664project3

Peter (Tianzhu) Liu

2022-11-25

Data Cleaning

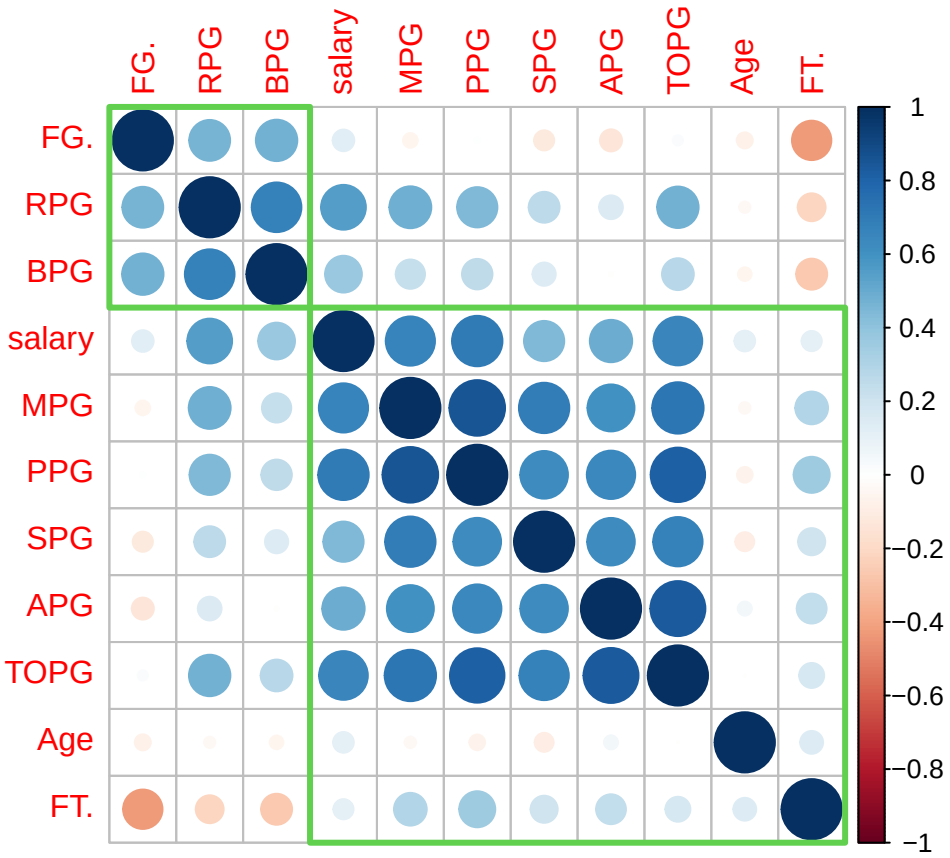
```
data <- read.csv(file = 'final_dataset2.csv')
data = data[data[,"G"] >= 40,]      #screen out less statistically significant data points
data = data[data[,"Age"] >= 24,]    #salary cap
data_reduced = data[,c('Player', 'Pos', 'salary','Age','FG.','FT.','MPG','PPG','APG','RPG','TOPG','BPG'
```

Correlation

```
library(corrplot)
```

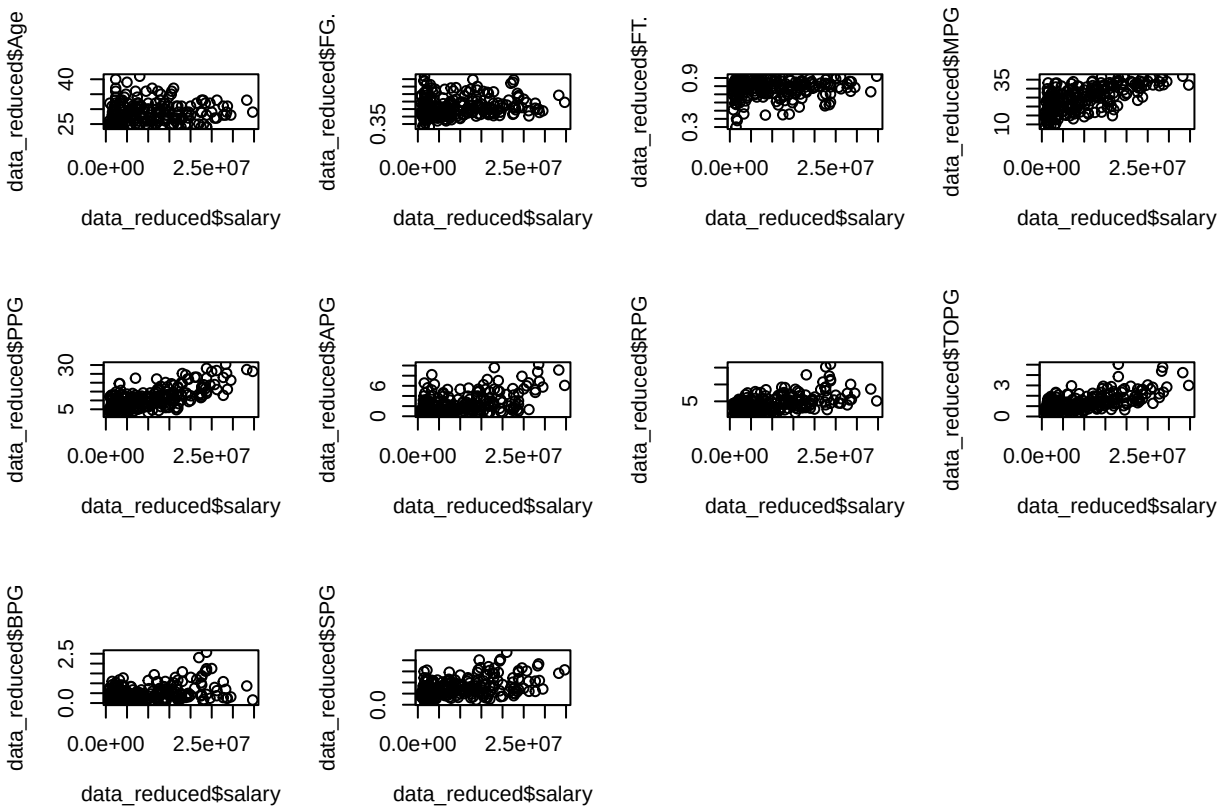
```
## corrplot 0.92 loaded
```

```
corrplot(cor(data_reduced[,c(3:13)]),
         method = "circle",
         order = "hclust",         # Ordering method of the matrix
         hclust.method = "ward.D", # If order = "hclust", is the cluster method to be used
         addrect = 2,              # If order = "hclust", number of cluster rectangles
         rect.col = 3,             # Color of the rectangles
         rect.lwd = 3)
```

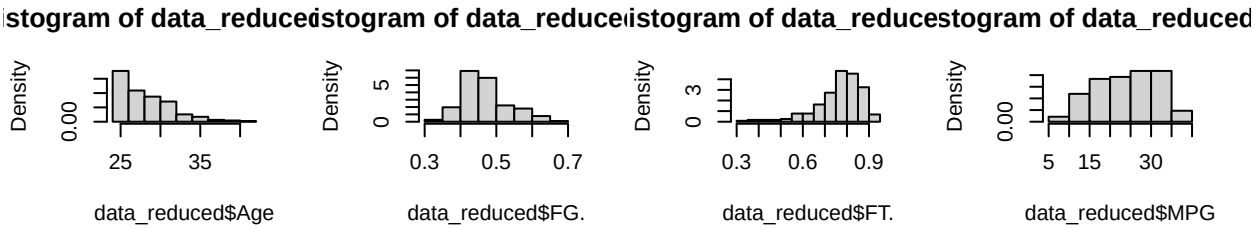


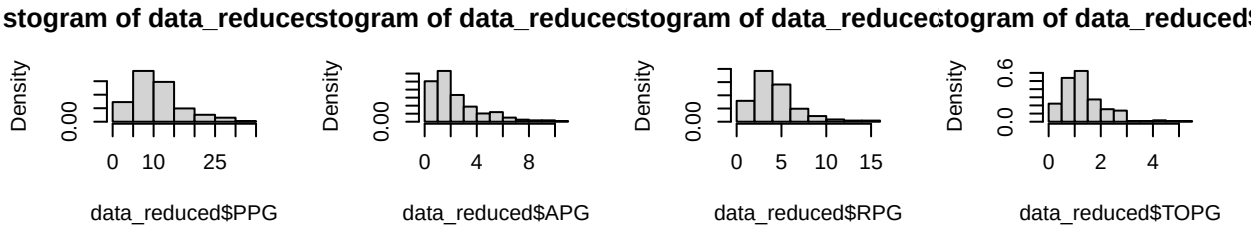
EDA

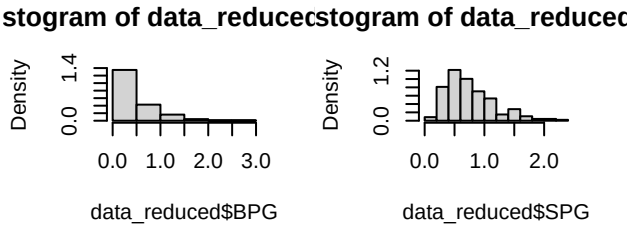
```
par(mfrow=c(3,4))
plot(data_reduced$salary, data_reduced$Age)
plot(data_reduced$salary, data_reduced$FG.)
plot(data_reduced$salary, data_reduced$FT.)
plot(data_reduced$salary, data_reduced$MPG)
plot(data_reduced$salary, data_reduced$PPG)
plot(data_reduced$salary, data_reduced$APG)
plot(data_reduced$salary, data_reduced$RPG)
plot(data_reduced$salary, data_reduced$TOPG)
plot(data_reduced$salary, data_reduced$BPG)
plot(data_reduced$salary, data_reduced$SPG)
```



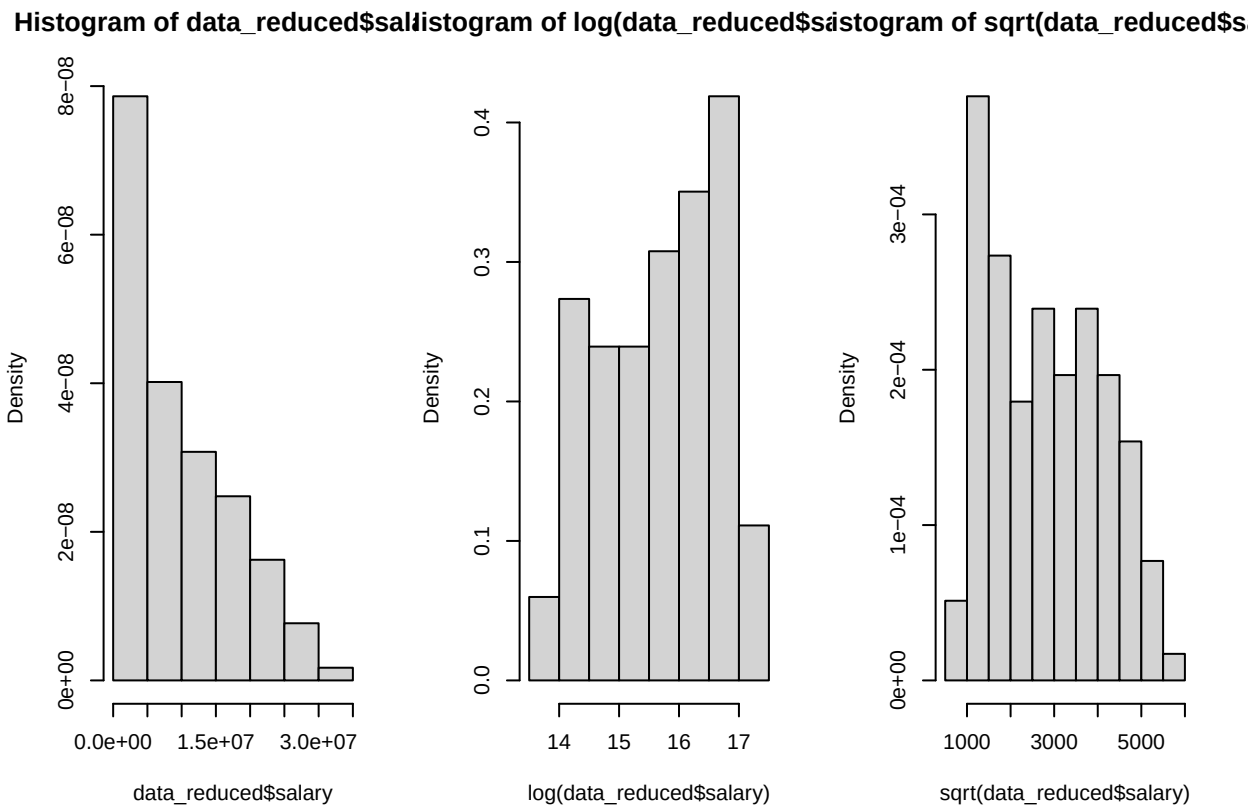

```
par(mfrow=c(3,4))
hist(data_reduced$Age,freq = F)
hist(data_reduced$FG.,freq = F)
hist(data_reduced$FT.,freq = F)
hist(data_reduced$MPG,freq = F)
hist(data_reduced$PPG,freq = F)
hist(data_reduced$APG,freq = F)
hist(data_reduced$RPG,freq = F)
hist(data_reduced$TOPG,freq = F)
hist(data_reduced$BPG,freq = F)
hist(data_reduced$SPG,freq = F)
```








```
par(mfrow=c(1,3))
hist(data_reduced$salary,freq = F)
hist(log(data_reduced$salary),freq = F)
hist(sqrt(data_reduced$salary),freq = F)
```

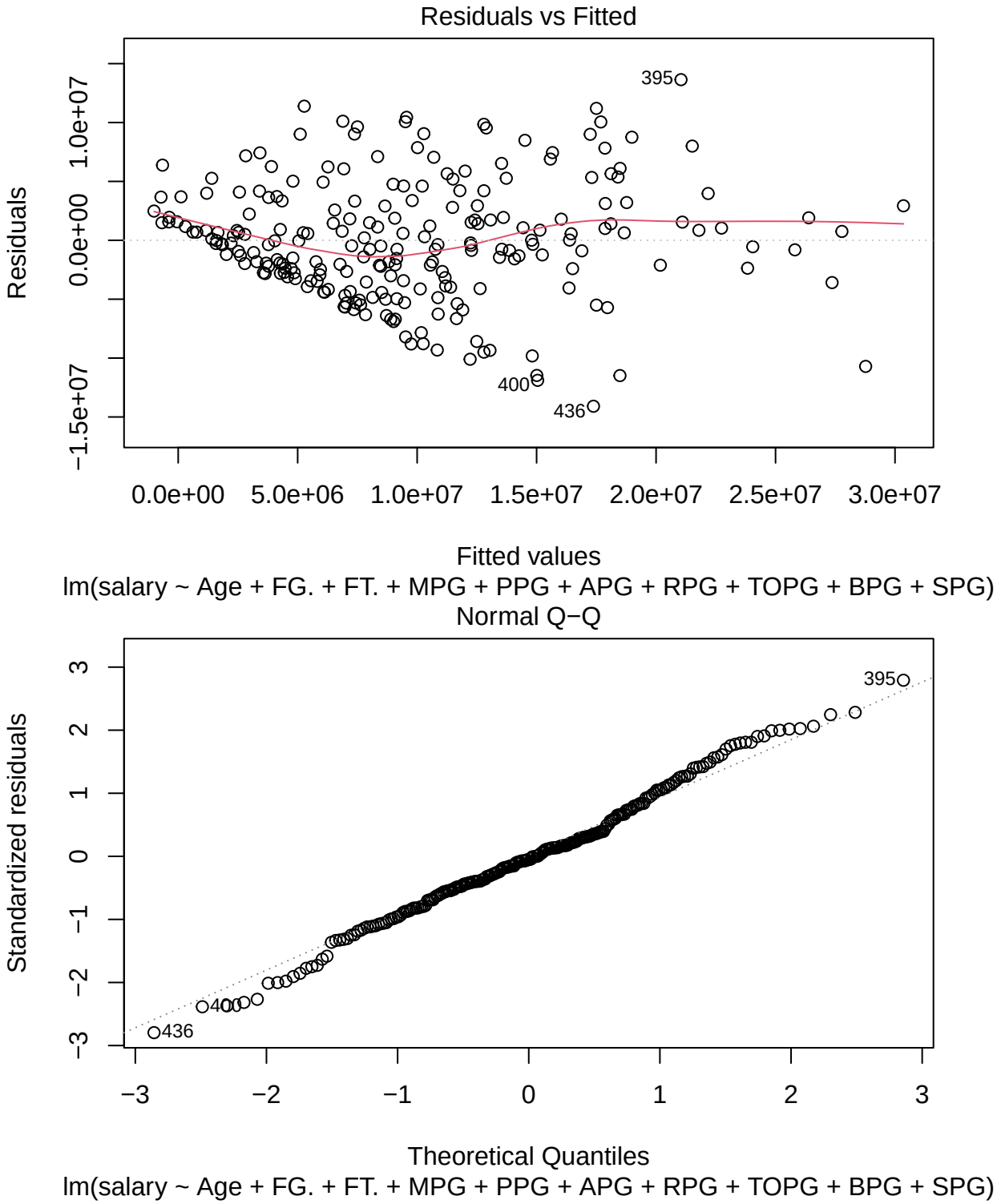


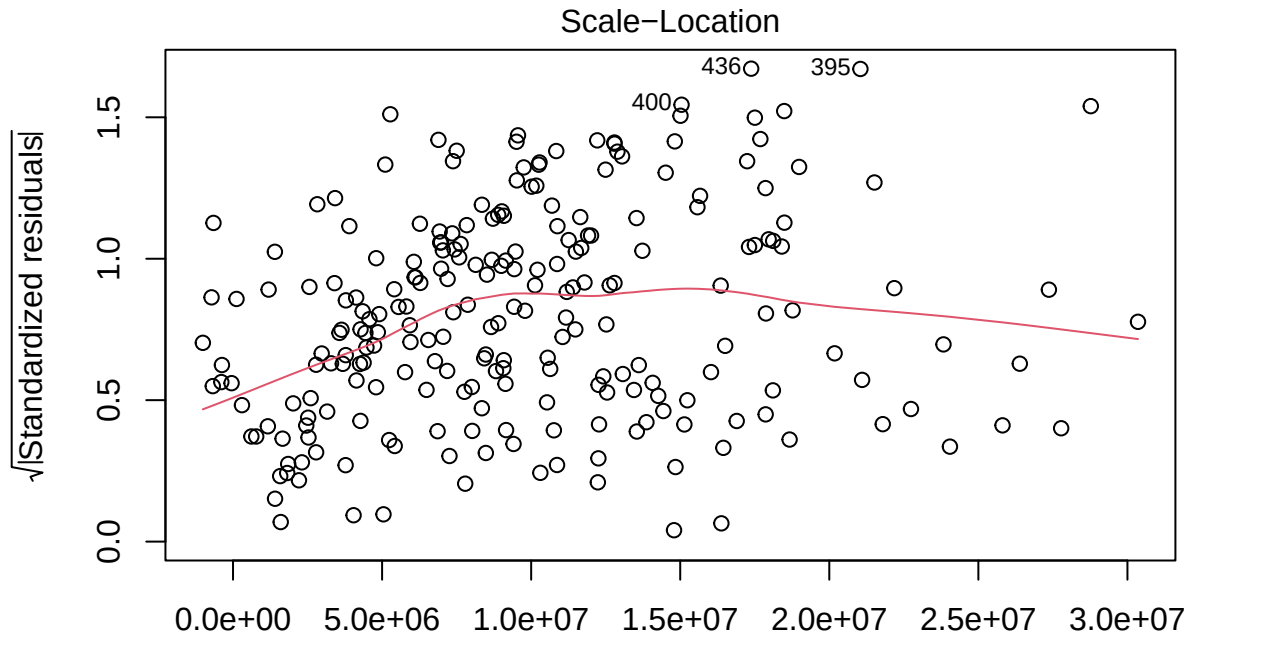

```
model1 = lm(salary~Age+FG.+FT.+MPG+PPG+APG+RPG+TOPG+BPG+SPG, data = data_reduced)
summary(model1)
```

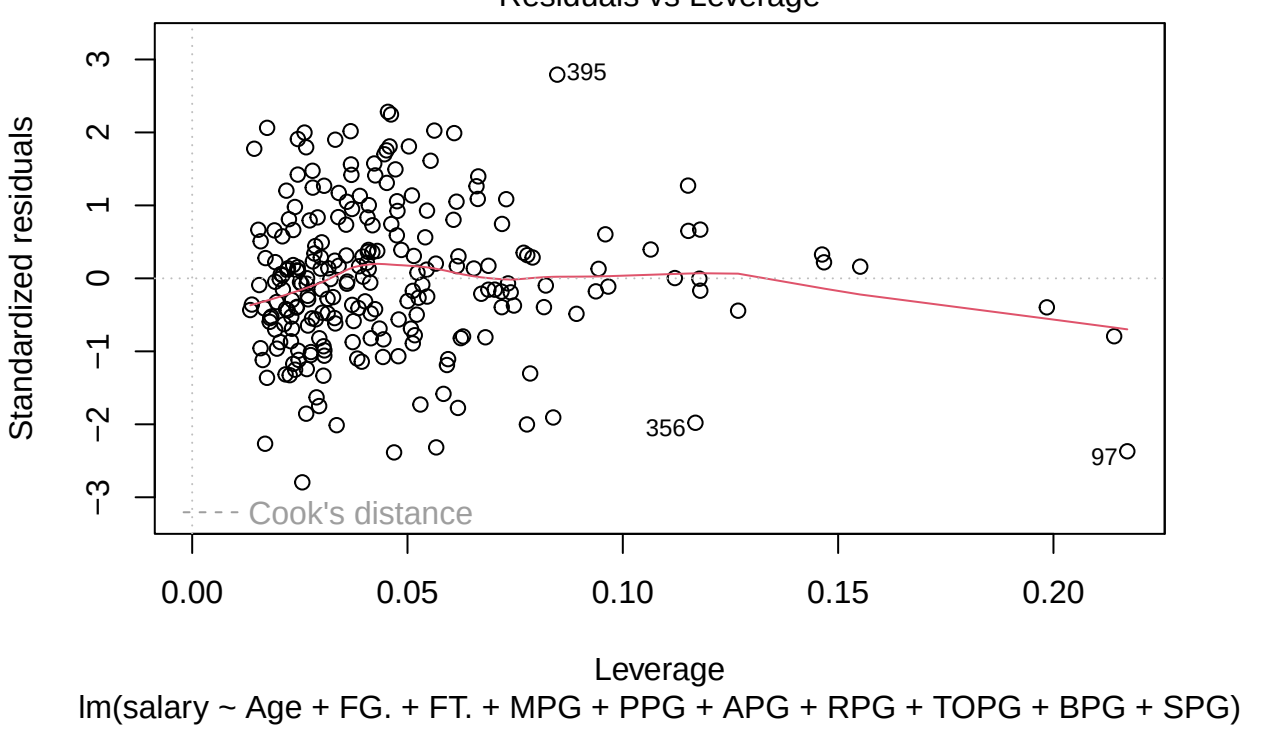
```
##
## Call:
## lm(formula = salary ~ Age + FG. + FT. + MPG + PPG + APG + RPG +
##     TOPG + BPG + SPG, data = data_reduced)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -14089723  -2994727   -253863   3077857  13638501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10742788    5588191  -1.922 0.055828 .
## Age            374641     101166   3.703 0.000268 ***
## FG.          -1376158    6705565  -0.205 0.837583
## FT.          -5556879    4163071  -1.335 0.183302
## MPG            178543     104891   1.702 0.090117 .
## PPG            595949     146358   4.072 6.48e-05 ***
## APG            464214     380792   1.219 0.224103
## RPG            688235     243074   2.831 0.005059 **
## TOPG           180454    1206349   0.150 0.881226
## BPG           1512513    1184947   1.276 0.203128
## SPG          -1172602    1282839  -0.914 0.361669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5106000 on 223 degrees of freedom
## Multiple R-squared:  0.6099, Adjusted R-squared:  0.5924
## F-statistic: 34.87 on 10 and 223 DF,  p-value: < 2.2e-16
```

```
plot(model1)
```





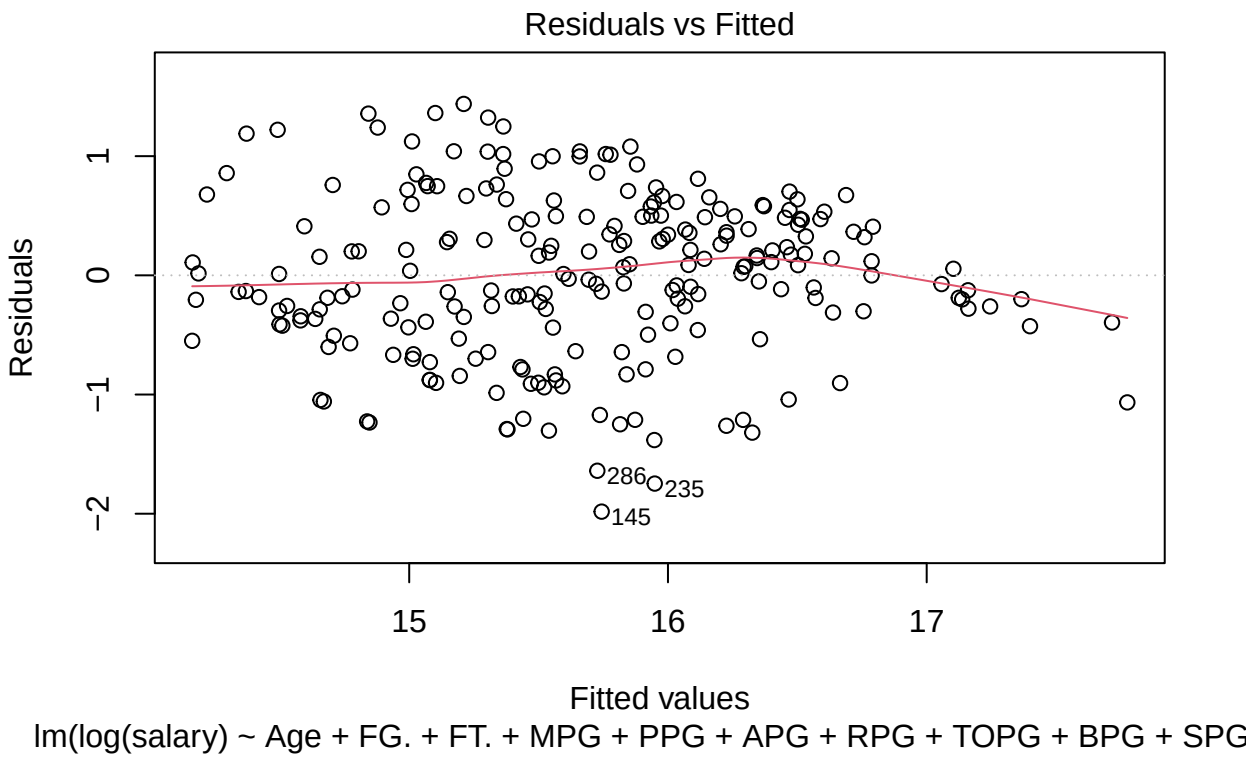


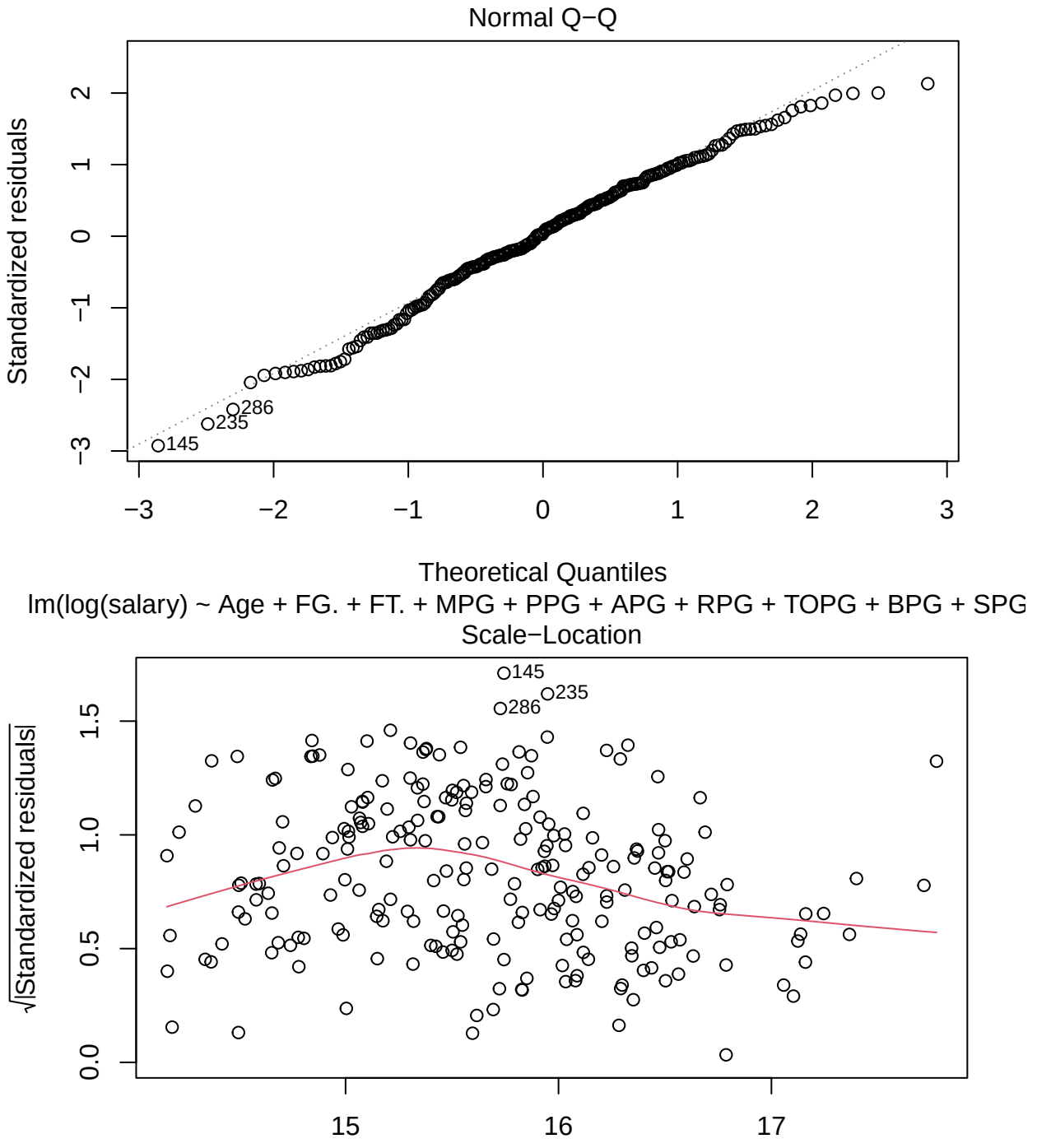

```
model2 = lm(log(salary)~ Age+FG.+FT.+MPG+PPG+APG+RPG+TOPG+BPG+SPG, data = data_reduced)
summary(model2)
```

```
##
## Call:
## lm(formula = log(salary) ~ Age + FG. + FT. + MPG + PPG + APG +
##     RPG + TOPG + BPG + SPG, data = data_reduced)
```

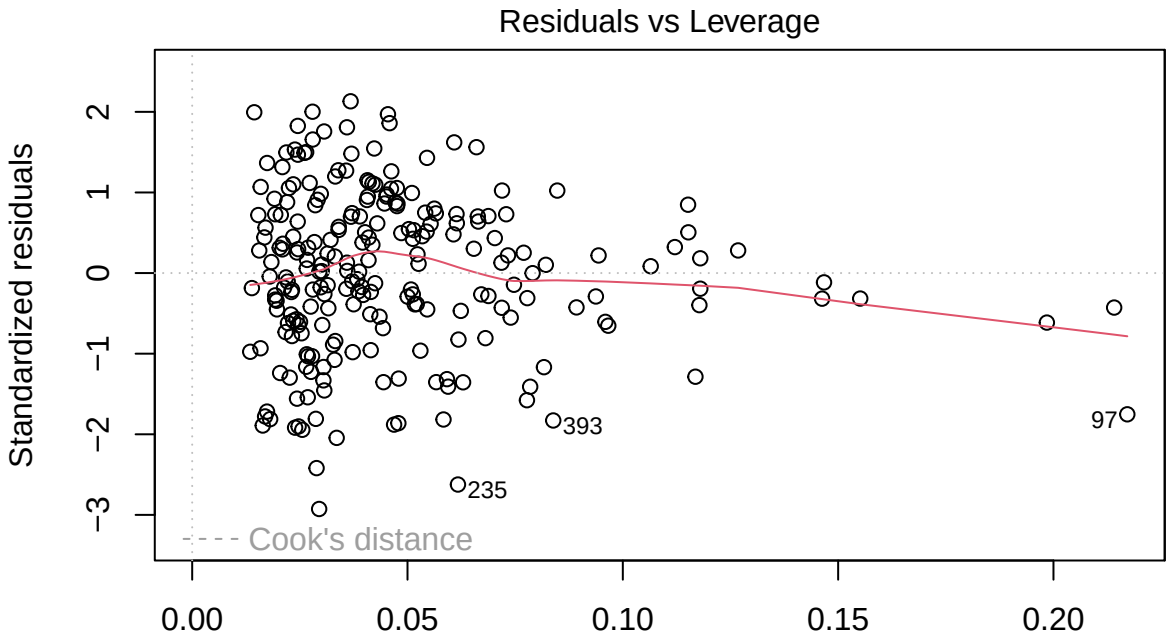
```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.98249 -0.40067  0.02802  0.49012  1.43804
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.53557    0.75246  15.331  < 2e-16 ***
## Age          0.06899    0.01362   5.065 8.57e-07 ***
## FG.          0.52017    0.90291   0.576   0.5651
## FT.         -0.17224    0.56056  -0.307   0.7589
## MPG          0.05889    0.01412   4.170 4.36e-05 ***
## PPG          0.02460    0.01971   1.248   0.2132
## APG         -0.04333    0.05127  -0.845   0.3990
## RPG          0.05911    0.03273   1.806   0.0723 .
## TOPG         0.18835    0.16244   1.160   0.2475
## BPG          0.08460    0.15955   0.530   0.5965
## SPG         -0.06184    0.17274  -0.358   0.7207
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6876 on 223 degrees of freedom
## Multiple R-squared:  0.5475, Adjusted R-squared:  0.5272
## F-statistic: 26.98 on 10 and 223 DF,  p-value: < 2.2e-16
```

```
plot(model2)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(salary) ~ Age + FG. + FT. + MPG + PPG + APG + RPG + TOPG + BPG + SPG
145
235
286

Scale−Location
√|Standardized residuals|
Fitted values
lm(log(salary) ~ Age + FG. + FT. + MPG + PPG + APG + RPG + TOPG + BPG + SPG
145
235
286



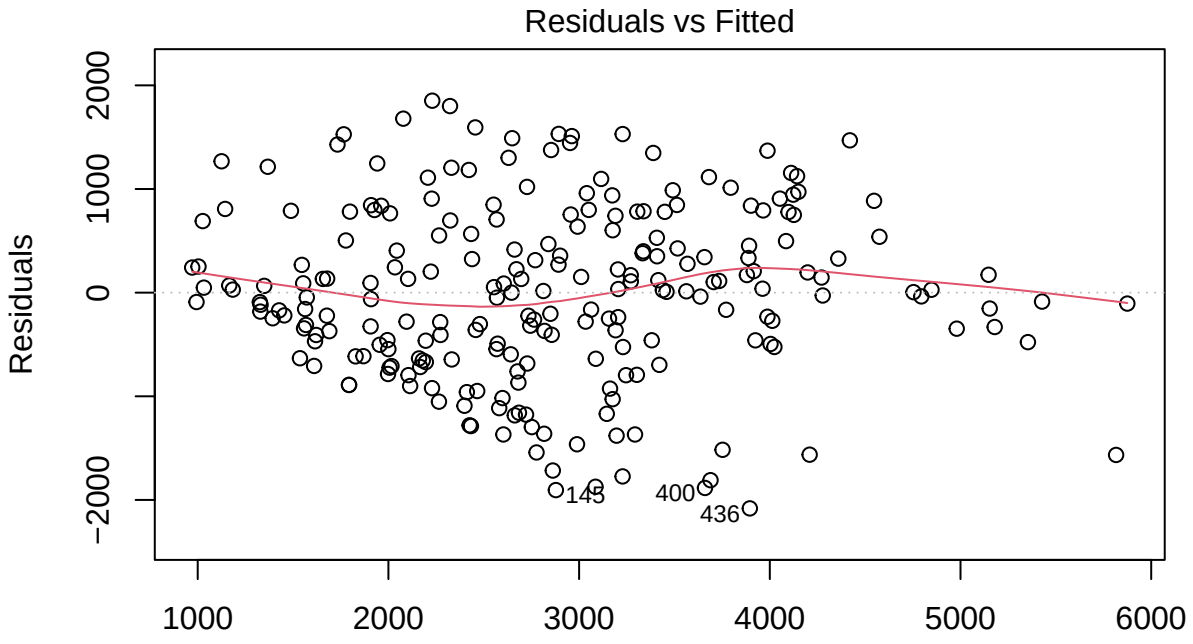
lm(log(salary) ~ Age + FG. + FT. + MPG + PPG + APG + RPG + TOPG + BPG + SPG

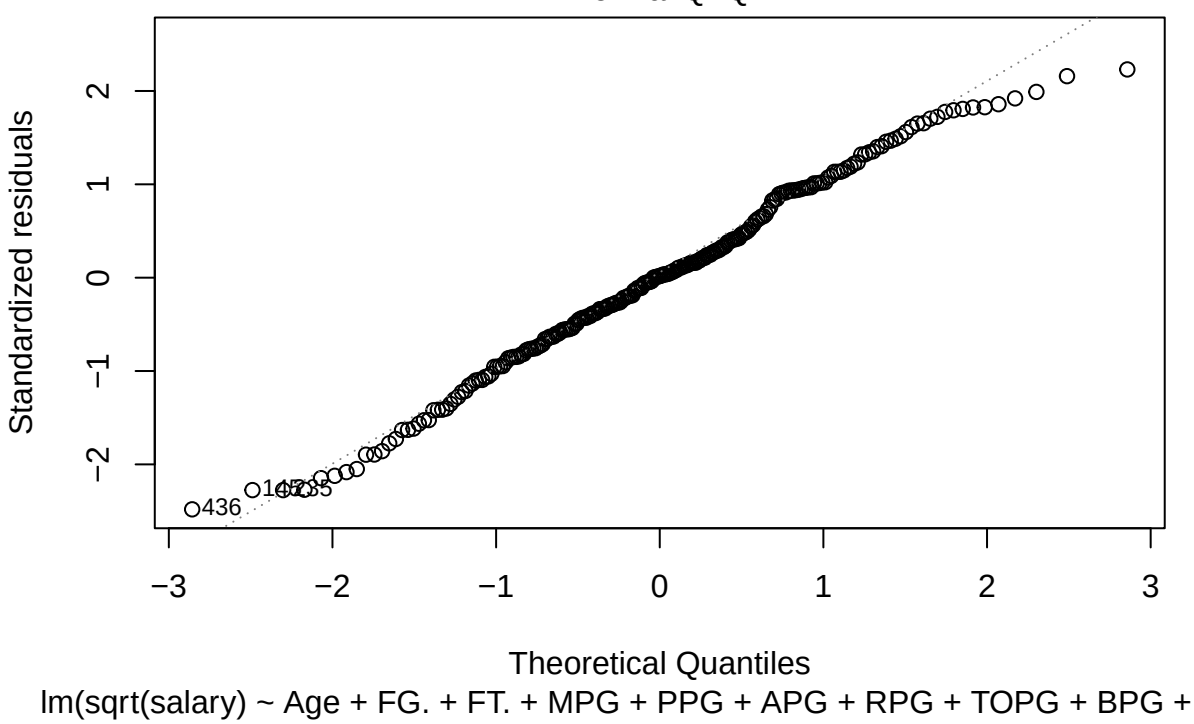
```
model3 = lm(sqrt(salary)~Age+FG.+FT.+MPG+PPG+APG+RPG+TOPG+BPG+SPG, data = data_reduced)
summary(model3)
```

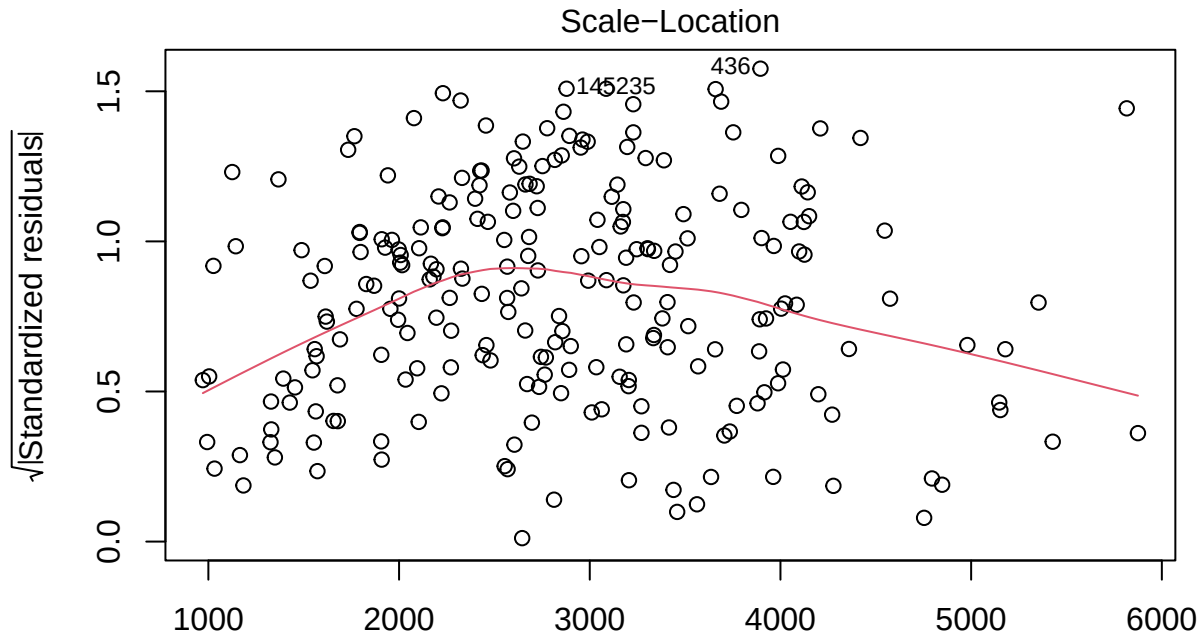
```
##
## Call:
## lm(formula = sqrt(salary) ~ Age + FG. + FT. + MPG + PPG + APG +
##     RPG + TOPG + BPG + SPG, data = data_reduced)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2081.39  -524.60    14.38   628.16  1851.96
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1495.535    929.840  -1.608  0.10917
## Age            74.377     16.833   4.418 1.55e-05 ***
## FG.           224.765   1115.764   0.201  0.84053
## FT.          -624.707    692.709  -0.902  0.36812
## MPG            55.008     17.453   3.152  0.00185 **
## PPG            64.437     24.353   2.646  0.00873 **
## APG             5.909     63.361   0.093  0.92578
## RPG            93.739     40.446   2.318  0.02138 *
## TOPG          147.406    200.729   0.734  0.46351
## BPG           193.064    197.168   0.979  0.32855
## SPG          -143.745    213.456  -0.673  0.50138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 849.7 on 223 degrees of freedom
## Multiple R-squared:  0.5892, Adjusted R-squared:  0.5708
```

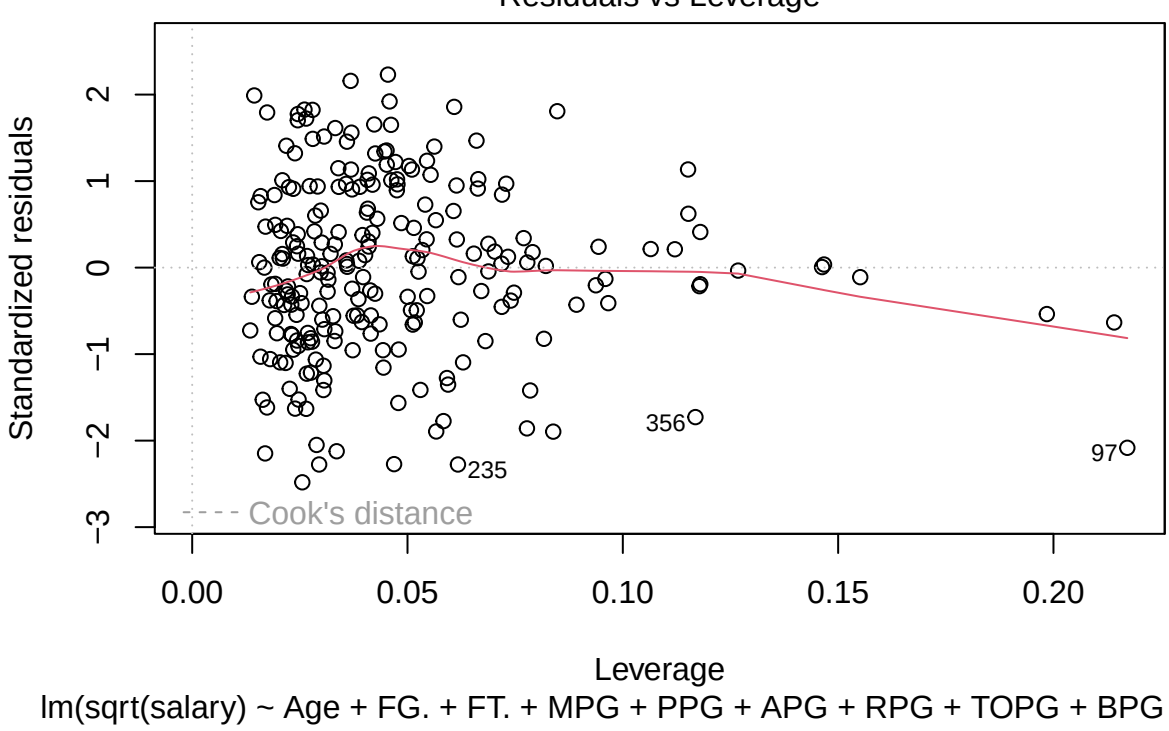
```
## F-statistic: 31.99 on 10 and 223 DF,  p-value: < 2.2e-16
```

```
plot(model3)
```



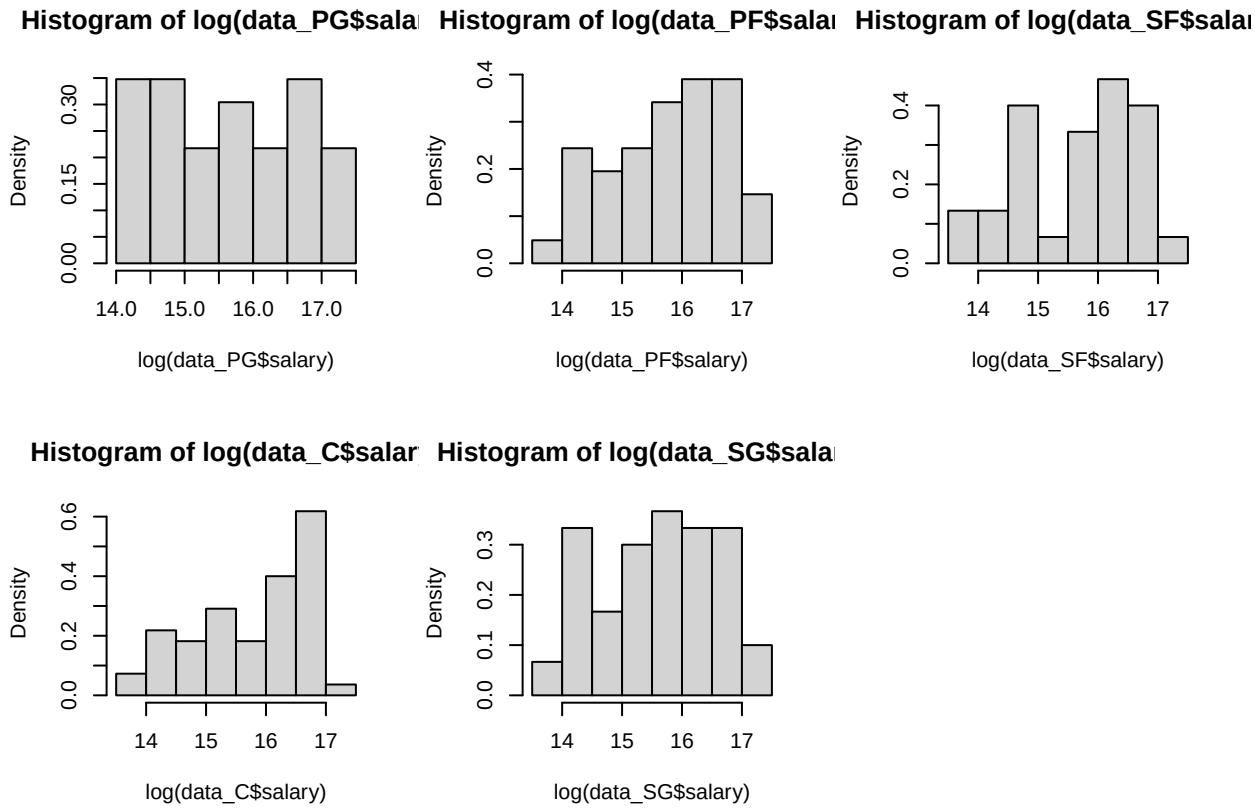






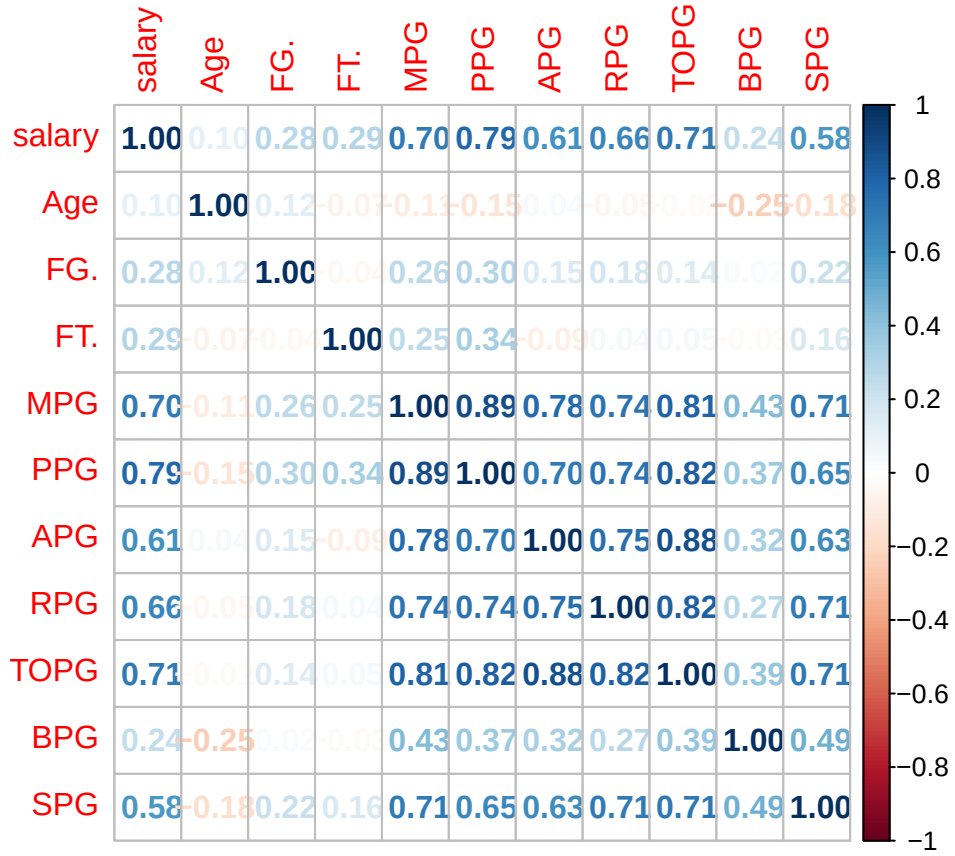

```
data_PG =  data_reduced[data_reduced[,2] == "PG", ]
data_PF =  data_reduced[data_reduced[,2] == "PF", ]
data_SF =  data_reduced[data_reduced[,2] == "SF", ]
data_C = data_reduced[data_reduced[,2] == "C", ]
data_SG = data_reduced[data_reduced[,2] == "SG", ]
```

```
par(mfrow=c(2,3))
hist(log(data_PG$salary),freq = F)
hist(log(data_PF$salary),freq = F)
hist(log(data_SF$salary),freq = F)
hist(log(data_C$salary),freq = F)
hist(log(data_SG$salary),freq = F)
```

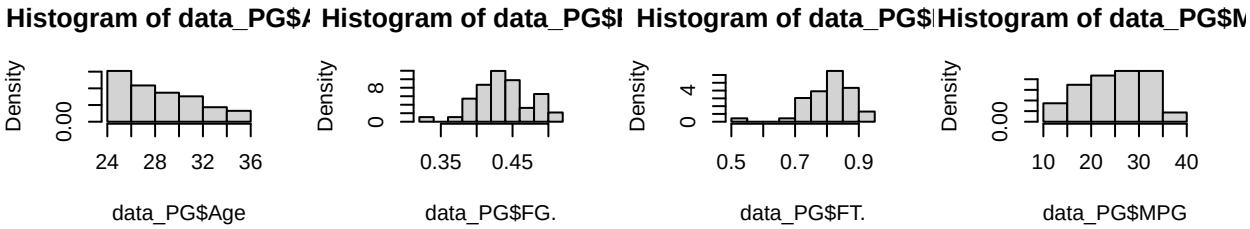


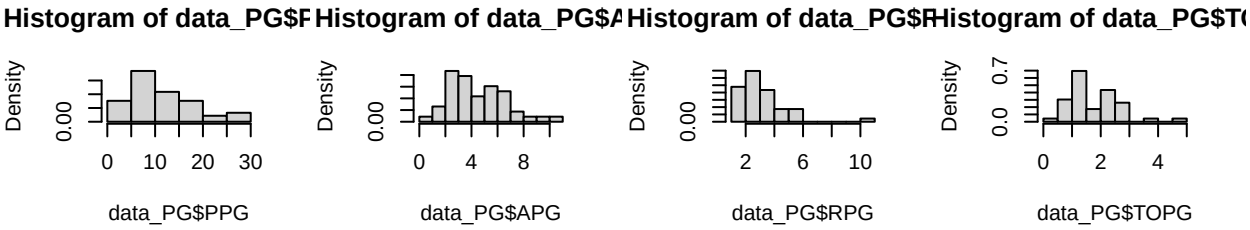
For the PG players,

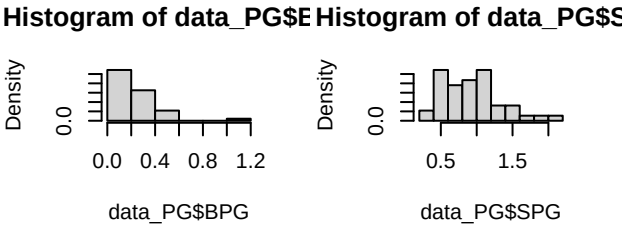
```
corrplot(cor(data_PG[,-c(1,2)]), method = 'number')
```




```
par(mfrow=c(3,4))
hist(data_PG$Age,freq = F)
hist(data_PG$FG.,freq = F)
hist(data_PG$FT.,freq = F)
hist(data_PG$MPG,freq = F)
hist(data_PG$PPG,freq = F)
hist(data_PG$APG,freq = F)
hist(data_PG$RPG,freq = F)
hist(data_PG$TOPG,freq = F)
hist(data_PG$BPG,freq = F)
hist(data_PG$SPG,freq = F)
```








```
lm1 = lm(log(data_PG$salary)~.,data_PG[,c(4:13)])
summary(lm1)
```

```
##
## Call:
## lm(formula = log(data_PG$salary) ~ ., data = data_PG[, c(4:13)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.39808 -0.44226  0.09293  0.43362  0.96283
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.37528    2.35990   4.820 2.76e-05 ***
## Age          0.10002    0.03472   2.881  0.00673 **
## FG.         -0.41153    3.22023  -0.128  0.89904
## FT.         -0.69889    2.01503  -0.347  0.73079
## MPG          0.04783    0.03971   1.205  0.23647
## PPG          0.08335    0.04592   1.815  0.07806 .
## APG         -0.10740    0.11712  -0.917  0.36538
## RPG         -0.10879    0.13097  -0.831  0.41179
## TOPG         0.29198    0.34866   0.837  0.40803
## BPG         -0.16434    0.68483  -0.240  0.81175
## SPG          0.28831    0.45191   0.638  0.52764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6838 on 35 degrees of freedom
```

```
## Multiple R-squared:  0.6561, Adjusted R-squared:  0.5578
## F-statistic: 6.677 on 10 and 35 DF,  p-value: 1.065e-05
```

```
#step(lm1)
lm2 = lm(log(data_PG$salary)~Age + PPG,data_PG)
summary(lm2)
```

```
##
## Call:
## lm(formula = log(data_PG$salary) ~ Age + PPG, data = data_PG)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4695 -0.3951 -0.0248  0.3409  1.2085
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.42999    0.89312  12.798 2.93e-16 ***
## Age          0.09529    0.02948   3.233  0.00236 **
## PPG          0.12242    0.01494   8.192 2.54e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6454 on 43 degrees of freedom
## Multiple R-squared:  0.6236, Adjusted R-squared:  0.6061
## F-statistic: 35.62 on 2 and 43 DF,  p-value: 7.537e-10
```

```
library(car)
```

```
## Loading required package: carData
```

```
which(dffits(lm2)>2*sqrt(2/46))
```

```
## 417
##  44
```

```
which(abs(covratio(lm2)-1)>(3*2/46))
```

```
##  61  74 107 142 179 212 250 270 286 359 376 386 395 410 436
##   1   4   8  10  14  18  21  23  24  31  35  37  41  43  46
```

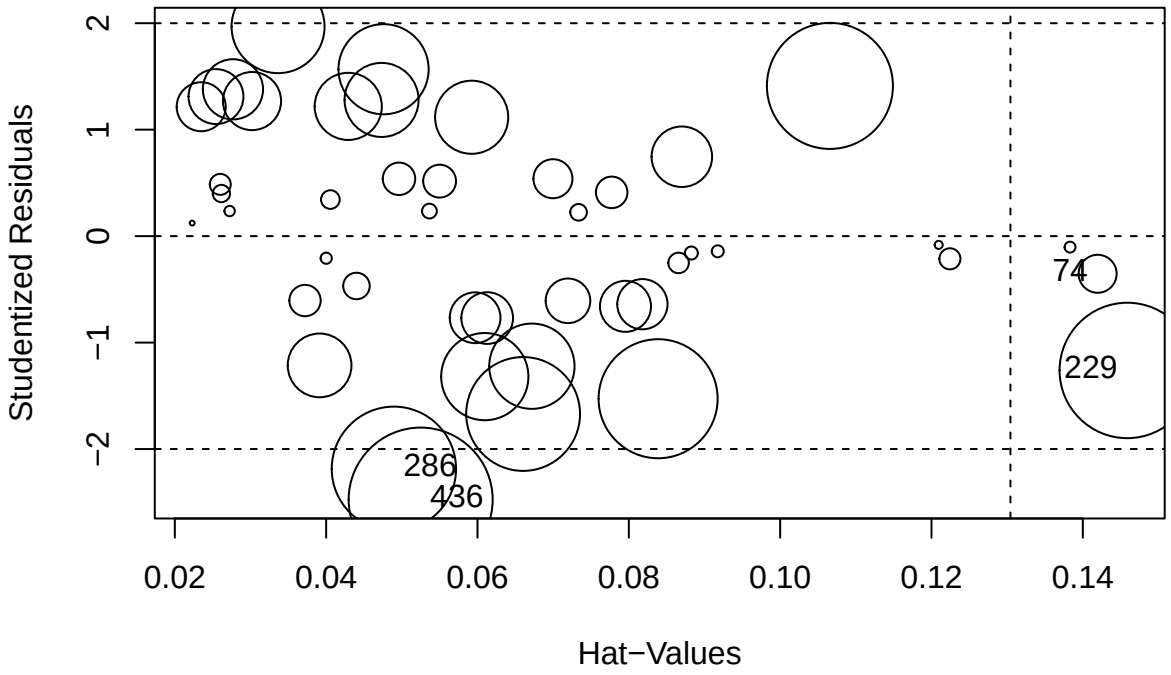
```
which(dfbetas(lm2)>(2/sqrt(46)))
```

```
## [1] 15 20 70 86 90
```

```
outlierTest(lm2)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 436 -2.474546           0.017463      0.80328
```

```
influencePlot(lm2)
```



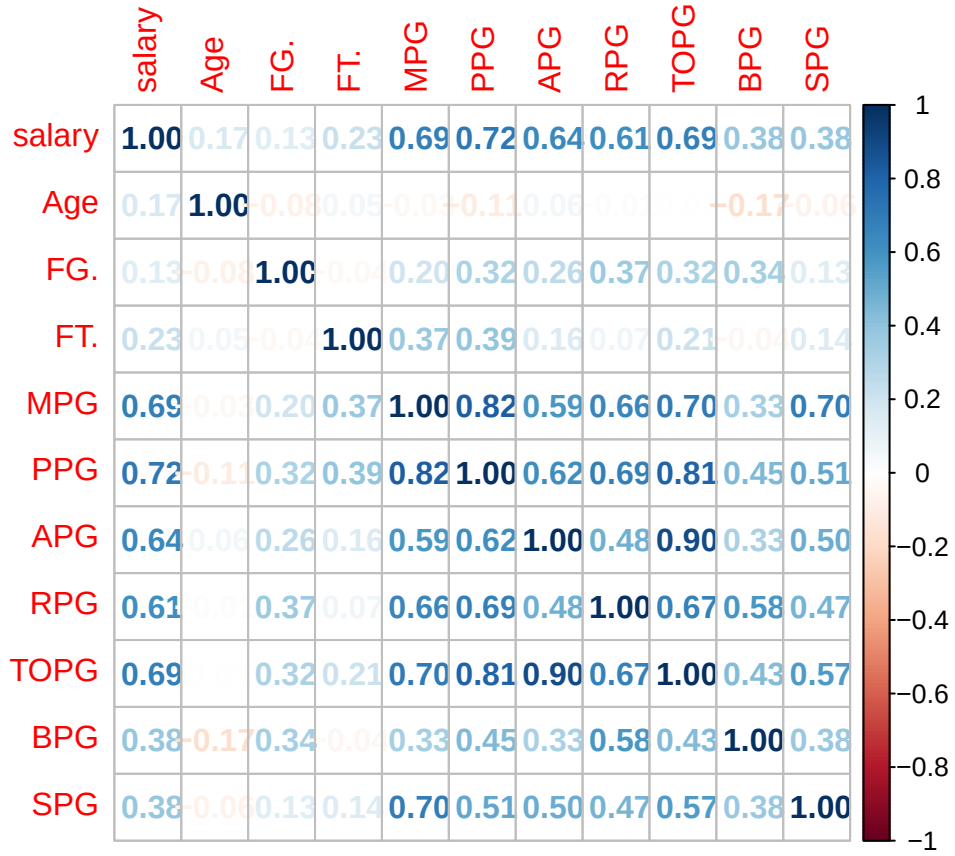

```
##         StudRes        Hat       CookD
## 74   -0.3536104 0.14195908 0.007039019
## 229  -1.2615690 0.14587631 0.089378045
## 286  -2.1852867 0.04895314 0.075322445
## 436  -2.4745462 0.05249088 0.101037513
```

```
#Same varaibles in the model as before
data_PG2 = data_PG[-c(74,229,286,436),]
lm3 = lm(log(data_PG2$salary)~.,data_PG2[,c(4:13)])
#step(lm3)
```

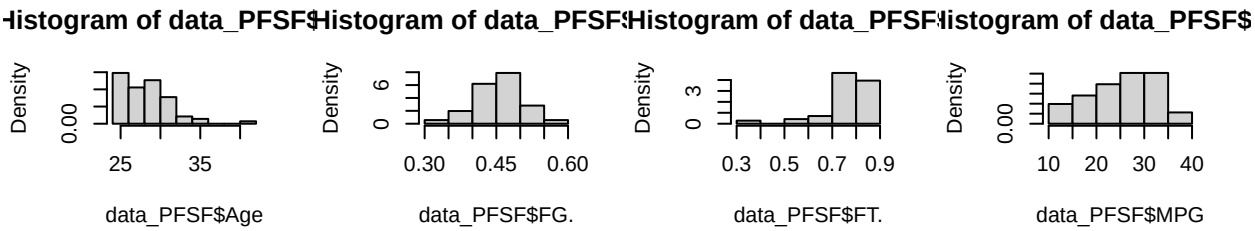
For the PF and SF players

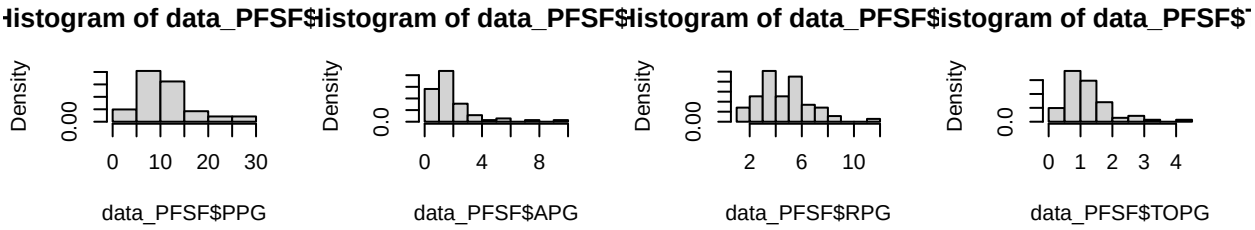
```
data_PFSF = rbind(data_PF,data_SF)
```

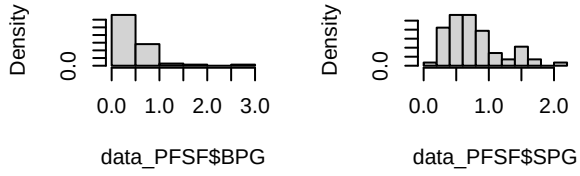
```
#corrplot(cor(data_PF[,-c(1,2)]), method = 'number')
#corrplot(cor(data_SF[,-c(1,2)]), method = 'number')
corrplot(cor(data_PFSF[,-c(1,2)]), method = 'number')
```




```
par(mfrow=c(3,4))
hist(data_PFSF$Age,freq = F)
hist(data_PFSF$FG.,freq = F)
hist(data_PFSF$FT.,freq = F)
hist(data_PFSF$MPG,freq = F)
hist(data_PFSF$PPG,freq = F)
hist(data_PFSF$APG,freq = F)
hist(data_PFSF$RPG,freq = F)
hist(data_PFSF$TOPG,freq = F)
hist(data_PFSF$BPG,freq = F)
hist(data_PFSF$SPG,freq = F)
```








```
lm5 = lm(log(data_PFSF$salary)~.,data_PFSF[,c(4:13)])
summary(lm5)
```

```
##
## Call:
## lm(formula = log(data_PFSF$salary) ~ ., data = data_PFSF[, c(4:13)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44059 -0.42070 -0.02365  0.40681  1.52853
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.71233    1.27303   9.200 4.55e-13 ***
## Age          0.08528    0.02521   3.383  0.00127 **
## FG.         -2.47504    1.73359  -1.428  0.15856
## FT.          0.52665    0.94119   0.560  0.57787
## MPG          0.06178    0.02746   2.250  0.02816 *
## PPG          0.02888    0.03634   0.795  0.42993
## APG          0.18765    0.13411   1.399  0.16690
## RPG          0.14029    0.07419   1.891  0.06347 .
## TOPG        -0.36571    0.39781  -0.919  0.36162
## BPG          0.10580    0.27159   0.390  0.69825
## SPG         -0.24347    0.30280  -0.804  0.42453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6604 on 60 degrees of freedom
```

```
## Multiple R-squared:  0.6077, Adjusted R-squared:  0.5423
## F-statistic: 9.293 on 10 and 60 DF,  p-value: 4.441e-09
```

```
#step(lm5)
lm6 = lm(log(data_PFSF$salary)~Age + MPG + RPG,data_PFSF)
summary(lm6)
```

```
##
## Call:
## lm(formula = log(data_PFSF$salary) ~ Age + MPG + RPG, data = data_PFSF)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.58465 -0.44792  0.03487  0.39978  1.32275
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.75400    0.74875  14.363  < 2e-16 ***
## Age          0.08876    0.02375   3.737 0.000387 ***
## MPG          0.07528    0.01501   5.016 4.12e-06 ***
## RPG          0.11457    0.05433   2.109 0.038713 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6549 on 67 degrees of freedom
## Multiple R-squared:  0.5691, Adjusted R-squared:  0.5498
## F-statistic:  29.5 on 3 and 67 DF,  p-value: 2.838e-12
```

```
library(car)
which(dffits(lm6)>2*sqrt(3/46))
```

```
## 185
##  15
```

```
which(abs(covratio(lm6)-1)>(3*3/46))
```

```
##  22 235 442
##   2  54  68
```

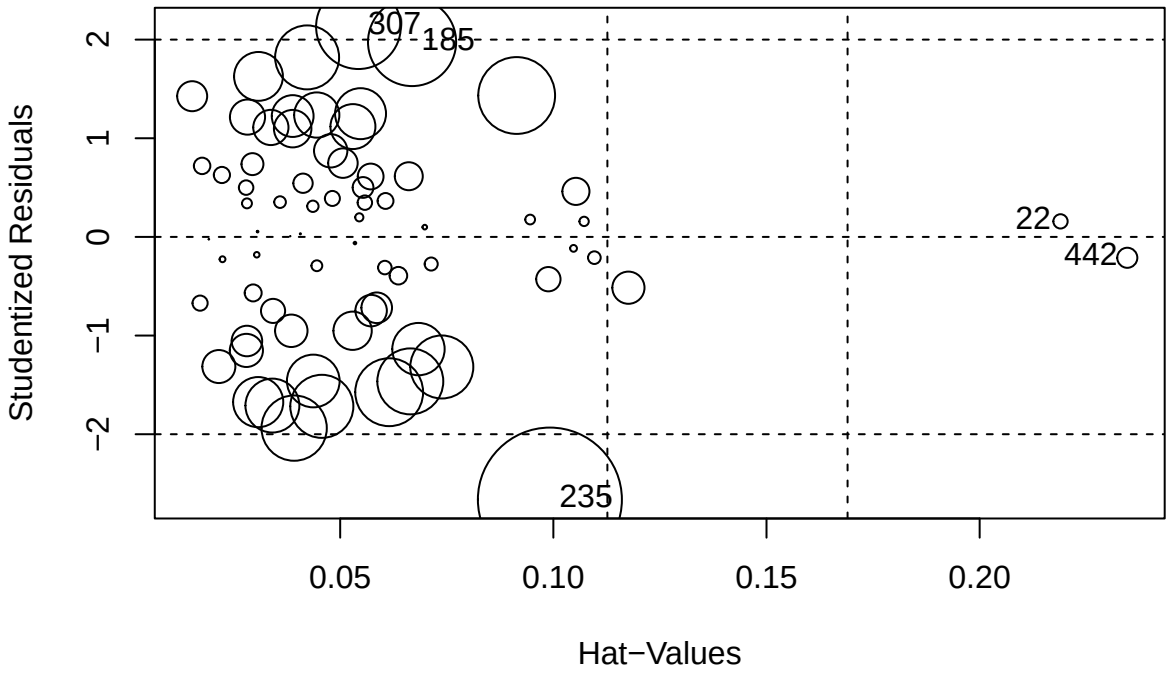
```
which(dfbetas(lm6)>(2/sqrt(46)))
```

```
## [1]  59  60 125 191 267
```

```
outlierTest(lm6)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 235 -2.662551           0.009732      0.69097
```

```
influencePlot(lm6)
```

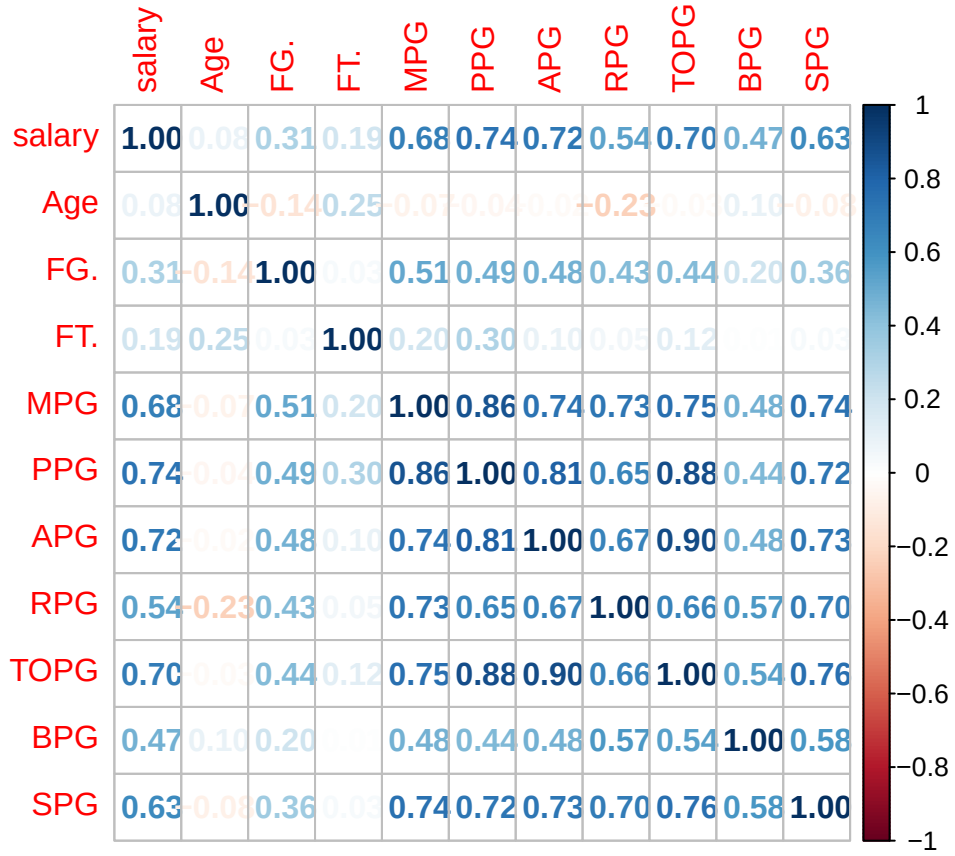



```
##        StudRes        Hat       CookD
## 22   0.1575431 0.21896403 0.001765258
## 185  1.9755474 0.06685735 0.067003324
## 235 -2.6625512 0.09918723 0.178887044
## 307  2.1309564 0.05428110 0.061888371
## 442 -0.2115417 0.23462983 0.003479205
```

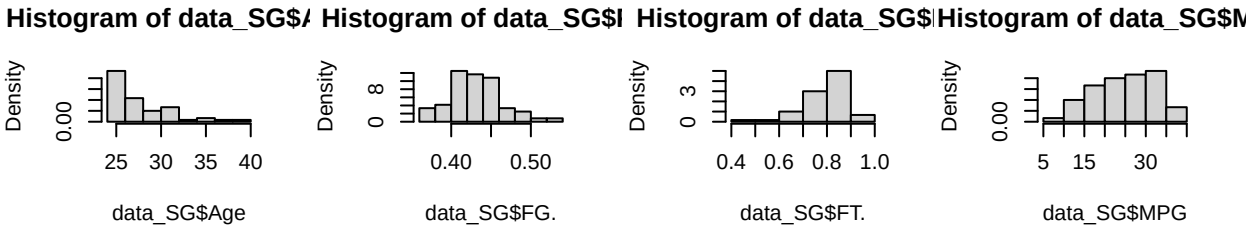
```
#Same varaibles in the model as before
data_PFSF2 = data_PFSF[-c(22,185,235,307,442),]
lm7 = lm(log(data_PFSF2$salary)~.,data_PFSF2[,c(4:13)])
#step(lm7)
```

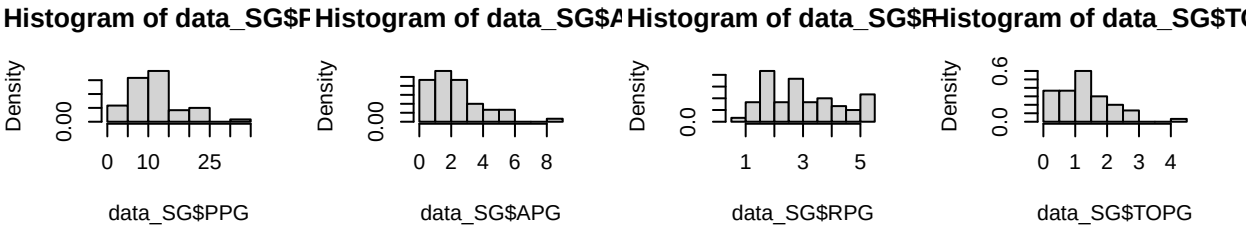
For the SG players,

```
corrplot(cor(data_SG[,-c(1,2)]), method = 'number')
```




```
par(mfrow=c(3,4))
hist(data_SG$Age,freq = F)
hist(data_SG$FG.,freq = F)
hist(data_SG$FT.,freq = F)
hist(data_SG$MPG,freq = F)
hist(data_SG$PPG,freq = F)
hist(data_SG$APG,freq = F)
hist(data_SG$RPG,freq = F)
hist(data_SG$TOPG,freq = F)
hist(data_SG$BPG,freq = F)
hist(data_SG$SPG,freq = F)
```






```
lm9 = lm(log(data_SG$salary)~.,data_SG[,c(4:13)])
summary(lm9)
```

```
##
## Call:
## lm(formula = log(data_SG$salary) ~ ., data = data_SG[, c(4:13)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71901 -0.48090  0.07355  0.44723  1.35337
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.02057    1.82435   7.137 4.07e-09 ***
## Age          0.03361    0.02964   1.134   0.2623
## FG.         -3.37909    3.37402  -1.002   0.3215
## FT.          1.28263    1.19683   1.072   0.2891
## MPG          0.06025    0.03004   2.006   0.0504 .
## PPG          0.01440    0.04865   0.296   0.7685
## APG          0.08445    0.13942   0.606   0.5475
## RPG         -0.09169    0.12972  -0.707   0.4830
## TOPG         0.09819    0.36982   0.266   0.7917
## BPG          0.32850    0.59001   0.557   0.5802
## SPG          0.30905    0.40284   0.767   0.4467
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7243 on 49 degrees of freedom
```

```
## Multiple R-squared:  0.5648, Adjusted R-squared:  0.4759
## F-statistic: 6.358 on 10 and 49 DF,  p-value: 3.582e-06
```

```
#step(lm9)
lm10 = lm(log(data_SG$salary)~ Age + MPG + APG,data_SG)
summary(lm10)
```

```
##
## Call:
## lm(formula = log(data_SG$salary) ~ Age + MPG + APG, data = data_SG)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.65029 -0.49816  0.05603  0.48046  1.48764
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.98625    0.81902  14.635  < 2e-16 ***
## Age          0.05364    0.02546   2.107 0.039631 *
## MPG          0.07172    0.01883   3.810 0.000348 ***
## APG          0.13763    0.08315   1.655 0.103463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7111 on 56 degrees of freedom
## Multiple R-squared:  0.5206, Adjusted R-squared:  0.4949
## F-statistic: 20.27 on 3 and 56 DF,  p-value: 5.089e-09
```

```
library(car)
which(dffits(lm10)>2*sqrt(3/60))
```

```
## 8 9
## 1 2
```

```
which(abs(covratio(lm10)-1)>(3*4/60))
```

```
##   8 145 180 191 265 455
##   1  23  28  29  41  60
```

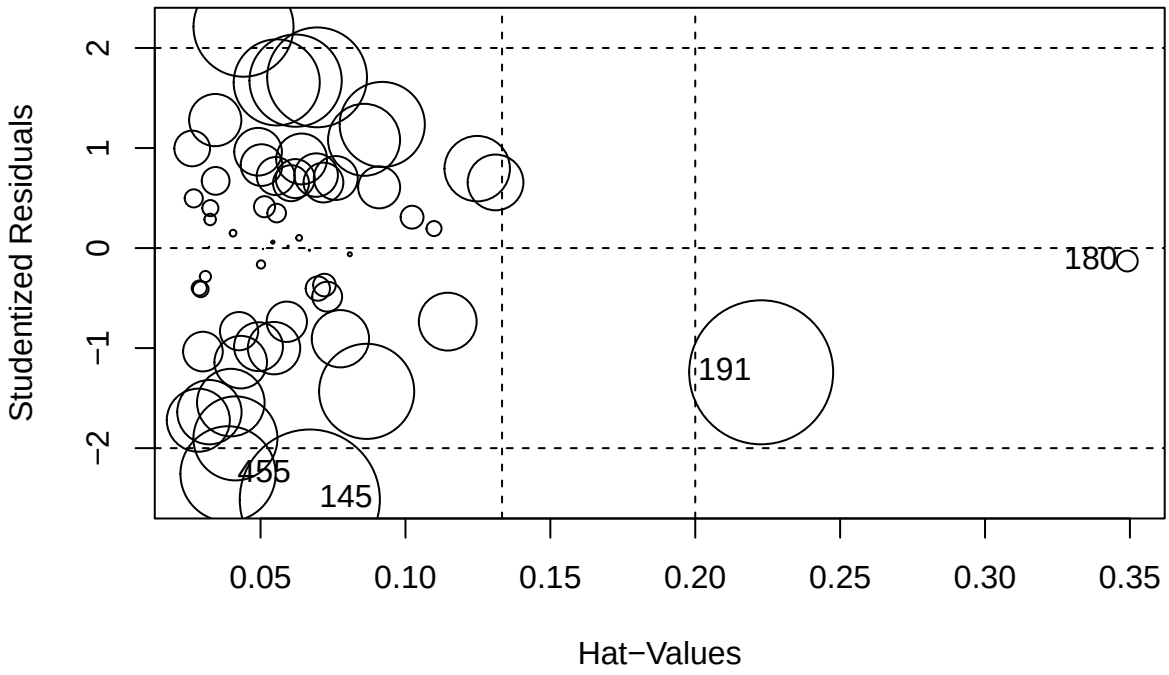
```
which(dfbetas(lm10)>(2/sqrt(60)))
```

```
## [1]   1   2  23  29  43  48 120 123 203
```

```
outlierTest(lm10)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 145 -2.514169           0.014887      0.89322
```

```
influencePlot(lm10)
```

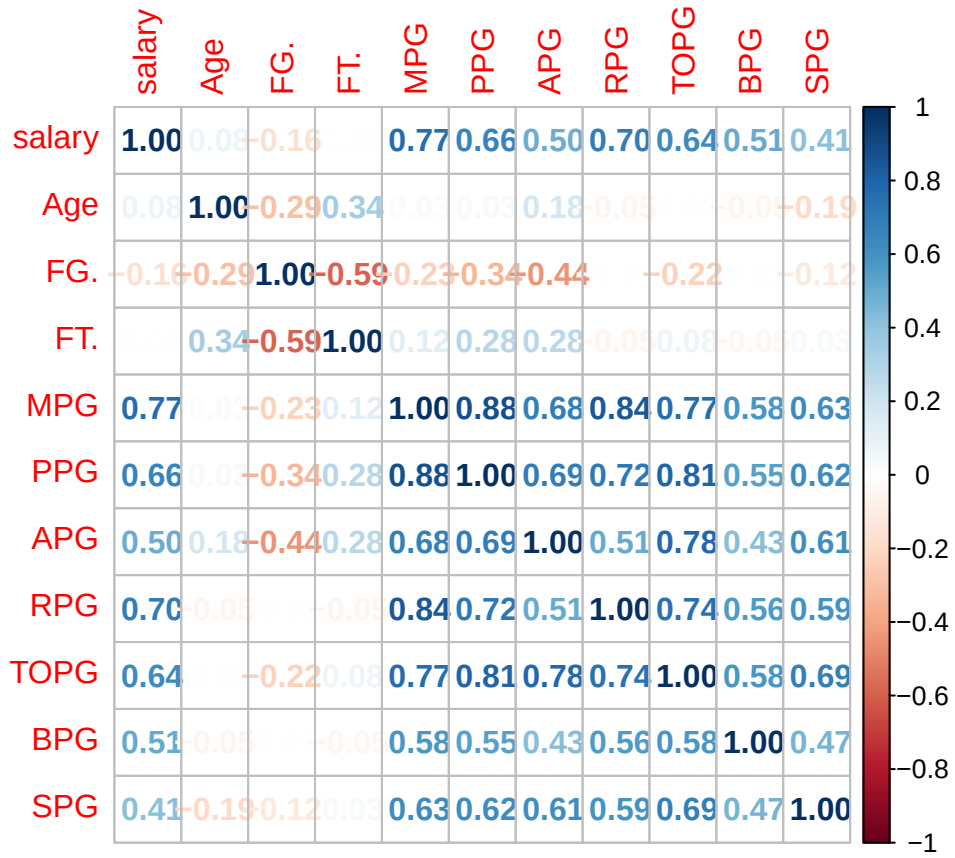



```
##         StudRes        Hat       CookD
## 145 -2.5141691 0.06699268 0.103621478
## 180 -0.1296745 0.34909042 0.002294875
## 191 -1.2418908 0.22274911 0.109440020
## 455 -2.2574579 0.03877990 0.047896606
```

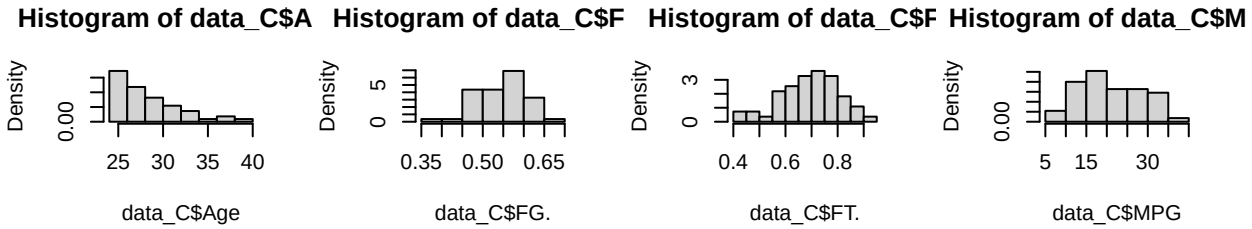
```
#Same varaibles in the model as before
data_SG2 = data_SG[-c(145,180,191,455),]
lm11 = lm(log(data_SG2$salary)~.,data_SG2[,c(4:13)])
#step(lm11)
```

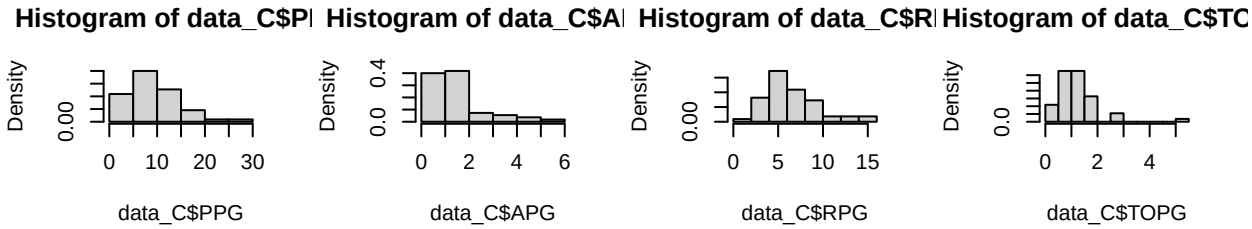
For the C players

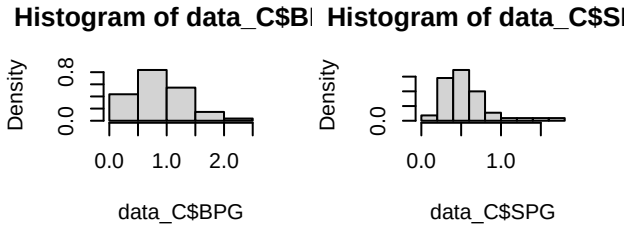
```
corrplot(cor(data_C[,-c(1,2)]), method = 'number')
```




```
par(mfrow=c(3,4))
hist(data_C$Age,freq = F)
hist(data_C$FG.,freq = F)
hist(data_C$FT.,freq = F)
hist(data_C$MPG,freq = F)
hist(data_C$PPG,freq = F)
hist(data_C$APG,freq = F)
hist(data_C$RPG,freq = F)
hist(data_C$TOPG,freq = F)
hist(data_C$BPG,freq = F)
hist(data_C$SPG,freq = F)
```








```
lm13 = lm(log(data_C$salary)~.,data_C[,c(4:13)])
summary(lm13)
```

```
##
## Call:
## lm(formula = log(data_C$salary) ~ ., data = data_C[, c(4:13)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45883 -0.41456 -0.06907  0.51299  1.12229
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.00575    1.90661   7.346 3.57e-09 ***
## Age          0.04307    0.03055   1.410 0.165622
## FG.         -1.90580    2.14520  -0.888 0.379155
## FT.         -0.89904    1.15835  -0.776 0.441821
## MPG          0.13786    0.03816   3.612 0.000774 ***
## PPG         -0.07064    0.05031  -1.404 0.167337
## APG         -0.23969    0.18699  -1.282 0.206615
## RPG         -0.01264    0.06734  -0.188 0.851973
## TOPG         0.45401    0.31038   1.463 0.150643
## BPG         -0.06569    0.28204  -0.233 0.816910
## SPG         -0.13963    0.49185  -0.284 0.777827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7032 on 44 degrees of freedom
```

```
## Multiple R-squared:  0.6094, Adjusted R-squared:  0.5207
## F-statistic: 6.865 on 10 and 44 DF,  p-value: 2.406e-06
```

```
#step(lm13)
lm14 = lm(log(data_C$salary)~Age + MPG + PPG+ APG + TOPG ,data_C)
summary(lm14)
```

```
##
## Call:
## lm(formula = log(data_C$salary) ~ Age + MPG + PPG + APG + TOPG,
##     data = data_C)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.50213 -0.32268 -0.08718  0.52865  1.06823
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.24263    0.81784  14.969  < 2e-16 ***
## Age          0.04483    0.02664   1.682   0.0988 .
## MPG          0.13213    0.02676   4.938 9.56e-06 ***
## PPG         -0.07345    0.04255  -1.726   0.0906 .
## APG         -0.20575    0.15191  -1.354   0.1818
## TOPG         0.40084    0.24666   1.625   0.1106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6765 on 49 degrees of freedom
## Multiple R-squared:  0.5975, Adjusted R-squared:  0.5564
## F-statistic: 14.55 on 5 and 49 DF,  p-value: 9.817e-09
```

```
library(car)
which(dffits(lm14)>2*sqrt(5/55))
```

```
## 4
## 1
```

```
which(abs(covratio(lm14)-1)>(3*5/55))
```

```
##   4  15  77  97 121 271 290 335 349 375 397 437
##   1   4  10  13  17  33  35  42  43  46  48  52
```

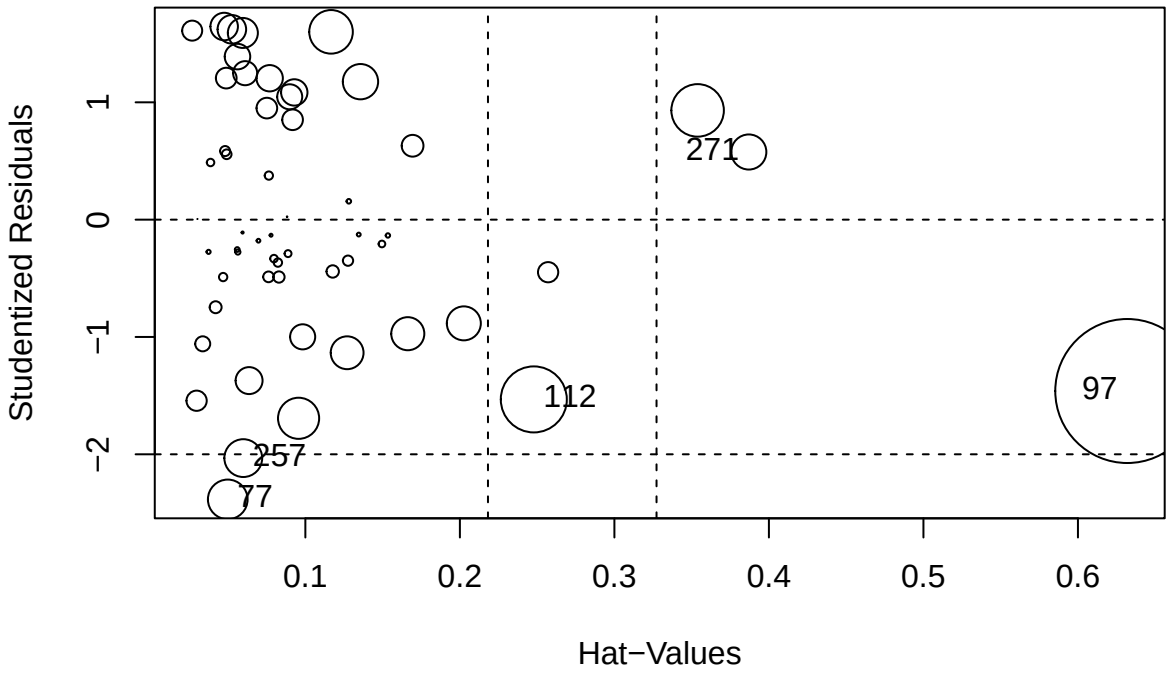
```
which(dfbetas(lm14)>(2/sqrt(55)))
```

```
##  [1]  16  52  65  79  85 108 123 172 194 198 217 221 291 299 328
```

```
outlierTest(lm14)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 77 -2.38438           0.021108           NA
```

```
influencePlot(lm14)
```




```
##       StudRes        Hat      CookD
## 77  -2.384380 0.04976428 0.04529258
## 97  -1.460625 0.63193724 0.59668766
## 112 -1.532146 0.24783823 0.12546553
## 257 -2.032457 0.05981689 0.04117218
## 271  0.576530 0.38697584 0.03545329
```

```
#Same varaibles in the model as before
data_C2 = data_C[-c(77,97,112,257,271),]
lm15 = lm(log(data_C2$salary)~.,data_C2[,c(4:13)])
#step(lm15)
```

PF

```
lmod1 = lm(log(data_PF$salary)~.,data_PF[,c(4:13)])
summary(lmod1)
```

```
##
## Call:
## lm(formula = log(data_PF$salary) ~ ., data = data_PF[, c(4:13)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1625 -0.3874 -0.1142  0.4695  1.0095
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.30427    1.59214   7.100 6.77e-08 ***
## Age          0.09409    0.04147   2.269   0.0306 *
## FG.         -2.31517    2.21292  -1.046   0.3038
## FT.          0.66553    1.02312   0.650   0.5203
## MPG          0.05307    0.03808   1.393   0.1737
```

```
## PPG          0.03248    0.04320   0.752   0.4579
## APG          0.34137    0.16426   2.078   0.0463 *
## RPG          0.24499    0.11319   2.165   0.0385 *
## TOPG        -0.86796    0.49761  -1.744   0.0914 .
## BPG          0.01858    0.33985   0.055   0.9568
## SPG         -0.25579    0.46112  -0.555   0.5832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6283 on 30 degrees of freedom
## Multiple R-squared:  0.6936, Adjusted R-squared:  0.5915
## F-statistic: 6.792 on 10 and 30 DF,  p-value: 1.975e-05
```

```
#step(lmod1)
lmod2 = lm(log(data_PF$salary)~ Age + MPG + APG + RPG + TOPG ,data_PF)
summary(lmod2)
```

```
##
## Call:
## lm(formula = log(data_PF$salary) ~ Age + MPG + APG + RPG + TOPG,
##     data = data_PF)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21872 -0.47668 -0.02955  0.42568  1.10189
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.91984    1.11232   9.817 1.37e-11 ***
## Age          0.08100    0.03706   2.185  0.03564 *
## MPG          0.07401    0.02647   2.796  0.00835 **
## APG          0.29177    0.15149   1.926  0.06226 .
## RPG          0.20414    0.09802   2.083  0.04466 *
## TOPG        -0.73215    0.42175  -1.736  0.09136 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6126 on 35 degrees of freedom
## Multiple R-squared:  0.6602, Adjusted R-squared:  0.6117
## F-statistic:  13.6 on 5 and 35 DF,  p-value: 2.136e-07
```

SF

```
lmod3 = lm(log(data_SF$salary)~.,data_SF[,c(4:13)])
summary(lmod3)
```

```
##
## Call:
## lm(formula = log(data_SF$salary) ~ ., data = data_SF[, c(4:13)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.66711 -0.34751 -0.06671  0.29731  1.37242
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 10.82433    2.96840   3.647  0.00172 **
## Age          0.11778    0.04665   2.525  0.02063 *
## FG.          0.20730    4.07044   0.051  0.95992
## FT.         -1.07448    2.52128  -0.426  0.67478
## MPG          0.08865    0.05711   1.552  0.13709
## PPG         -0.01268    0.08101  -0.157  0.87729
## APG         -0.19051    0.36463  -0.522  0.60738
## RPG         -0.09161    0.23577  -0.389  0.70193
## TOPG         0.88853    1.01848   0.872  0.39388
## BPG          0.12819    0.59110   0.217  0.83062
## SPG         -0.35881    0.54569  -0.658  0.51872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8075 on 19 degrees of freedom
## Multiple R-squared:  0.5542, Adjusted R-squared:  0.3196
## F-statistic: 2.362 on 10 and 19 DF,  p-value: 0.05129
```

```
#step(lmod3)
lmod4 = lm(log(data_SF$salary)~ Age + MPG + TOPG ,data_SF)
summary(lmod4)
```

```
##
## Call:
## lm(formula = log(data_SF$salary) ~ Age + MPG + TOPG, data = data_SF)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63686 -0.49200  0.01551  0.29576  1.42490
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.68042    1.11337   9.593    5e-10 ***
## Age          0.10649    0.03348   3.181  0.00378 **
## MPG          0.05701    0.02772   2.057  0.04985 *
## TOPG         0.41064    0.30520   1.345  0.19008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7049 on 26 degrees of freedom
## Multiple R-squared:  0.5351, Adjusted R-squared:  0.4814
## F-statistic: 9.974 on 3 and 26 DF,  p-value: 0.0001496
```