P8130 Final Project Code

```
library(tidyverse)
library(readr)
library(readxl)
library(GGally)
library(mgcv)
library(ggplot2)
library(glmnet)
library(caret)
library(modelr)
library(mgcv)
library(broom)
```

Load Data

```
data <- readxl::read_excel("body_density_data.xlsx") %>%
  select (-bodyfat_brozek, -bodyfat_siri, -id)
```

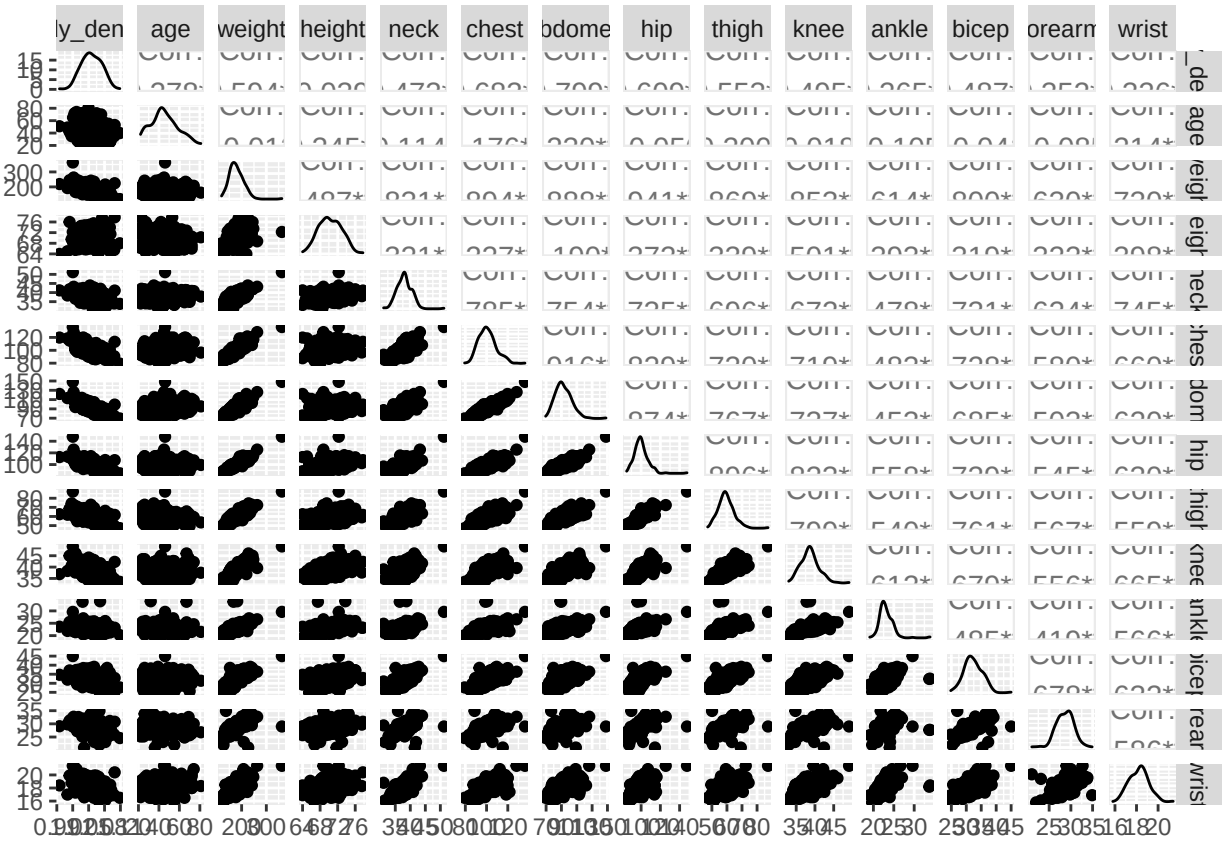
Descriptive Statistics

```
summary(data)
```

```
##   body_density        age            weight          height     
##  Min.   :0.995   Min.   :22.00   Min.   :118.5   Min.   :64.00  
##  1st Qu.:1.041   1st Qu.:35.75   1st Qu.:159.0   1st Qu.:68.25  
##  Median :1.055   Median :43.00   Median :176.5   Median :70.00  
##  Mean   :1.056   Mean   :44.88   Mean   :178.9   Mean   :70.31  
##  3rd Qu.:1.070   3rd Qu.:54.00   3rd Qu.:197.0   3rd Qu.:72.25  
##  Max.   :1.109   Max.   :81.00   Max.   :363.1   Max.   :77.75  
##       neck            chest           abdomen            hip       
##  Min.   :31.10   Min.   : 79.30   Min.   : 69.40   Min.   : 85.0  
##  1st Qu.:36.40   1st Qu.: 94.35   1st Qu.: 84.58   1st Qu.: 95.5  
##  Median :38.00   Median : 99.65   Median : 90.95   Median : 99.3  
##  Mean   :37.99   Mean   :100.82   Mean   : 92.56   Mean   : 99.9  
##  3rd Qu.:39.42   3rd Qu.:105.38   3rd Qu.: 99.33   3rd Qu.:103.5  
##  Max.   :51.20   Max.   :136.20   Max.   :148.10   Max.   :147.7  
##      thigh            knee            ankle           bicep      
##  Min.   :47.20   Min.   :33.00   Min.   :19.1   Min.   :24.80  
##  1st Qu.:56.00   1st Qu.:36.98   1st Qu.:22.0   1st Qu.:30.20  
##  Median :59.00   Median :38.50   Median :22.8   Median :32.05  
##  Mean   :59.41   Mean   :38.59   Mean   :23.1   Mean   :32.27  
##  3rd Qu.:62.35   3rd Qu.:39.92   3rd Qu.:24.0   3rd Qu.:34.33  
##  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.:16.60  
##  Median :28.70   Median :17.10  
##  Mean   :28.66   Mean   :17.15  
##  3rd Qu.:30.00   3rd Qu.:17.70  
##  Max.   :34.90   Max.   :21.40
```

```
##  Min.   :15.80
##  1st Qu.:17.60
##  Median :18.30
##  Mean   :18.23
##  3rd Qu.:18.80
##  Max.   :21.40
```

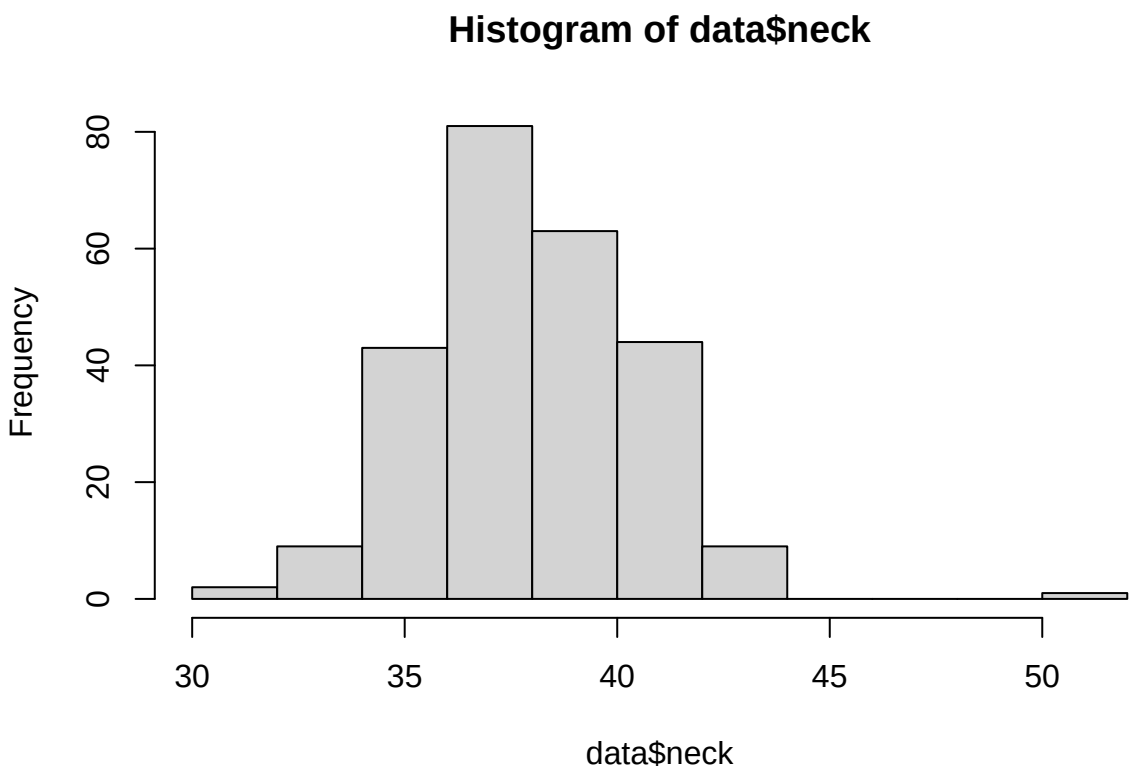
```
ggpairs(data)
```



neck, ankle, abdomen, weight might need transformation. But we will come back to it aftere model diagnostic

neck

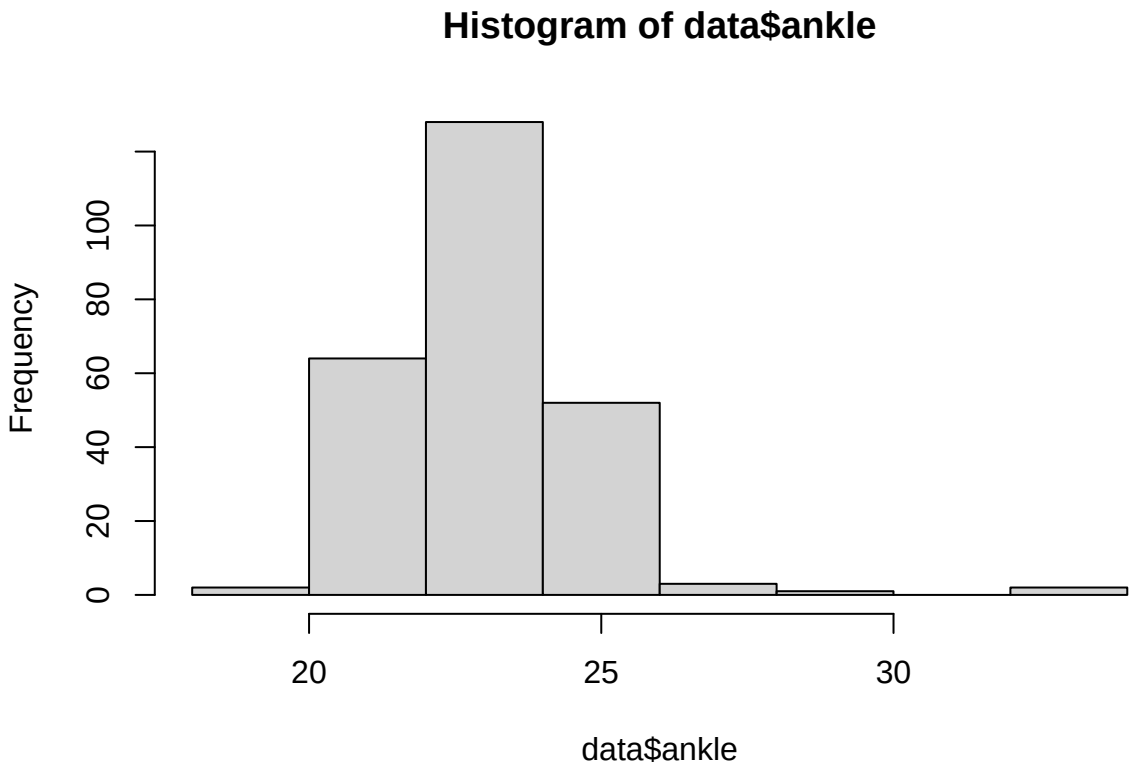
```
hist(data$neck)
```



It's just an outlier

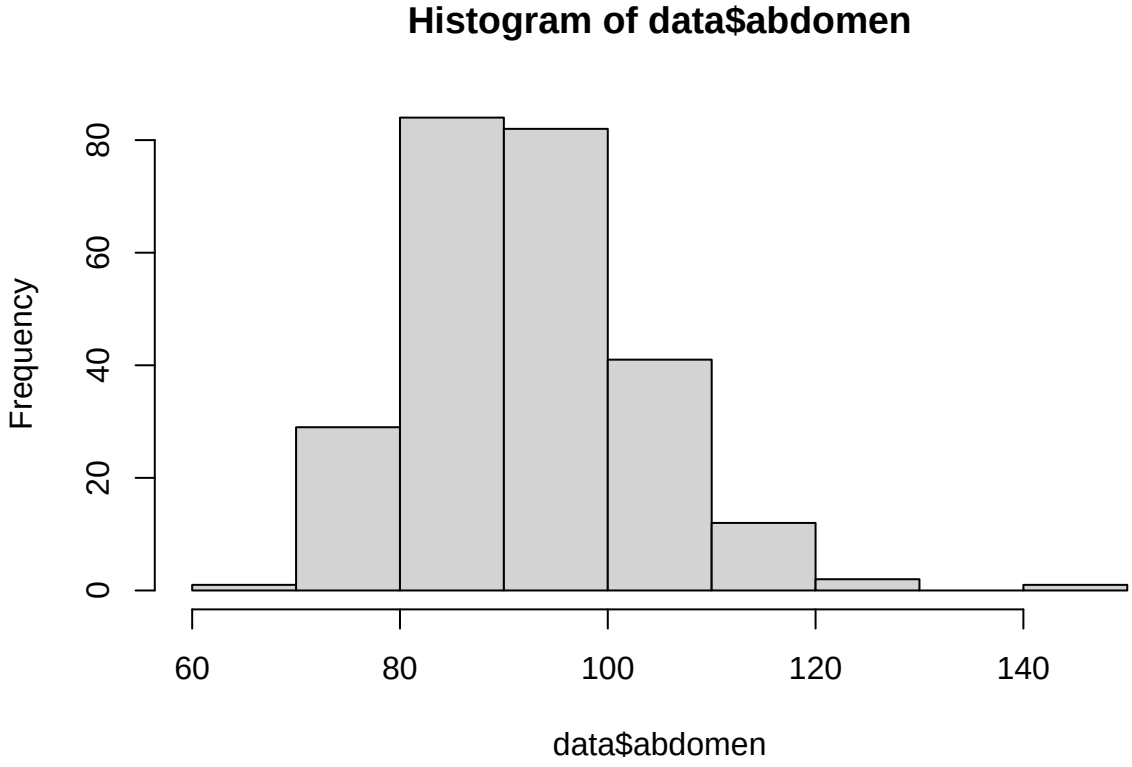
ankle

```
hist(data$ankle)
```



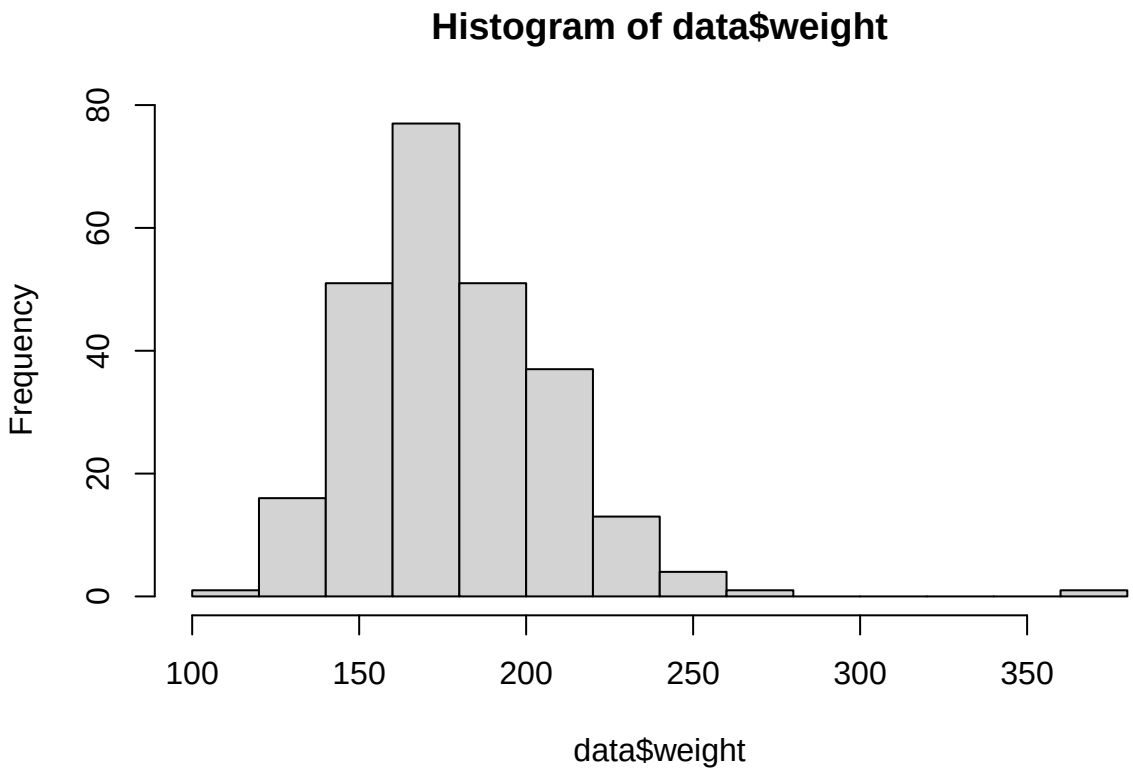
abdomen

```
hist(data$abdomen)
```



weight

```
hist(data$weight)
```



All outliers above are denoting a specific group of participants.

Participant #39 is an outlier. This participant has outlying measurement for weight and abdomen. Since he/she also has a larger chest circumference, the data point is less likely to be a measurement error. It could has useful info.

No transformation for all for nows

Variable Selection

Backward selection

```
mult.fit <- lm(body_density ~ ., data = data)
step(mult.fit, direction='backward')
```

```
## Start:  AIC=-2306.93
## body_density ~ age + weight + height + neck + chest + abdomen +
##     hip + thigh + knee + ankle + bicep + forearm + wrist
##
##           Df Sum of Sq      RSS     AIC
## - knee     1 0.0000009 0.023846 -2308.9
## - height   1 0.0000095 0.023855 -2308.8
## - chest    1 0.0000232 0.023868 -2308.7
## - ankle    1 0.0001512 0.023996 -2307.3
## <none>                 0.023845 -2306.9
```

```
## - bicep    1 0.0002048 0.024050 -2306.8
## - hip      1 0.0002362 0.024081 -2306.4
## - age      1 0.0002811 0.024126 -2306.0
## - weight   1 0.0003018 0.024147 -2305.8
## - thigh    1 0.0003499 0.024195 -2305.3
## - neck     1 0.0003995 0.024245 -2304.7
## - forearm  1 0.0004916 0.024337 -2303.8
## - wrist    1 0.0010602 0.024905 -2298.0
## - abdomen  1 0.0117925 0.035638 -2207.7
##
## Step:  AIC=-2308.92
## body_density ~ age + weight + height + neck + chest + abdomen +
##     hip + thigh + ankle + bicep + forearm + wrist
##
##           Df Sum of Sq      RSS     AIC
## - height   1 0.0000086 0.023855 -2310.8
## - chest    1 0.0000234 0.023870 -2310.7
## - ankle    1 0.0001610 0.024007 -2309.2
## <none>                 0.023846 -2308.9
## - bicep    1 0.0002040 0.024050 -2308.8
## - hip      1 0.0002352 0.024081 -2308.4
## - weight   1 0.0003015 0.024148 -2307.8
## - age      1 0.0003093 0.024155 -2307.7
## - thigh    1 0.0003991 0.024245 -2306.7
## - neck     1 0.0004089 0.024255 -2306.6
## - forearm  1 0.0004986 0.024345 -2305.7
## - wrist    1 0.0010613 0.024907 -2299.9
## - abdomen  1 0.0117917 0.035638 -2209.7
##
## Step:  AIC=-2310.83
## body_density ~ age + weight + neck + chest + abdomen + hip +
##     thigh + ankle + bicep + forearm + wrist
##
##           Df Sum of Sq      RSS     AIC
## - chest    1 0.0000165 0.023871 -2312.7
## - ankle    1 0.0001673 0.024022 -2311.1
## <none>                 0.023855 -2310.8
## - bicep    1 0.0002206 0.024075 -2310.5
## - hip      1 0.0002267 0.024081 -2310.4
## - age      1 0.0003118 0.024167 -2309.6
## - neck     1 0.0004011 0.024256 -2308.6
## - thigh    1 0.0004393 0.024294 -2308.2
## - forearm  1 0.0004947 0.024350 -2307.7
## - weight   1 0.0006990 0.024554 -2305.6
## - wrist    1 0.0010744 0.024929 -2301.7
## - abdomen  1 0.0132581 0.037113 -2201.4
##
## Step:  AIC=-2312.66
## body_density ~ age + weight + neck + abdomen + hip + thigh +
##     ankle + bicep + forearm + wrist
##
##           Df Sum of Sq      RSS     AIC
## - ankle    1 0.0001706 0.024042 -2312.9
## <none>                 0.023871 -2312.7
```

```
## - bicep    1 0.0002106 0.024082 -2312.4
## - hip      1 0.0002142 0.024085 -2312.4
## - age      1 0.0003066 0.024178 -2311.4
## - neck     1 0.0004116 0.024283 -2310.3
## - thigh    1 0.0004774 0.024349 -2309.7
## - forearm  1 0.0004804 0.024352 -2309.6
## - weight   1 0.0008772 0.024748 -2305.6
## - wrist    1 0.0010639 0.024935 -2303.7
## - abdomen  1 0.0175263 0.041397 -2175.9
##
## Step:  AIC=-2312.86
## body_density ~ age + weight + neck + abdomen + hip + thigh +
##     bicep + forearm + wrist
##
##           Df Sum of Sq      RSS     AIC
## <none>                 0.024042 -2312.9
## - bicep    1 0.0001923 0.024234 -2312.8
## - hip      1 0.0002267 0.024269 -2312.5
## - age      1 0.0002803 0.024322 -2311.9
## - forearm  1 0.0004752 0.024517 -2309.9
## - neck     1 0.0004938 0.024536 -2309.7
## - thigh    1 0.0005100 0.024552 -2309.6
## - weight   1 0.0007406 0.024782 -2307.2
## - wrist    1 0.0009172 0.024959 -2305.4
## - abdomen  1 0.0173814 0.041423 -2177.8
##
## Call:
## lm(formula = body_density ~ age + weight + neck + abdomen + hip +
##     thigh + bicep + forearm + wrist, data = data)
##
## Coefficients:
## (Intercept)          age       weight         neck      abdomen          hip
##   1.1490232   -0.0001206    0.0002587    0.0011730   -0.0022186    0.0004887
##       thigh        bicep      forearm        wrist
##  -0.0007050   -0.0005451   -0.0009971    0.0036030
```

Forward selection

```
intercept_only_fit <- lm(body_density ~ 1, data=data)
step(intercept_only_fit, direction = 'forward', scope = formula(mult.fit))
```

```
## Start:  AIC=-1995.68
## body_density ~ 1
##
##           Df Sum of Sq      RSS     AIC
## + abdomen  1  0.058031 0.032880 -2250.0
## + chest    1  0.042359 0.048552 -2151.8
## + hip      1  0.033754 0.057157 -2110.6
## + weight   1  0.032083 0.058828 -2103.4
## + thigh    1  0.027811 0.063100 -2085.7
## + knee     1  0.022279 0.068632 -2064.5
## + bicep    1  0.021571 0.069340 -2061.9
```

```
## + neck     1  0.020337 0.070575 -2057.5
## + forearm  1  0.011242 0.079669 -2026.9
## + wrist    1  0.009645 0.081266 -2021.9
## + age      1  0.007008 0.083903 -2013.9
## + ankle    1  0.006379 0.084532 -2012.0
## <none>                 0.090911 -1995.7
## + height   1  0.000140 0.090771 -1994.1
##
## Step:  AIC=-2249.97
## body_density ~ abdomen
##
##           Df Sum of Sq      RSS     AIC
## + weight   1 0.0057258 0.027154 -2296.2
## + wrist    1 0.0042415 0.028638 -2282.8
## + neck     1 0.0035347 0.029345 -2276.6
## + height   1 0.0034367 0.029443 -2275.8
## + hip      1 0.0030517 0.029828 -2272.5
## + knee     1 0.0017569 0.031123 -2261.8
## + chest    1 0.0013594 0.031520 -2258.6
## + ankle    1 0.0010813 0.031799 -2256.4
## + age      1 0.0008402 0.032040 -2254.5
## + thigh    1 0.0007782 0.032102 -2254.0
## + bicep    1 0.0006199 0.032260 -2252.8
## + forearm  1 0.0003102 0.032570 -2250.4
## <none>                 0.032880 -2250.0
##
## Step:  AIC=-2296.18
## body_density ~ abdomen + weight
##
##           Df  Sum of Sq      RSS     AIC
## + wrist    1 0.00100555 0.026149 -2303.7
## + thigh    1 0.00069040 0.026464 -2300.7
## + bicep    1 0.00051011 0.026644 -2299.0
## + neck     1 0.00050946 0.026645 -2299.0
## + forearm  1 0.00038040 0.026774 -2297.7
## <none>                  0.027154 -2296.2
## + age      1 0.00007895 0.027075 -2294.9
## + knee     1 0.00007007 0.027084 -2294.8
## + height   1 0.00005514 0.027099 -2294.7
## + ankle    1 0.00005081 0.027103 -2294.7
## + chest    1 0.00001698 0.027137 -2294.3
## + hip      1 0.00000066 0.027154 -2294.2
##
## Step:  AIC=-2303.69
## body_density ~ abdomen + weight + wrist
##
##           Df  Sum of Sq      RSS     AIC
## + forearm  1 0.00075008 0.025399 -2309.0
## + bicep    1 0.00068754 0.025461 -2308.4
## + thigh    1 0.00038582 0.025763 -2305.4
## <none>                  0.026149 -2303.7
## + ankle    1 0.00019290 0.025956 -2303.6
## + knee     1 0.00014287 0.026006 -2303.1
## + neck     1 0.00013465 0.026014 -2303.0
```

```
## + height   1 0.00004558 0.026103 -2302.1
## + hip      1 0.00004433 0.026104 -2302.1
## + age      1 0.00003201 0.026117 -2302.0
## + chest    1 0.00000110 0.026148 -2301.7
##
## Step:  AIC=-2309.03
## body_density ~ abdomen + weight + wrist + forearm
##
##          Df  Sum of Sq      RSS     AIC
## + bicep   1 3.1451e-04 0.025084 -2310.2
## + thigh   1 2.8539e-04 0.025113 -2309.9
## + neck    1 2.8054e-04 0.025118 -2309.8
## + ankle   1 2.2026e-04 0.025178 -2309.2
## <none>                 0.025399 -2309.0
## + knee    1 1.3829e-04 0.025260 -2308.4
## + age     1 8.9663e-05 0.025309 -2307.9
## + chest   1 3.0326e-05 0.025368 -2307.3
## + height  1 1.6199e-05 0.025382 -2307.2
## + hip     1 1.4807e-05 0.025384 -2307.2
##
## Step:  AIC=-2310.17
## body_density ~ abdomen + weight + wrist + forearm + bicep
##
##          Df  Sum of Sq      RSS     AIC
## + neck    1 0.00036636 0.024718 -2311.9
## + ankle   1 0.00025025 0.024834 -2310.7
## <none>                 0.025084 -2310.2
## + thigh   1 0.00017170 0.024912 -2309.9
## + knee    1 0.00015528 0.024929 -2309.7
## + age     1 0.00009321 0.024991 -2309.1
## + chest   1 0.00004448 0.025040 -2308.6
## + hip     1 0.00001697 0.025067 -2308.3
## + height  1 0.00000000 0.025084 -2308.2
##
## Step:  AIC=-2311.88
## body_density ~ abdomen + weight + wrist + forearm + bicep + neck
##
##          Df  Sum of Sq      RSS     AIC
## <none>                 0.024718 -2311.9
## + ankle   1 1.7676e-04 0.024541 -2311.7
## + thigh   1 1.4301e-04 0.024575 -2311.3
## + age     1 1.3500e-04 0.024583 -2311.3
## + knee    1 8.5366e-05 0.024632 -2310.8
## + hip     1 6.4373e-05 0.024653 -2310.5
## + chest   1 2.2083e-05 0.024696 -2310.1
## + height  1 8.5180e-06 0.024709 -2310.0
##
## Call:
## lm(formula = body_density ~ abdomen + weight + wrist + forearm +
##     bicep + neck, data = data)
##
## Coefficients:
## (Intercept)      abdomen       weight        wrist      forearm        bicep
```

```
##   1.1680224   -0.0023211    0.0003356    0.0031416   -0.0009754   -0.0007586
##        neck
##   0.0009874
```

Stepwise regression

```
step(intercept_only_fit, direction = 'both', scope = formula(mult.fit), data = data)
```

```
## Start:  AIC=-1995.68
## body_density ~ 1
##
##           Df Sum of Sq      RSS     AIC
## + abdomen  1  0.058031 0.032880 -2250.0
## + chest    1  0.042359 0.048552 -2151.8
## + hip      1  0.033754 0.057157 -2110.6
## + weight   1  0.032083 0.058828 -2103.4
## + thigh    1  0.027811 0.063100 -2085.7
## + knee     1  0.022279 0.068632 -2064.5
## + bicep    1  0.021571 0.069340 -2061.9
## + neck     1  0.020337 0.070575 -2057.5
## + forearm  1  0.011242 0.079669 -2026.9
## + wrist    1  0.009645 0.081266 -2021.9
## + age      1  0.007008 0.083903 -2013.9
## + ankle    1  0.006379 0.084532 -2012.0
## <none>                 0.090911 -1995.7
## + height   1  0.000140 0.090771 -1994.1
##
## Step:  AIC=-2249.97
## body_density ~ abdomen
##
##           Df Sum of Sq      RSS     AIC
## + weight   1  0.005726 0.027154 -2296.2
## + wrist    1  0.004242 0.028638 -2282.8
## + neck     1  0.003535 0.029345 -2276.6
## + height   1  0.003437 0.029443 -2275.8
## + hip      1  0.003052 0.029828 -2272.5
## + knee     1  0.001757 0.031123 -2261.8
## + chest    1  0.001359 0.031520 -2258.6
## + ankle    1  0.001081 0.031799 -2256.4
## + age      1  0.000840 0.032040 -2254.5
## + thigh    1  0.000778 0.032102 -2254.0
## + bicep    1  0.000620 0.032260 -2252.8
## + forearm  1  0.000310 0.032570 -2250.4
## <none>                 0.032880 -2250.0
## - abdomen  1  0.058031 0.090911 -1995.7
##
## Step:  AIC=-2296.18
## body_density ~ abdomen + weight
##
##           Df Sum of Sq      RSS     AIC
## + wrist    1  0.001006 0.026149 -2303.7
## + thigh    1  0.000690 0.026464 -2300.7
## + bicep    1  0.000510 0.026644 -2299.0
```

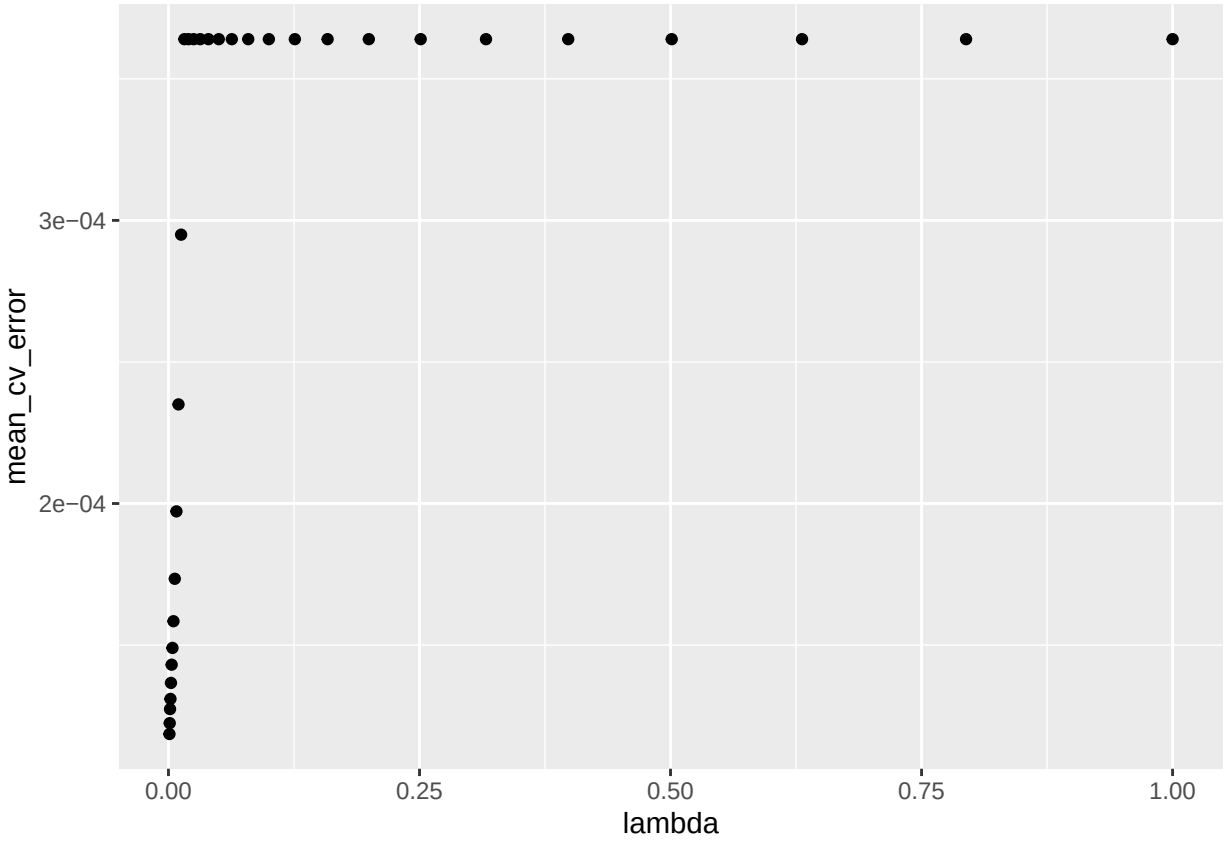
```
## + neck     1  0.000509 0.026645 -2299.0
## + forearm  1  0.000380 0.026774 -2297.7
## <none>                 0.027154 -2296.2
## + age      1  0.000079 0.027075 -2294.9
## + knee     1  0.000070 0.027084 -2294.8
## + height   1  0.000055 0.027099 -2294.7
## + ankle    1  0.000051 0.027103 -2294.7
## + chest    1  0.000017 0.027137 -2294.3
## + hip      1  0.000001 0.027154 -2294.2
## - weight   1  0.005726 0.032880 -2250.0
## - abdomen  1  0.031674 0.058828 -2103.4
##
## Step:  AIC=-2303.69
## body_density ~ abdomen + weight + wrist
##
##           Df Sum of Sq      RSS     AIC
## + forearm  1 0.0007501 0.025399 -2309.0
## + bicep    1 0.0006875 0.025461 -2308.4
## + thigh    1 0.0003858 0.025763 -2305.4
## <none>                 0.026149 -2303.7
## + ankle    1 0.0001929 0.025956 -2303.6
## + knee     1 0.0001429 0.026006 -2303.1
## + neck     1 0.0001346 0.026014 -2303.0
## + height   1 0.0000456 0.026103 -2302.1
## + hip      1 0.0000443 0.026104 -2302.1
## + age      1 0.0000320 0.026117 -2302.0
## + chest    1 0.0000011 0.026148 -2301.7
## - wrist    1 0.0010055 0.027154 -2296.2
## - weight   1 0.0024898 0.028638 -2282.8
## - abdomen  1 0.0304182 0.056567 -2111.2
##
## Step:  AIC=-2309.03
## body_density ~ abdomen + weight + wrist + forearm
##
##           Df Sum of Sq      RSS     AIC
## + bicep    1 0.0003145 0.025084 -2310.2
## + thigh    1 0.0002854 0.025113 -2309.9
## + neck     1 0.0002805 0.025118 -2309.8
## + ankle    1 0.0002203 0.025178 -2309.2
## <none>                 0.025399 -2309.0
## + knee     1 0.0001383 0.025260 -2308.4
## + age      1 0.0000897 0.025309 -2307.9
## + chest    1 0.0000303 0.025368 -2307.3
## + height   1 0.0000162 0.025382 -2307.2
## + hip      1 0.0000148 0.025384 -2307.2
## - forearm  1 0.0007501 0.026149 -2303.7
## - wrist    1 0.0013752 0.026774 -2297.7
## - weight   1 0.0031502 0.028549 -2281.6
## - abdomen  1 0.0311622 0.056561 -2109.3
##
## Step:  AIC=-2310.17
## body_density ~ abdomen + weight + wrist + forearm + bicep
##
##           Df Sum of Sq      RSS     AIC
```

```
## + neck     1 0.0003664 0.024718 -2311.9
## + ankle    1 0.0002503 0.024834 -2310.7
## <none>                 0.025084 -2310.2
## + thigh    1 0.0001717 0.024912 -2309.9
## + knee     1 0.0001553 0.024929 -2309.7
## + age      1 0.0000932 0.024991 -2309.1
## - bicep    1 0.0003145 0.025399 -2309.0
## + chest    1 0.0000445 0.025040 -2308.6
## - forearm  1 0.0003770 0.025461 -2308.4
## + hip      1 0.0000170 0.025067 -2308.3
## + height   1 0.0000000 0.025084 -2308.2
## - wrist    1 0.0014136 0.026498 -2298.3
## - weight   1 0.0034253 0.028509 -2279.9
## - abdomen  1 0.0313477 0.056432 -2107.8
##
## Step:  AIC=-2311.88
## body_density ~ abdomen + weight + wrist + forearm + bicep + neck
##
##           Df Sum of Sq      RSS     AIC
## <none>                 0.024718 -2311.9
## + ankle    1  0.000177 0.024541 -2311.7
## + thigh    1  0.000143 0.024575 -2311.3
## + age      1  0.000135 0.024583 -2311.3
## + knee     1  0.000085 0.024632 -2310.8
## + hip      1  0.000064 0.024653 -2310.5
## - neck     1  0.000366 0.025084 -2310.2
## + chest    1  0.000022 0.024696 -2310.1
## + height   1  0.000009 0.024709 -2310.0
## - bicep    1  0.000400 0.025118 -2309.8
## - forearm  1  0.000469 0.025187 -2309.1
## - wrist    1  0.000843 0.025561 -2305.4
## - weight   1  0.002751 0.027468 -2287.3
## - abdomen  1  0.031689 0.056407 -2106.0
##
## Call:
## lm(formula = body_density ~ abdomen + weight + wrist + forearm +
##     bicep + neck, data = data)
##
## Coefficients:
## (Intercept)      abdomen       weight        wrist      forearm        bicep
##   1.1680224   -0.0023211    0.0003356    0.0031416   -0.0009754   -0.0007586
##        neck
##   0.0009874
```

Lasso

```
set.seed(200)
lamba_seq <- 10^seq(-3, 0,by = .1)
train_predictors <- data %>% select(-body_density)
cv_object <- cv.glmnet(as.matrix(train_predictors), data$body_density,
                   lambda = lamba_seq, nfolds = 10)
```

```
tibble(lambda  = cv_object$lambda,
       mean_cv_error = cv_object$cvm) %>%
  ggplot(aes(x = lambda, y = mean_cv_error)) +
  geom_point()
```




```
cv_object$lambda.min
```

```
## [1] 0.001
```

```
lasso_fit <- glmnet(as.matrix(train_predictors),
                    data$body_density,
                    lambda = cv_object$lambda.min)
coef(lasso_fit)
```

```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                         s0
## (Intercept)  1.106666e+00
## age         -5.216849e-05
## weight       .
## height       6.873219e-04
## neck         .
## chest        .
## abdomen     -1.441851e-03
## hip          .
## thigh        .
## knee         .
## ankle        .
## bicep        .
```

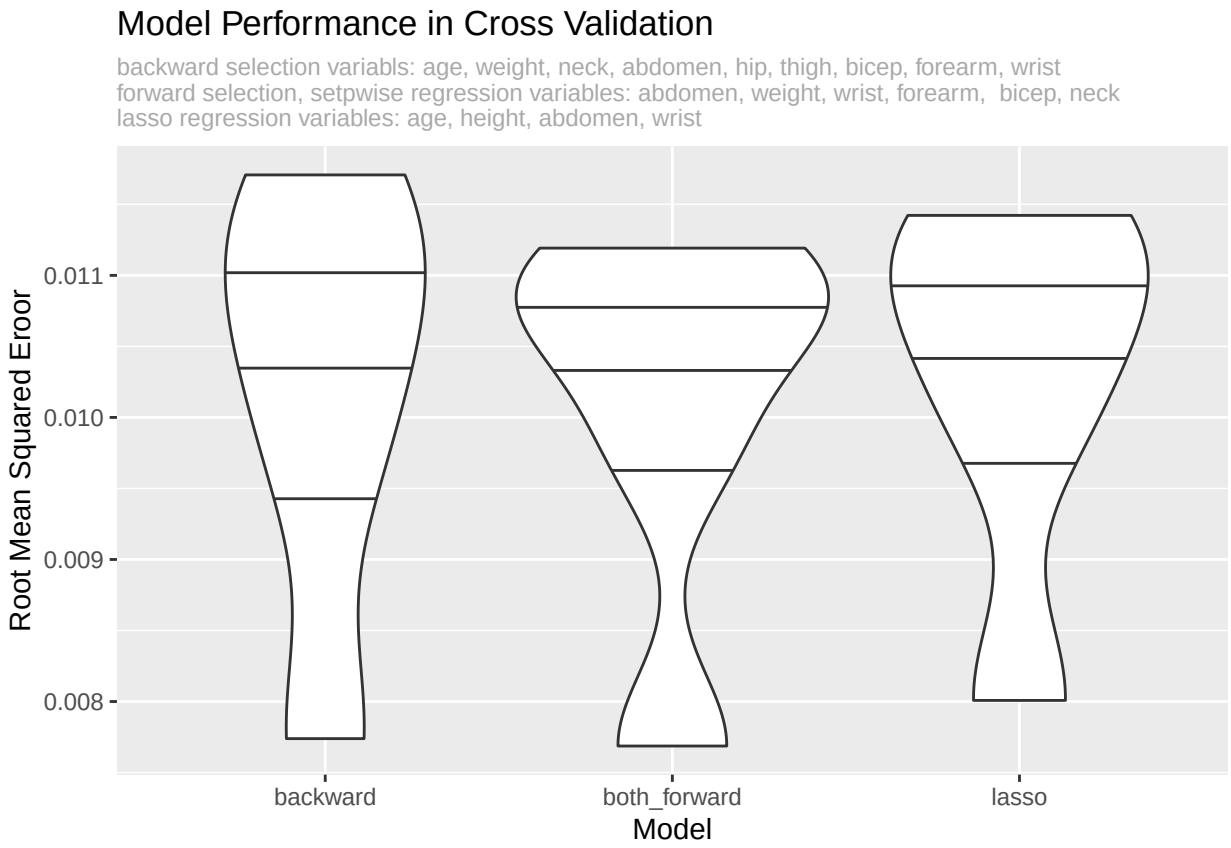
```
## forearm      .
## wrist        1.995518e-03
```

Cross Validation

```
cv_df =
  crossv_mc(data, 5) %>%
  mutate(
    train = map(train, as_tibble),
    test = map(test, as_tibble))

cv_df =
  cv_df %>%
  mutate(
    backward_mod  = map(train, ~lm(body_density ~ age + weight + neck + abdomen + hip + thigh + bicep +
    both_forward_mod  = map(train, ~lm(body_density ~ abdomen + weight + wrist + forearm + bicep + neck
    lasso_mod  = map(train, ~lm(body_density ~ age + height + abdomen + wrist, data = as_tibble(.x)))) %
  mutate(
    rmse_backward = map2_dbl(backward_mod, test, ~rmse(model = .x, data = .y)),
    rmse_both_forward  = map2_dbl(both_forward_mod, test, ~rmse(model = .x, data = .y)),
    rmse_lasso = map2_dbl(lasso_mod, test, ~rmse(model = .x, data = .y)))
```

```
cv_df %>%
  select(starts_with("rmse")) %>%
  pivot_longer(
    everything(),
    names_to = "model",
    values_to = "rmse",
    names_prefix = "rmse_") %>%
  mutate(model = fct_inorder(model)) %>%
  ggplot(aes(x = model, y = rmse)) + geom_violin(draw_quantiles = c(0.25, 0.5, 0.75)) +
  ggtitle("Model Performance in Cross Validation",
          subtitle = "backward selection variabls: age, weight, neck, abdomen, hip, thigh, bicep, forea
  ylab("Root Mean Squared Eroor") +
  xlab("Model") +
  theme(plot.subtitle=element_text(size=9, color = "darkgray"))
```




```
rmse <- cv_df %>%
  select(starts_with("rmse"))

sapply(rmse, mean, na.rm = T) %>%
  knitr::kable(digits = 6, col.names = "mean",
               caption = "Cross Validation Mean RMSE.")
```

Table 1: Cross Validation Mean RMSE.

	mean
rmse_backward	0.010180
rmse_both_forward	0.010045
rmse_lasso	0.010249

Compare adjusted R^2

Backward selection model

```
lm(body_density ~ age + weight + neck + abdomen + hip + thigh + bicep + forearm + wrist, data = data)  %

##
## Call:
## lm(formula = body_density ~ age + weight + neck + abdomen + hip +
##     thigh + bicep + forearm + wrist, data = data)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.021639 -0.007151 -0.000122  0.006605  0.036365
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.149e+00  2.730e-02  42.082  < 2e-16 ***
## age         -1.206e-04  7.181e-05  -1.680  0.09429 .
## weight       2.587e-04  9.477e-05   2.730  0.00679 **
## neck         1.173e-03  5.261e-04   2.229  0.02670 *
## abdomen     -2.219e-03  1.677e-04 -13.227  < 2e-16 ***
## hip          4.887e-04  3.235e-04   1.511  0.13219
## thigh       -7.050e-04  3.111e-04  -2.266  0.02435 *
## bicep       -5.451e-04  3.918e-04  -1.391  0.16540
## forearm     -9.972e-04  4.559e-04  -2.187  0.02969 *
## wrist        3.603e-03  1.186e-03   3.038  0.00264 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009967 on 242 degrees of freedom
## Multiple R-squared:  0.7355, Adjusted R-squared:  0.7257
## F-statistic: 74.79 on 9 and 242 DF,  p-value: < 2.2e-16
```

Forward selection, stepwise regression model

```
lm(body_density ~ abdomen + weight + wrist + forearm + bicep + neck, data = data) %>% summary()
```

```
##
## Call:
## lm(formula = body_density ~ abdomen + weight + wrist + forearm +
##     bicep + neck, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.021082 -0.007581  0.000447  0.006907  0.036327
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.168e+00  1.792e-02  65.180  < 2e-16 ***
## abdomen     -2.321e-03  1.310e-04 -17.723  < 2e-16 ***
## weight       3.355e-04  6.426e-05   5.221 3.79e-07 ***
## wrist        3.142e-03  1.087e-03   2.890  0.00419 **
## forearm     -9.754e-04  4.523e-04  -2.157  0.03202 *
## bicep       -7.586e-04  3.808e-04  -1.992  0.04748 *
## neck         9.874e-04  5.181e-04   1.906  0.05787 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01004 on 245 degrees of freedom
## Multiple R-squared:  0.7281, Adjusted R-squared:  0.7215
## F-statistic: 109.4 on 6 and 245 DF,  p-value: < 2.2e-16
```

Lasso

```
lm(body_density ~ age + height + abdomen + wrist, data = data) %>% summary()
```

```
##
## Call:
## lm(formula = body_density ~ age + height + abdomen + wrist, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.020823 -0.007395 -0.000003  0.006912  0.040711
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.075e+00  1.958e-02  54.872  < 2e-16 ***
## age         -1.284e-04  5.766e-05  -2.227   0.0269 *
## height       7.539e-04  2.950e-04   2.556   0.0112 *
## abdomen     -1.670e-03  7.824e-05 -21.352  < 2e-16 ***
## wrist        4.851e-03  9.876e-04   4.913 1.64e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0104 on 247 degrees of freedom
## Multiple R-squared:  0.7064, Adjusted R-squared:  0.7016
## F-statistic: 148.5 on 4 and 247 DF,  p-value: < 2.2e-16
```

final model: both + forward

Final model

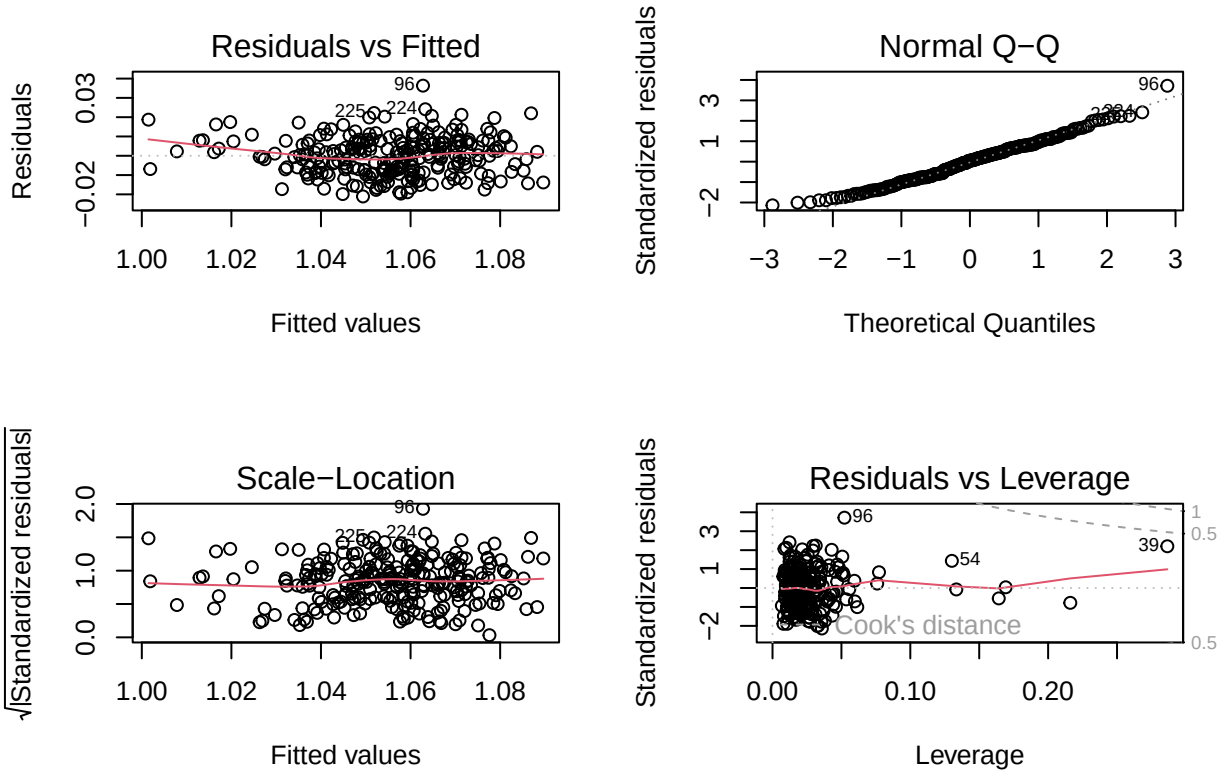
```
final <- lm(body_density ~ abdomen + weight + wrist + forearm + bicep + neck, data = data)
```

```
final %>%
  summary() %>%
  tidy() %>%
  knitr::kable(digits = 4,
               caption = "Linear Model Result")
```

Table 2: Linear Model Result

term	estimate	std.error	statistic	p.value
(Intercept)	1.1680	0.0179	65.1800	0.0000
abdomen	-0.0023	0.0001	-17.7229	0.0000
weight	0.0003	0.0001	5.2215	0.0000
wrist	0.0031	0.0011	2.8904	0.0042
forearm	-0.0010	0.0005	-2.1565	0.0320
bicep	-0.0008	0.0004	-1.9920	0.0475
neck	0.0010	0.0005	1.9056	0.0579

```
par(mfrow = c(2,2))
plot(final)
```




```
mean((data$body_density - predict.lm(final, data)) ^ 2)
```

```
## [1] 9.808605e-05
```