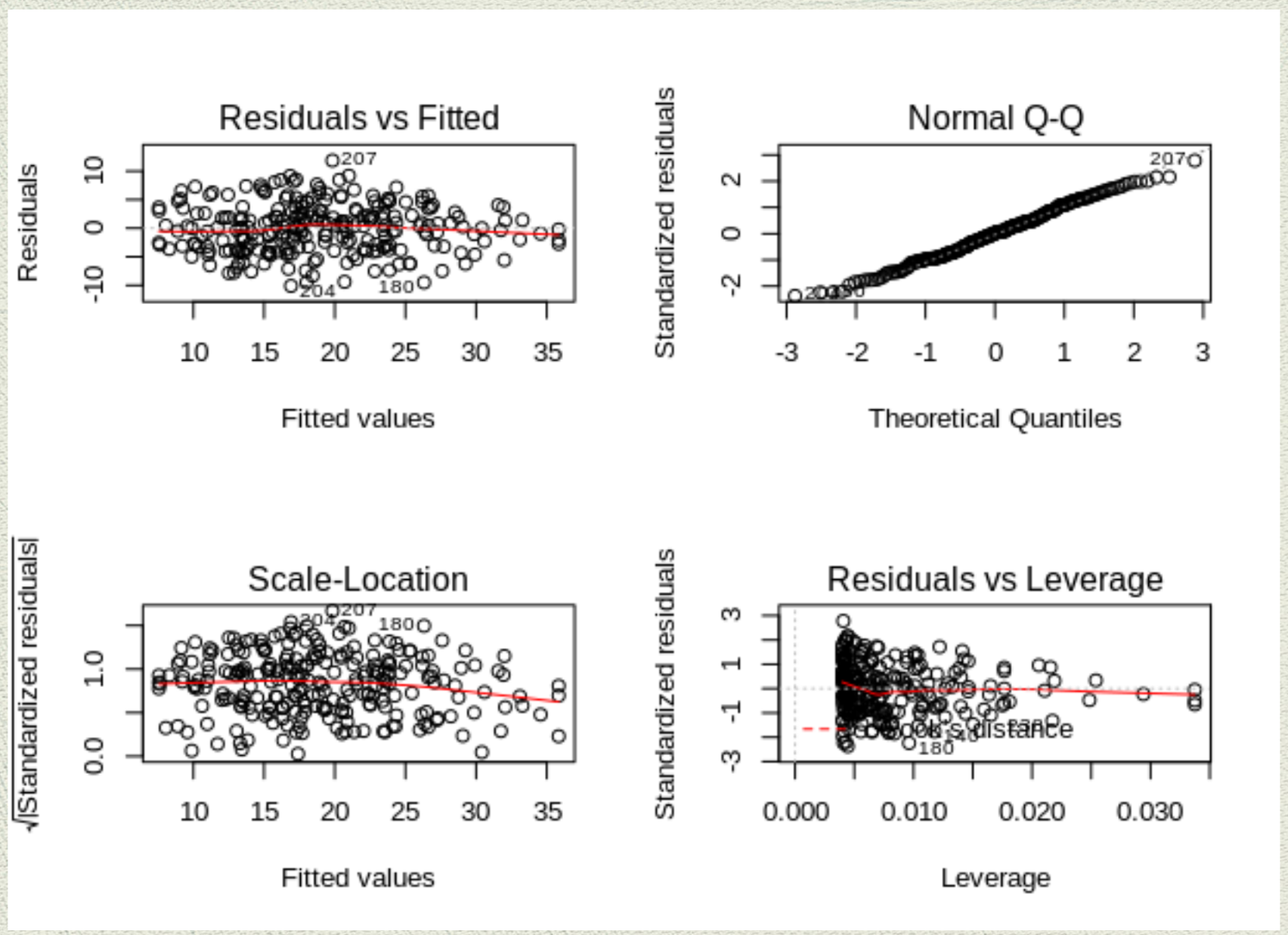
Model 2: Bodyfat ~ Weight * Abdomen

R-square : 0.5293 Cv-mean: 26.9237



Model 3: Bodyfat ~ Abdomen

R-square : 0.6785 Cv-mean: 18.3760



Comparison

BIC

BODYFAT ~ AGE + HEIGHT + ABDOMEN + WRIST

R-square : 0.7336 Cv-mean: 15.6591

Model 1

BODYFAT ~ WEIGHT + ABDOMEN

R-square : 0.7181 Cv-mean: 16.2636

Model 2

BODYFAT ~ WEIGHT * ABDOMEN

R-square : 0.5293 Cv-mean: 26.9237

Model 3

BODYFAT ~ ABDOMEN

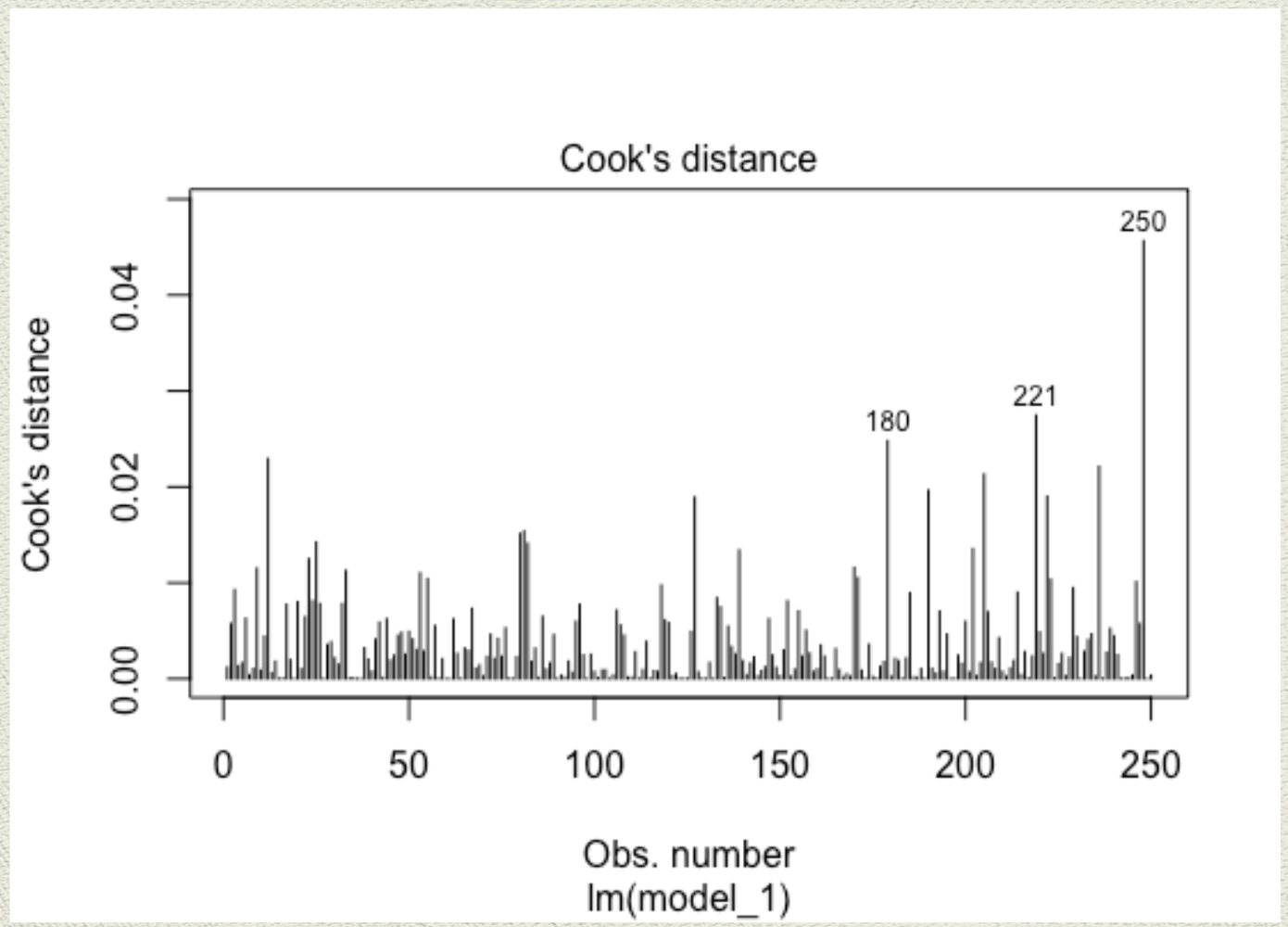
R-square : 0.6785 Cv-mean: 18.3760

Model One

BODY FAT = 2.94338 + 0.05205 * AGE + 0.69798 * ABDOMEN - 1.71840 * WRIST - 0.27799 * HEIGHT

R-square : 0.7336 Cv-mean: 15.6591 P-value < 2.2e-16

Possible Rule of Thumb:
Your body fat equal to
Your abdomen (cm)
multiply by 0.7
minus wrist (cm) multiply
by 1.7 minus height(inch)
multiply by 0.28 add 3.0



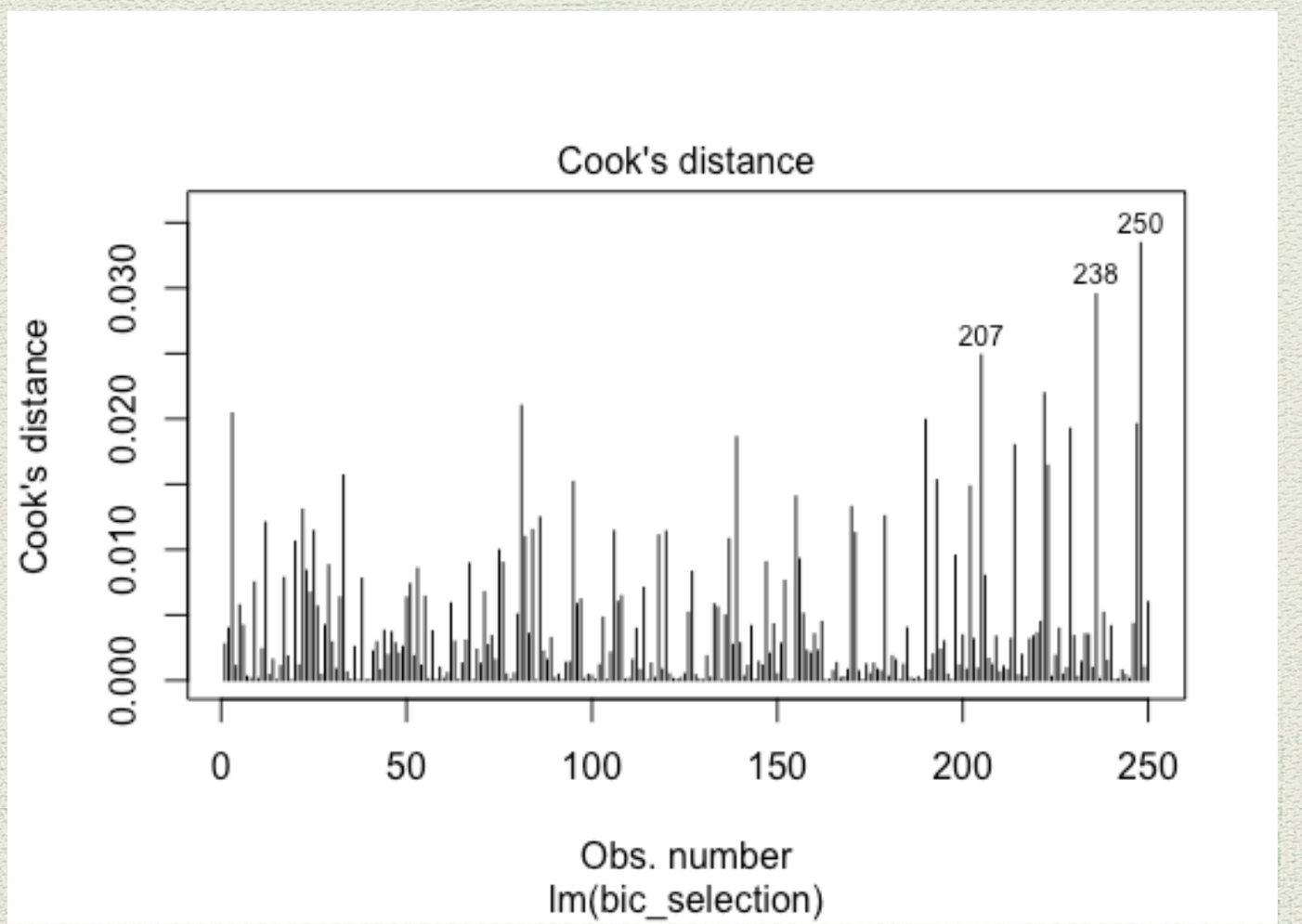
Model Two

BODY FAT = -41.60272 – 0.11553 * WEIGHT
+ 0.87794 * ABDOMEN

R-square : 0.7181 Cv-mean: 16.2636 P-value < 2.2e-16

Possible Rule of Thumb:

Your body fat equal to
Your abdomen (cm)
multiply by 0.88
minus weight (lbs)
multiply by 0.12 minus
41.6



Strengths:

1. To deal with outliers, except for using the 2 formulas to correct or delete abnormal points, we also use 1% and 99% quantile of each variables to replace the value under or beyond these 2 quantiles, which is a reasonable way to reduce outliers.

2. We use Cross Validation to make our result more convincible.

3. Our two models are all linear, which are simple and with only two or four significant variables, which are all easy to measure.

Weaknesses

1. For simplicity, the number of predictor variables are not enough.

2. The method we use to deal with outliers may to some extent hide the true situation of the sample.

Summary

Finally, we have two linear model. Model one contains 4 variables which are age, wrist, height and abdomen. Model two contains 2 variables which are abdomen and weight. All the coefficients also significant at the 0.05 level. They don't have a large different and they are both simple and accurate. So we actually used both of them when we built shiny app. Users can select either of them to calculate body fat. But because model two have less variables, so we prefer model 2 if we can only choose one.

Thank You !

Here is our shiny app:
https://lingfengzhu.shinyapps.io/bodyfat/