

# STA302 Final Project rmd

STA304 - Fall 2023 -Assignment 2

Hugo Feng, Chenxu Mao, Yakun Wang

December 11, 2023

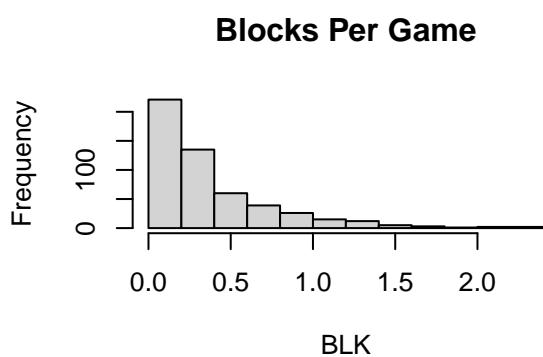
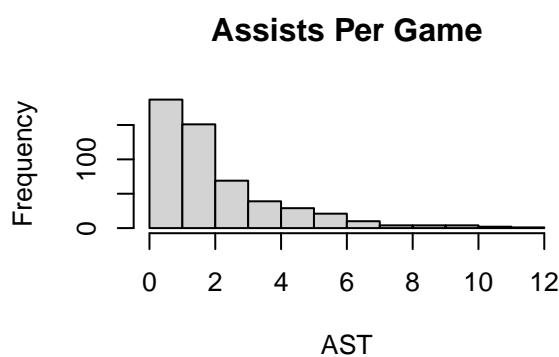
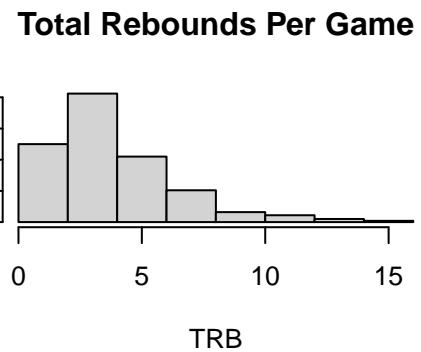
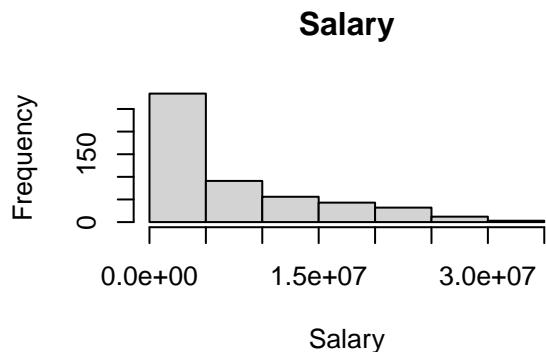
## Model Validation

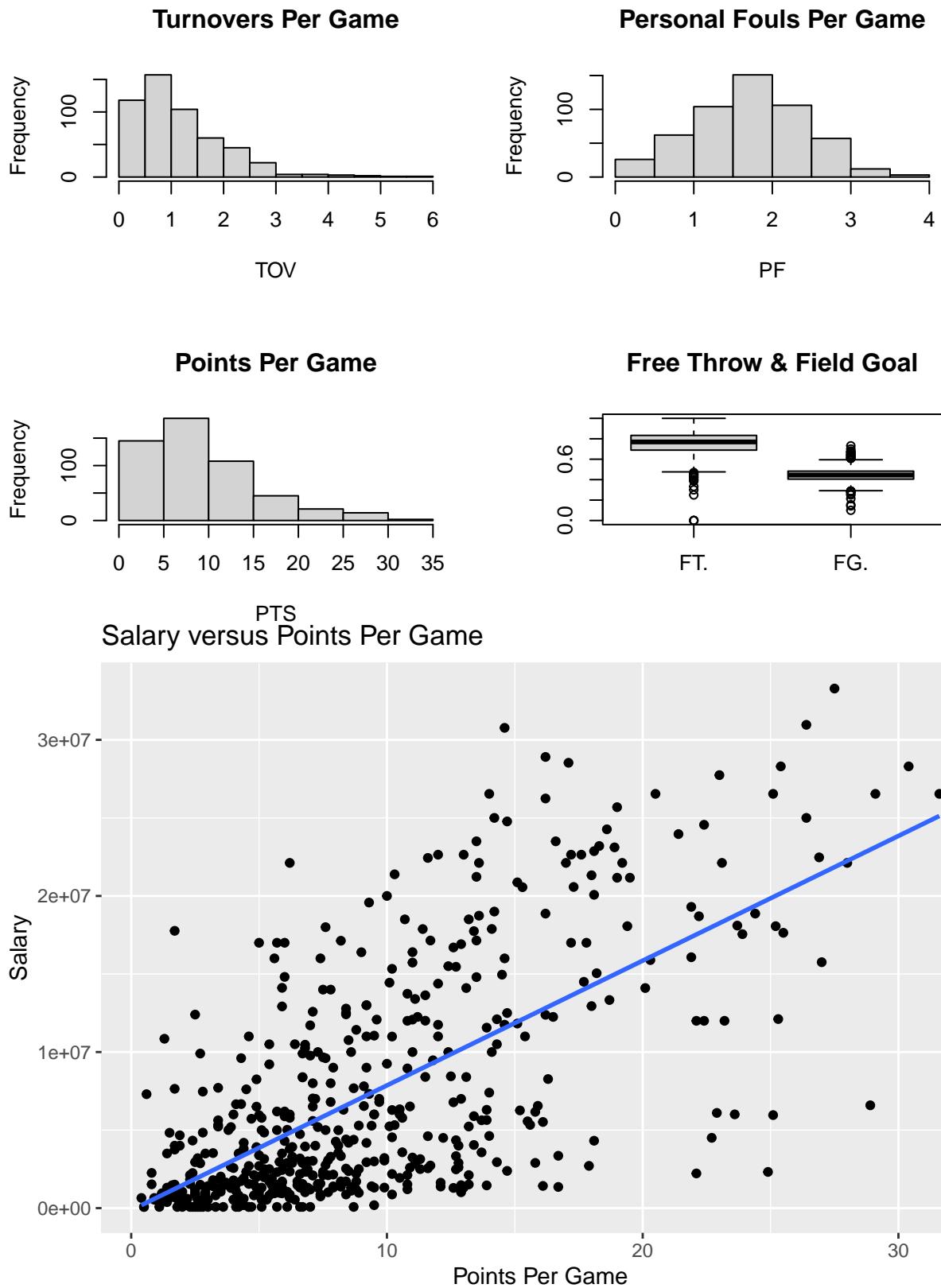
## Result

```
##      FT.          FG.          TRB          AST
## Min. :0.0000  Min. :0.1000  Min. : 0.200  Min. : 0.000
## 1st Qu.:0.6890 1st Qu.:0.4050 1st Qu.: 2.100 1st Qu.: 0.800
## Median :0.7690 Median :0.4450 Median : 3.300 Median : 1.500
## Mean   :0.7463 Mean   :0.4485 Mean   : 3.843 Mean   : 2.095
## 3rd Qu.:0.8330 3rd Qu.:0.4830 3rd Qu.: 5.100 3rd Qu.: 2.700
## Max.   :1.0000 Max.   :0.7310 Max.   :15.200 Max.   :11.200
##      BLK          TOV          PF          PTS
## Min. :0.0000  Min. :0.00  Min. :0.000  Min. : 0.400
## 1st Qu.:0.1000 1st Qu.:0.60 1st Qu.:1.300 1st Qu.: 4.700
## Median :0.3000 Median :1.00 Median :1.800 Median : 7.700
## Mean   :0.4152 Mean   :1.22 Mean   :1.759 Mean   : 9.225
## 3rd Qu.:0.5000 3rd Qu.:1.60 3rd Qu.:2.300 3rd Qu.:12.700
## Max.   :2.4000 Max.   :5.70 Max.   :3.900 Max.   :31.600
##      Salary
## Min.   : 77250
## 1st Qu.:1505942
## Median :4187599
## Mean   :7242215
## 3rd Qu.:11709603
## Max.   :33285709

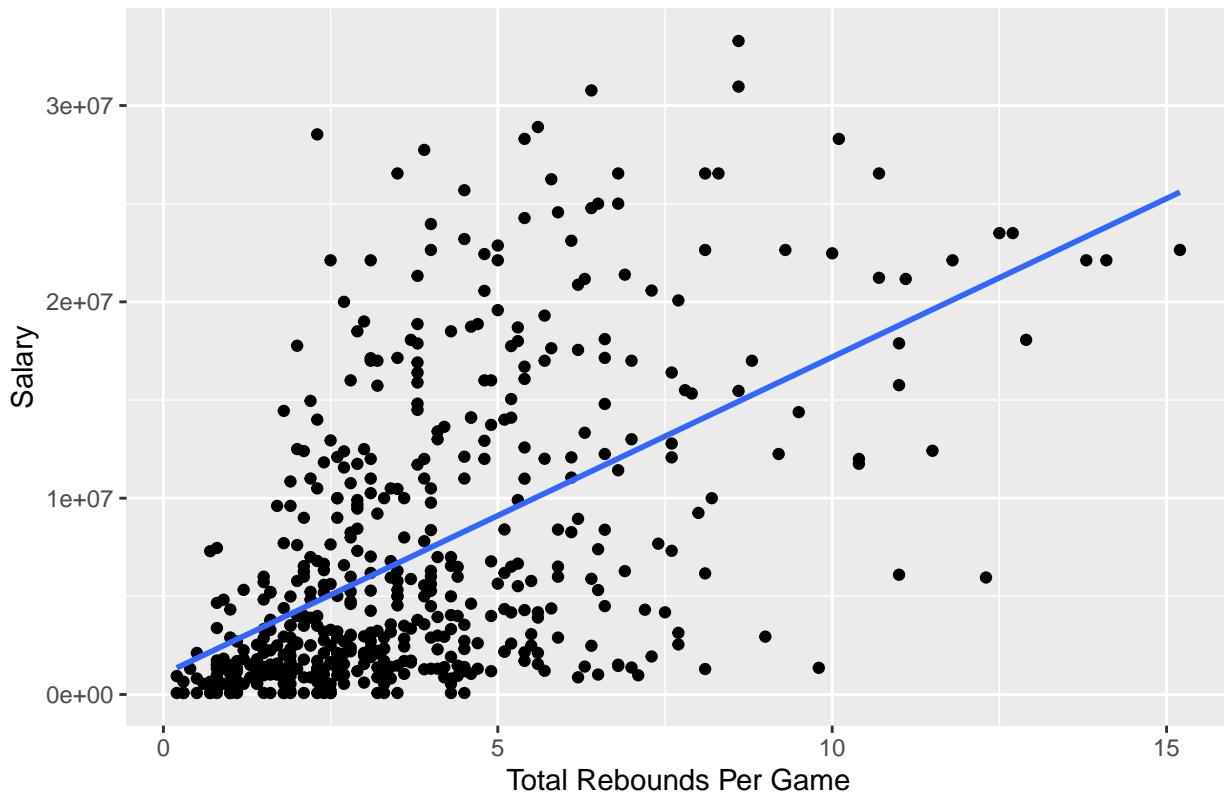
##      FT.          FG.          TRB          AST
## Min. :0.0000  Min. :0.0000  Min. : 0.000  Min. : 0.000
## 1st Qu.:0.6677 1st Qu.:0.4052 1st Qu.: 2.000 1st Qu.: 0.800
## Median :0.7600 Median :0.4440 Median : 3.200 Median : 1.300
## Mean   :0.7378 Mean   :0.4518 Mean   : 3.791 Mean   : 1.913
## 3rd Qu.:0.8197 3rd Qu.:0.4930 3rd Qu.: 4.800 3rd Qu.: 2.400
## Max.   :1.0000 Max.   :0.7500 Max.   :16.000 Max.   :10.700
##      BLK          TOV          PF          PTS
## Min. :0.0000  Min. :0.000  Min. :0.000  Min. : 0.500
## 1st Qu.:0.1000 1st Qu.:0.600 1st Qu.:1.300 1st Qu.: 4.600
## Median :0.3000 Median :0.900 Median :1.800 Median : 7.400
## Mean   :0.4247 Mean   :1.131 Mean   :1.789 Mean   : 8.994
## 3rd Qu.:0.5000 3rd Qu.:1.500 3rd Qu.:2.300 3rd Qu.:12.175
## Max.   :2.7000 Max.   :5.000 Max.   :3.800 Max.   :36.100
```

```
##      Salary
##  Min.   :  56845
##  1st Qu.: 1512601
##  Median : 3362124
##  Mean   : 6750262
##  3rd Qu.:10000000
##  Max.   :37457154
```



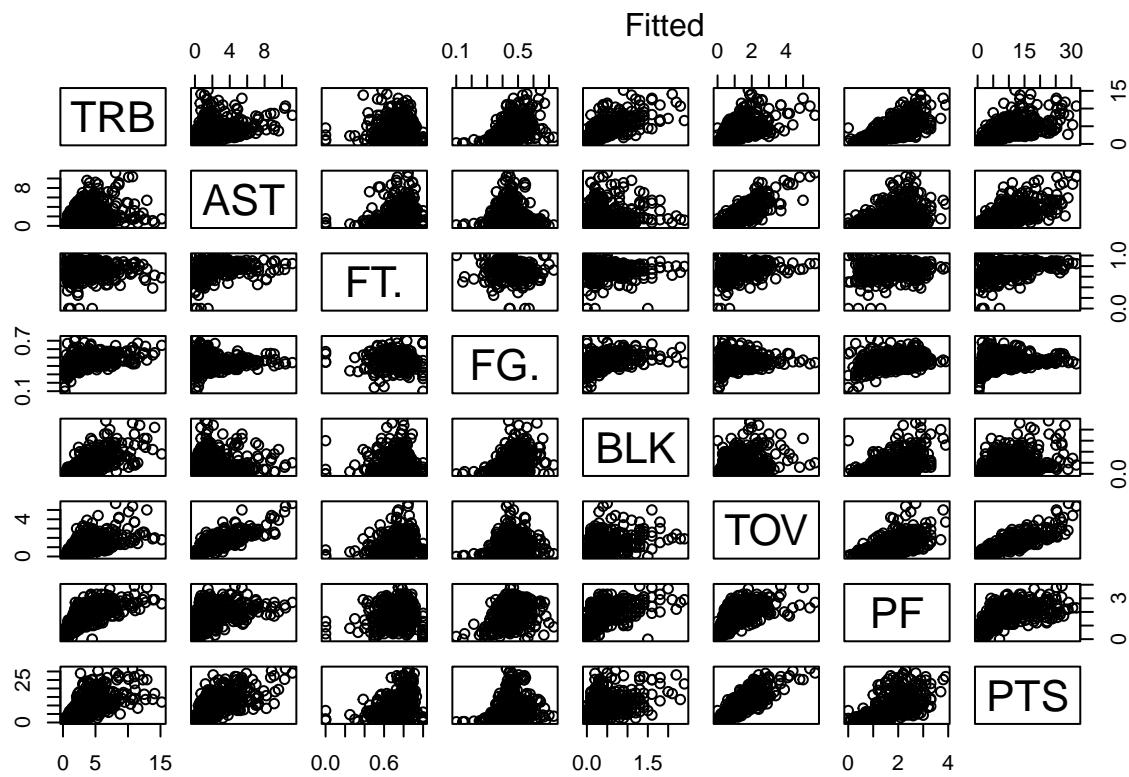
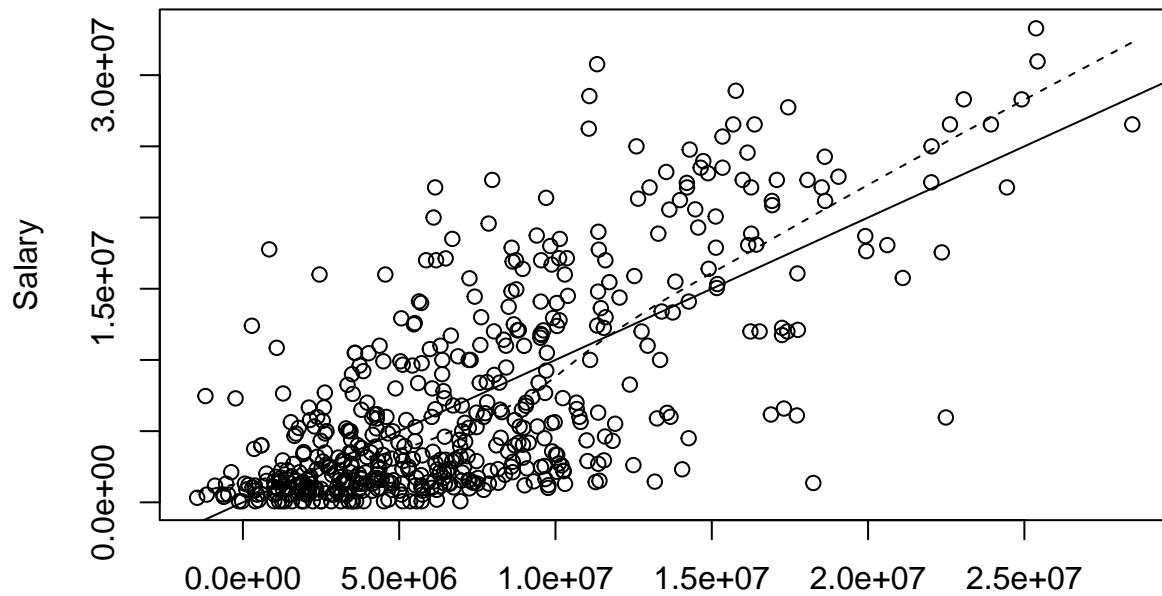


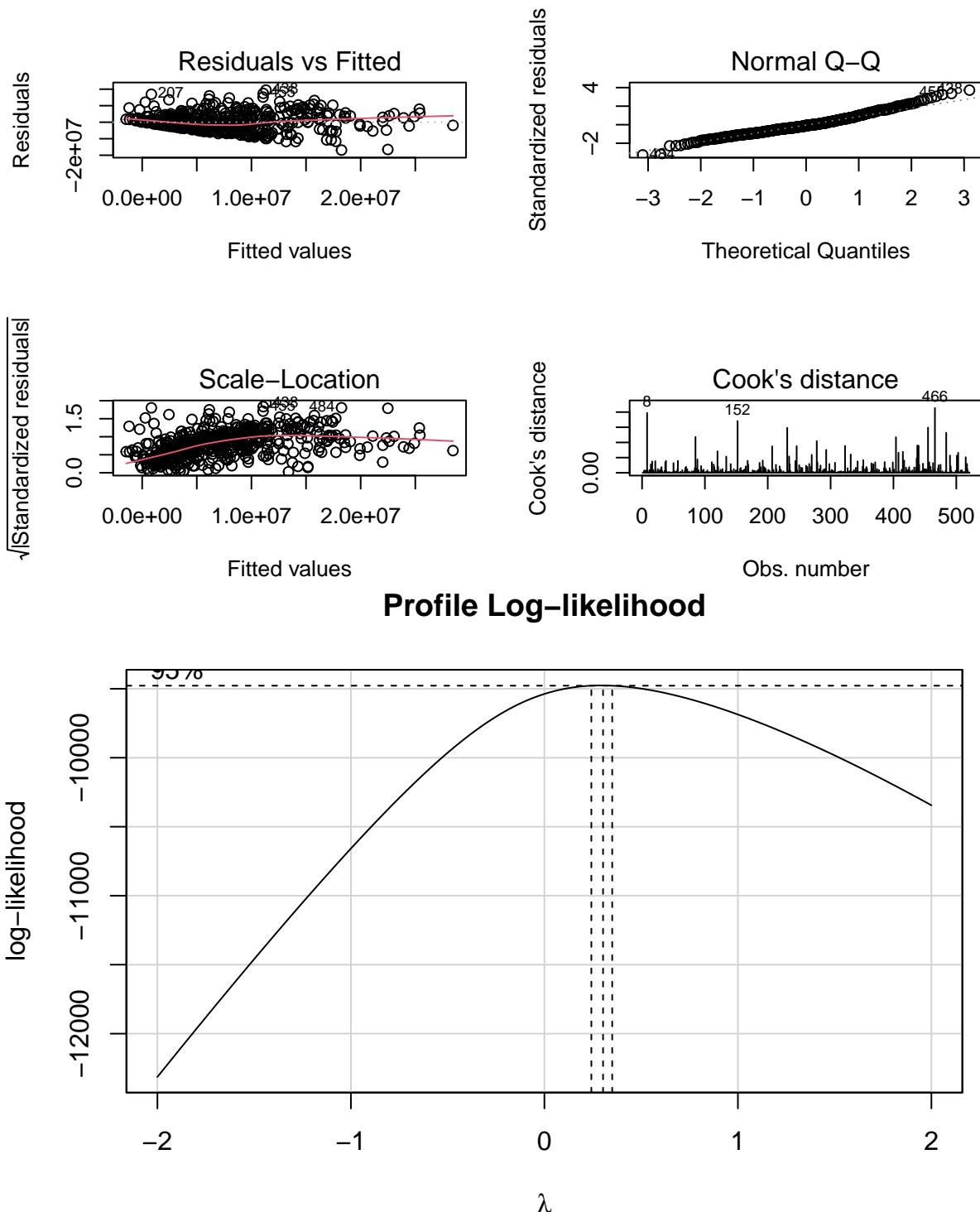
Salary versus Total Rebounds Per Game



```
##
## Call:
## lm(formula = Salary ~ FT. + FG. + TRB + AST + BLK + TOV + PF +
##     PTS + Pos1, data = training)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -16889653 -3404776 -480893  3017934 19438826
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1755171   2307031 -0.761  0.44713
## FT.         2055954   1917042  1.072  0.28402
## FG.        -3595997   3616806 -0.994  0.32057
## TRB         1170077   168899  6.928 1.29e-11 ***
## AST         1376843   264628  5.203 2.84e-07 ***
## BLK         387939    844119  0.460  0.64601
## TOV        -2260163   802720 -2.816  0.00506 **
## PF          -1248029   489869 -2.548  0.01114 *
## PTS         557712    79651  7.002 7.98e-12 ***
## Pos1        455295   180947  2.516  0.01217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5252000 on 511 degrees of freedom
## Multiple R-squared:  0.511, Adjusted R-squared:  0.5024
## F-statistic: 59.33 on 9 and 511 DF, p-value: < 2.2e-16
```

## Salary versus Fitted





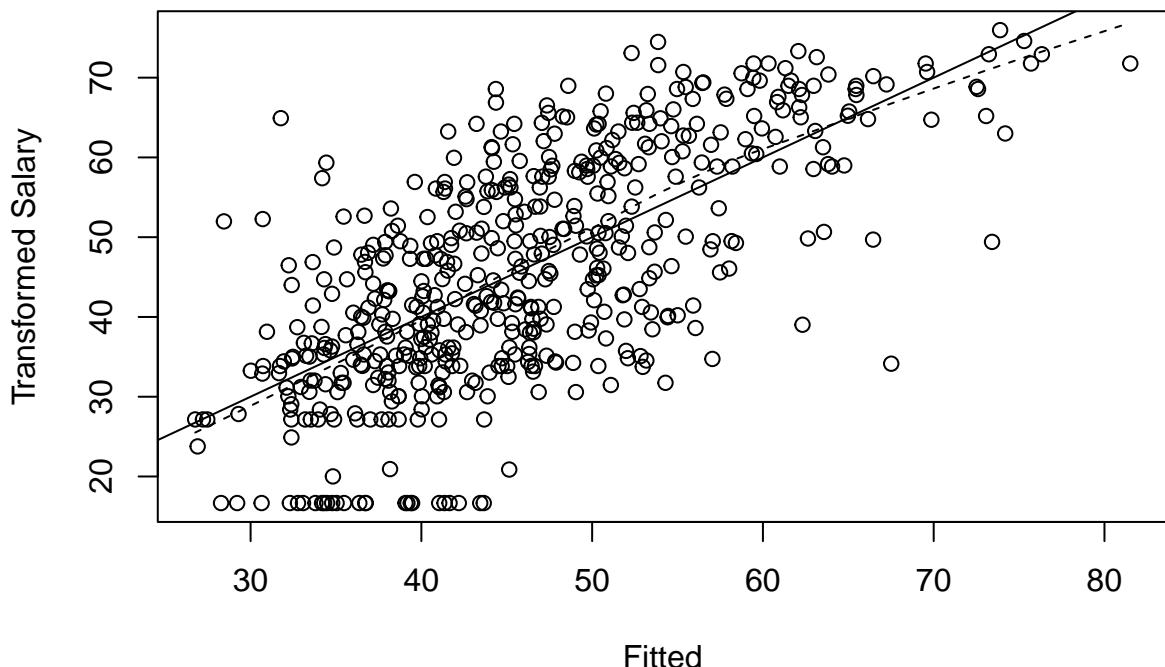
```
##
## Call:
## lm(formula = transformed_Salary ~ FT. + FG. + TRB + AST + BLK +
##     TOV + PF + PTS + Pos1, data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -11500 -10500 -10500 -10500 -10500
```

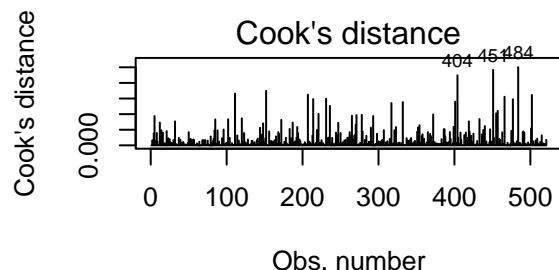
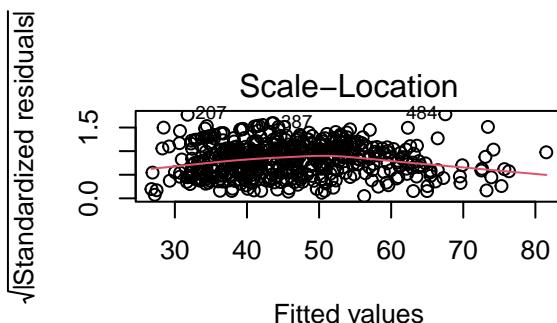
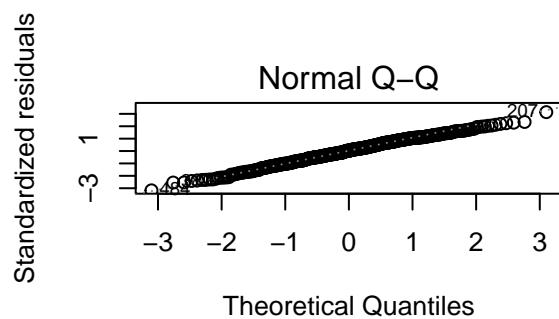
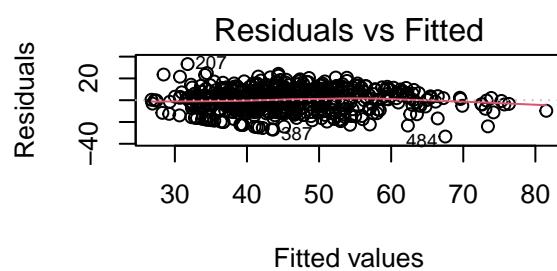
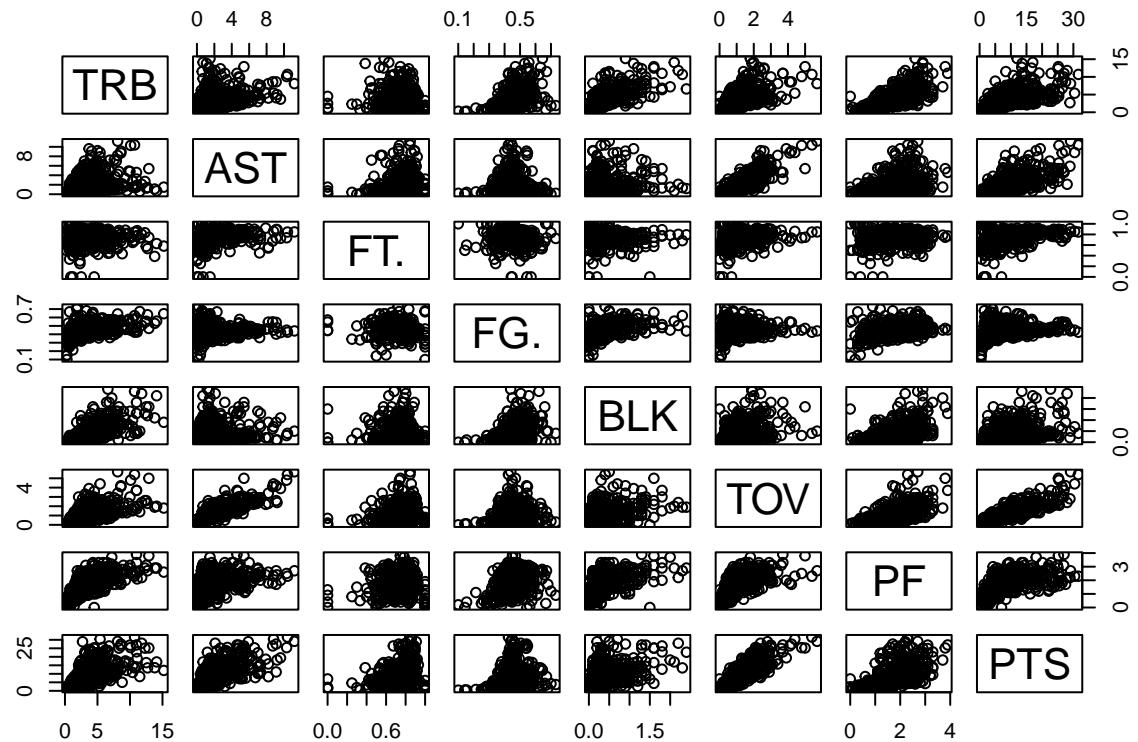
```

## -33.374 -7.184 0.371 7.860 33.153
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.1521    4.6818   4.518 7.76e-06 ***
## FT.         7.6374    3.8904   1.963  0.05017 .
## FG.        -2.7614    7.3398  -0.376  0.70691
## TRB         1.8678    0.3428   5.449 7.87e-08 ***
## AST         2.1329    0.5370   3.972 8.16e-05 ***
## BLK        -0.7280    1.7130  -0.425  0.67103
## TOV        -3.2672    1.6290  -2.006  0.04542 *
## PF          1.3384    0.9941   1.346  0.17880
## PTS         0.8475    0.1616   5.243 2.32e-07 ***
## Pos1        0.9882    0.3672   2.691  0.00735 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.66 on 511 degrees of freedom
## Multiple R-squared: 0.4655, Adjusted R-squared: 0.4561
## F-statistic: 49.45 on 9 and 511 DF, p-value: < 2.2e-16

```

### Transformed Salary versus Fitted





```
##  
## Call:  
## lm(formula = transformed_Salary ~ TRB + AST + TOV + PTS + Pos1,  
##      data = training)  
##  
## Residuals:  
##      Min        1Q    Median        3Q       Max  
## -32.373   -7.454    0.017    8.300   30.557
```

```

## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 26.8416   1.4768  18.176 < 2e-16 ***
## TRB          1.7745   0.2633   6.739 4.29e-11 ***
## AST          2.0868   0.5297   3.939 9.29e-05 *** 
## TOV         -2.8752   1.5568  -1.847  0.06535 .  
## PTS          0.9482   0.1521   6.236 9.38e-10 *** 
## Pos1         0.9451   0.3652   2.588  0.00994 ** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 10.68 on 515 degrees of freedom
## Multiple R-squared:  0.4585, Adjusted R-squared:  0.4533 
## F-statistic: 87.23 on 5 and 515 DF,  p-value: < 2.2e-16

## 
## Call:
## lm(formula = transformed_Salary ~ TRB + AST + PTS + Pos1, data = training)
## 
## Residuals:
##      Min       1Q     Median       3Q      Max      
## -31.3094  -7.3441  -0.0042   8.0116  29.8696
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 26.8177   1.4802  18.118 < 2e-16 ***
## TRB          1.5845   0.2430   6.522 1.66e-10 ***
## AST          1.3794   0.3668   3.761 0.000189 *** 
## PTS          0.8071   0.1318   6.124 1.81e-09 *** 
## Pos1         0.9553   0.3660   2.610 0.009322 ** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 10.71 on 516 degrees of freedom
## Multiple R-squared:  0.455, Adjusted R-squared:  0.4507 
## F-statistic: 107.7 on 4 and 516 DF,  p-value: < 2.2e-16

## Analysis of Variance Table
## 
## Response: transformed_Salary
##             Df Sum Sq Mean Sq F value    Pr(>F)    
## TRB          1 31077 31077.5 270.9424 < 2.2e-16 ***
## AST          1 11898 11897.6 103.7271 < 2.2e-16 ***
## PTS          1  5648  5647.6  49.2372 7.163e-12 ***
## Pos1         1    781    781.2   6.8112  0.009322 ** 
## Residuals 516 59186   114.7                
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

select model4

## named integer(0)

##  32  37  38  69  76  85  87  88  90  98 112 115 134 152 180 189 192 194 195 222
##  32  37  38  69  76  85  87  88  90  98 112 115 134 152 180 189 192 194 195 222

```

```

## 231 234 290 291 332 342 352 371 385 404 413 415 417 428 440 441 466 472 474 481
## 231 234 290 291 332 342 352 371 385 404 413 415 417 428 440 441 466 472 474 481
## 498 499 501 503 517 518
## 498 499 501 503 517 518

## named integer(0)

## 32 85 88 152 221 231 236 270 290 317 323 332 352 401 404 440 455 457 466 475
## 32 85 88 152 221 231 236 270 290 317 323 332 352 401 404 440 455 457 466 475
## 484
## 484

## [1] "Beta 0"
## 4 12 50 96 102 104 111 152 193 207 221 247 265 271 290 354 360 368 372 401
## 4 12 50 96 102 104 111 152 193 207 221 247 265 271 290 354 360 368 372 401
## 419 422 425 439 447 451 455 457 463 475 477 484
## 419 422 425 439 447 451 455 457 463 475 477 484
## [1] "Beta 1"
## 37 85 152 178 185 187 192 231 244 290 307 332 352 365 399 404 415 433 436 440
## 37 85 152 178 185 187 192 231 244 290 307 332 352 365 399 404 415 433 436 440
## 455 475 484 501
## 455 475 484 501
## [1] "Beta 2"
## 14 16 21 25 32 50 88 112 120 134 146 152 189 192 221 246 261 262 279 291
## 14 16 21 25 32 50 88 112 120 134 146 152 189 192 221 246 261 262 279 291
## 323 332 340 401 408 413 425 440 447 463 466 477 484 490
## 323 332 340 401 408 413 425 440 447 463 466 477 484 490
## [1] "Beta 3"
## 14 21 37 50 85 88 120 152 169 178 192 207 231 262 266 278 279 307 323 332
## 14 21 37 50 85 88 120 152 169 178 192 207 231 262 266 278 279 307 323 332
## 352 401 404 413 440 455 466 484
## 352 401 404 413 440 455 466 484
## [1] "Beta 4"
## 5 12 14 25 39 50 85 88 94 102 214 219 221 236 262 263 265 270 271 279
## 5 12 14 25 39 50 85 88 94 102 214 219 221 236 262 263 265 270 271 279
## 284 317 321 323 351 353 368 422 425 447 451 455 457 463 466 475 483
## 284 317 321 323 351 353 368 422 425 447 451 455 457 463 466 475 483

##      TRB       AST       PTS      Pos1
## 1.673402 2.264794 3.009965 1.168712

```

## Appendix 1: Result

```

##
## Call:
## lm(formula = Salary ~ FT. + FG. + TRB + AST + BLK + TOV + PF +
##     PTS + Pos1, data = test)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18806508 -3251908 -820219  2863915 20963462
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -103388    2033942 -0.051  0.95947

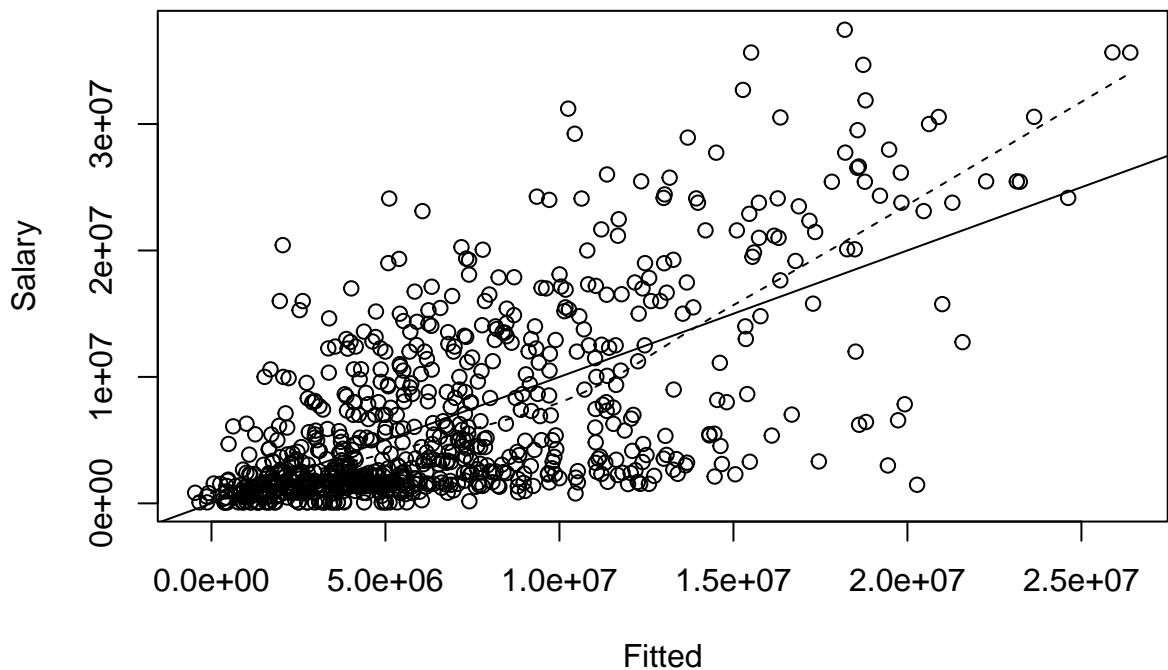
```

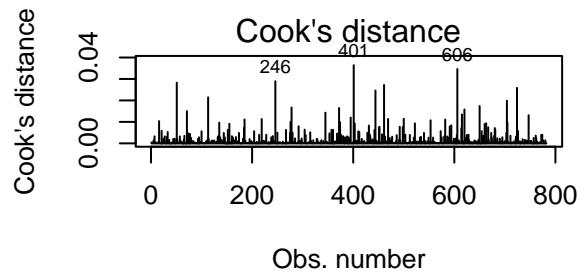
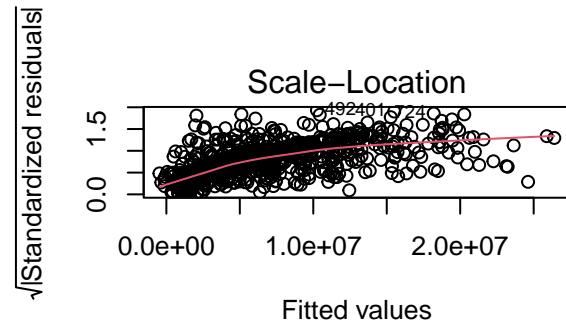
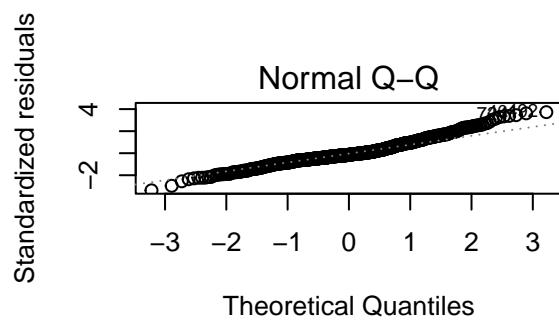
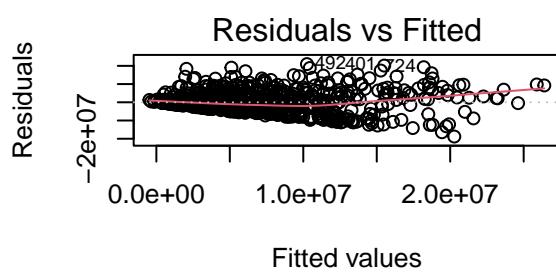
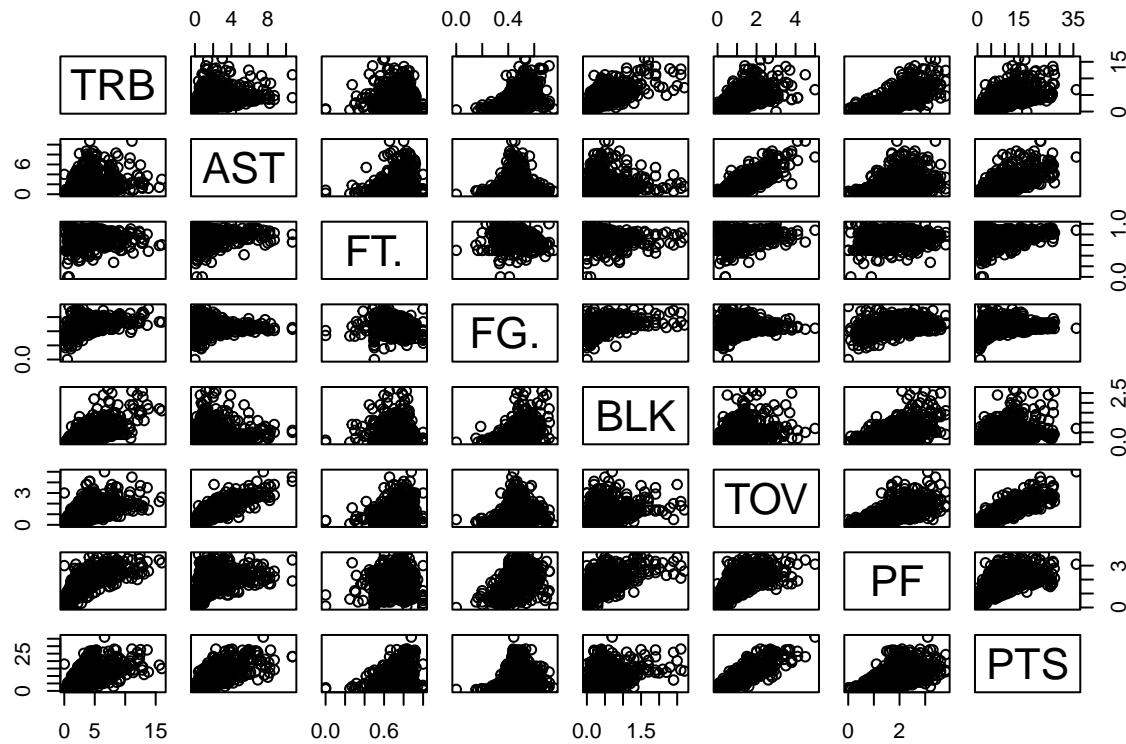
```

## FT.      -1138031   1888009  -0.603  0.54684
## FG.      -562715    2942444  -0.191  0.84839
## TRB       996903    150872   6.608  7.28e-11 *** 
## AST      1187438    245897   4.829  1.65e-06 ***
## BLK      -394730    724264  -0.545  0.58591
## TOV      -929384    736956  -1.261  0.20765
## PF       -1156753   433018  -2.671  0.00771 ** 
## PTS      491914     74977   6.561  9.80e-11 *** 
## Pos1     258572     164369   1.573  0.11610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5645000 on 772 degrees of freedom
## Multiple R-squared:  0.4361, Adjusted R-squared:  0.4295
## F-statistic: 66.33 on 9 and 772 DF,  p-value: < 2.2e-16

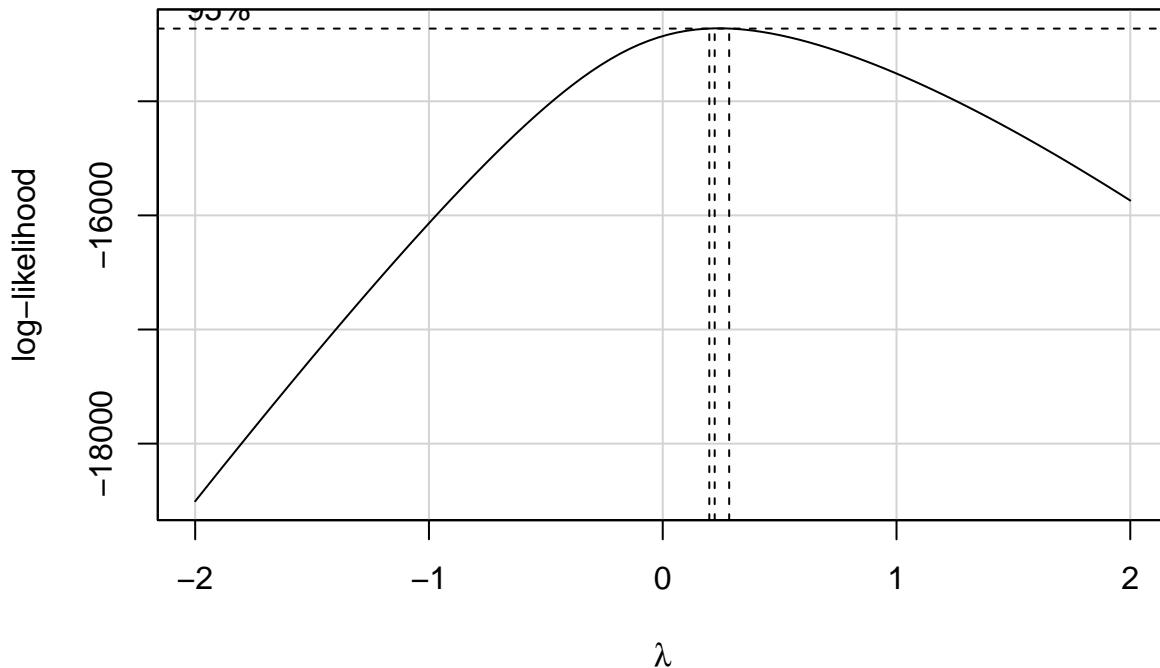
```

### Salary versus Fitted





## Profile Log-likelihood

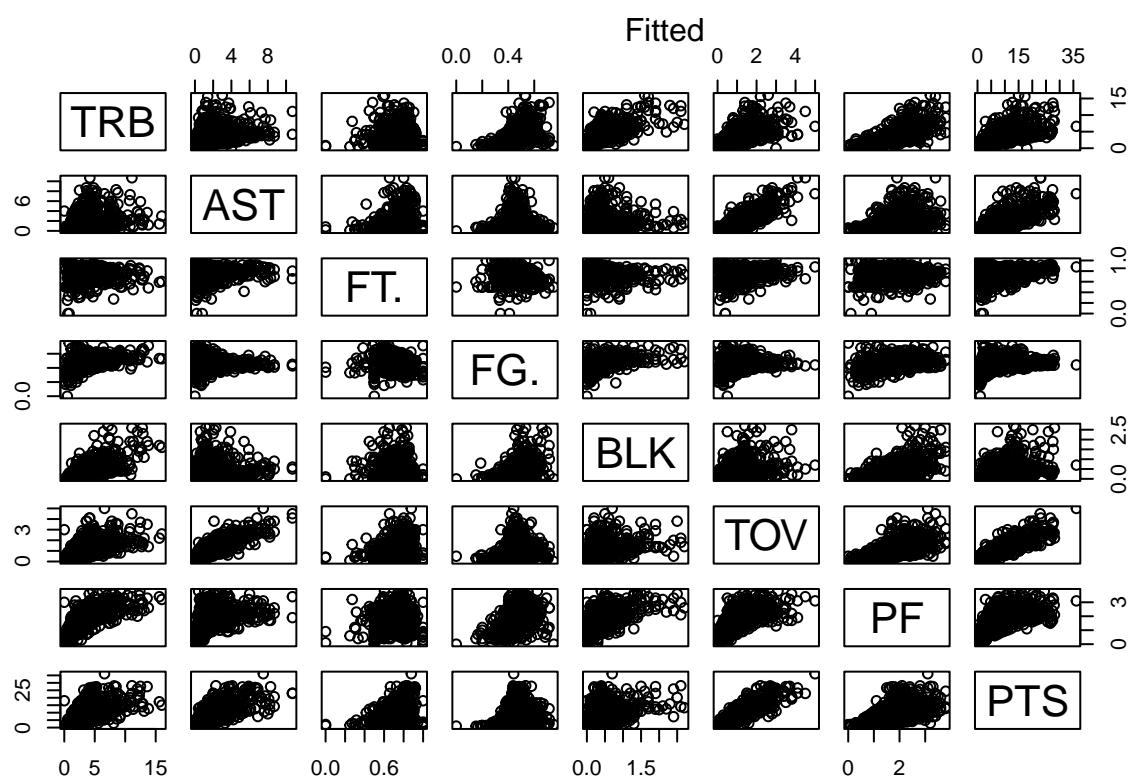
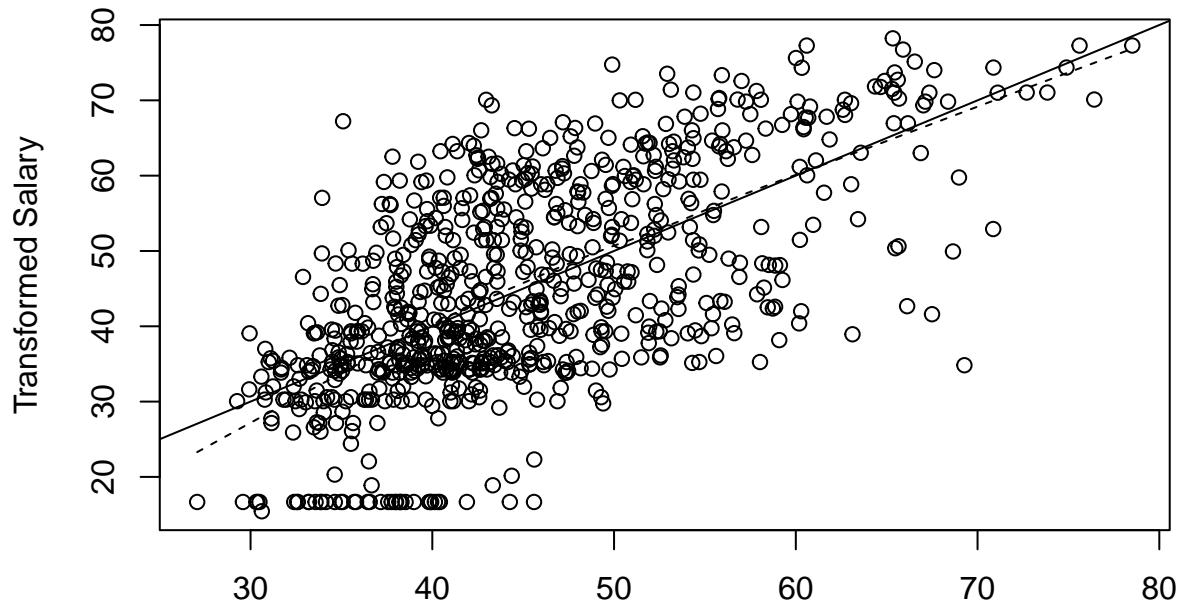


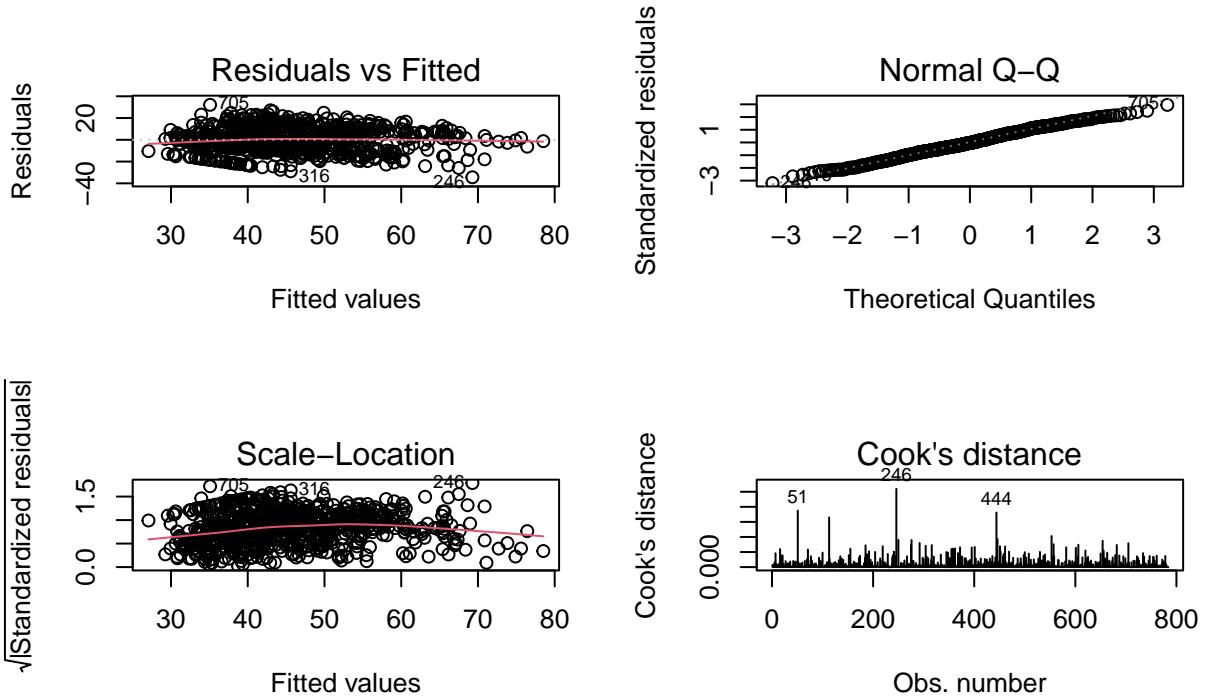
```

## 
## Call:
## lm(formula = transformed_Salary ~ FT. + FG. + TRB + AST + BLK +
##     TOV + PF + PTS + Pos1, data = test)
## 
## Residuals:
##      Min      1Q Median      3Q      Max 
## -34.448 -7.212 -0.438  7.916 32.130 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 24.6336   3.9361   6.258 6.44e-10 ***
## FT.          1.2935   3.6537   0.354 0.723405    
## FG.          4.9192   5.6942   0.864 0.387912    
## TRB          1.6013   0.2920   5.485 5.62e-08 ***
## AST          1.8078   0.4759   3.799 0.000157 ***  
## BLK          -1.9425   1.4016  -1.386 0.166170    
## TOV          -1.6512   1.4261  -1.158 0.247305    
## PF           1.0108   0.8380   1.206 0.228084    
## PTS          0.7977   0.1451   5.498 5.23e-08 ***  
## Pos1         0.5053   0.3181   1.589 0.112568    
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 10.92 on 772 degrees of freedom
## Multiple R-squared:  0.4098, Adjusted R-squared:  0.4029 
## F-statistic: 59.56 on 9 and 772 DF,  p-value: < 2.2e-16

```

### Transformed Salary versus Fitted





```

## 
## Call:
## lm(formula = transformed_Salary ~ TRB + AST + PTS, data = test)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -33.422  -7.045  -0.599   8.174  31.057 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 29.9090    0.7780  38.443 < 2e-16 ***
## TRB         1.4692    0.2032   7.229 1.17e-12 ***
## AST         1.2910    0.3248   3.975 7.71e-05 ***
## PTS         0.7970    0.1114   7.154 1.95e-12 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 10.93 on 778 degrees of freedom
## Multiple R-squared:  0.4041, Adjusted R-squared:  0.4018 
## F-statistic: 175.9 on 3 and 778 DF,  p-value: < 2.2e-16
## 
## Analysis of Variance Table
## 
## Response: transformed_Salary
##              Df Sum Sq Mean Sq F value    Pr(>F)    
## TRB          1 40162  40162 335.959 < 2.2e-16 ***
## AST          1 16787  16787 140.428 < 2.2e-16 ***
## PTS          1  6118   6118  51.182 1.947e-12 ***
## Residuals  778 93006     120                        
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## named integer(0)

```

```

## 31 33 66 74 113 137 140 357 372 401 404 431 457 469 489 522 540 552 555 606
## 31 33 66 74 113 137 140 357 372 401 404 431 457 469 489 522 540 552 555 606
## 615 665 685 703 704 747
## 615 665 685 703 704 747

## named integer(0)

## 32 85 88 152 221 231 236 270 290 317 323 332 352 401 404 440 455 457 466 475
## 32 85 88 152 221 231 236 270 290 317 323 332 352 401 404 440 455 457 466 475
## 484
## 484

## [1] "Beta 0"
## 51 113 218 219 229 246 250 262 320 334 372 381 384 408 425 459 527 537 553 566
## 51 113 218 219 229 246 250 262 320 334 372 381 384 408 425 459 527 537 553 566
## 567 582 601 631 705 736 755
## 567 582 601 631 705 736 755
## [1] "Beta 1"
## 7 21 31 77 95 113 125 135 155 174 184 195 204 219 246 306 317 372 431 444
## 7 21 31 77 95 113 125 135 155 174 184 195 204 219 246 306 317 372 431 444
## 451 452 457 461 500 532 549 553 582 668 679 726 761
## 451 452 457 461 500 532 549 553 582 668 679 726 761
## [1] "Beta 2"
## 36 135 240 246 276 304 358 372 401 428 457 469 477 498 540 553 606 747
## 36 135 240 246 276 304 358 372 401 428 457 469 477 498 540 553 606 747
## [1] "Beta 3"
## 7 16 21 33 84 132 174 222 246 275 358 360 372 388 401 403 444 452 457 469
## 7 16 21 33 84 132 174 222 246 275 358 360 372 388 401 403 444 452 457 469
## 474 477 498 540 599 605 606 679 724 739 761
## 474 477 498 540 599 605 606 679 724 739 761

##      TRB        AST        PTS
## 1.663052 1.990210 2.884788

##
## Call:
## lm(formula = transformed_Salary ~ TRB + AST + PTS + Pos1, data = training)
## 

## Residuals:
##      Min       1Q   Median       3Q      Max 
## -31.3094  -7.3441  -0.0042   8.0116  29.8696 
## 

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 26.8177    1.4802  18.118 < 2e-16 ***
## TRB          1.5845    0.2430   6.522 1.66e-10 ***
## AST          1.3794    0.3668   3.761 0.000189 *** 
## PTS          0.8071    0.1318   6.124 1.81e-09 *** 
## Pos1         0.9553    0.3660   2.610 0.009322 ** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 

## Residual standard error: 10.71 on 516 degrees of freedom
## Multiple R-squared:  0.455, Adjusted R-squared:  0.4507 
## F-statistic: 107.7 on 4 and 516 DF,  p-value: < 2.2e-16
## 
```

```
## Call:  
## lm(formula = transformed_Salary ~ TRB + AST + PTS, data = test)  
##  
## Residuals:  
##      Min       1Q   Median      3Q      Max  
## -33.422  -7.045  -0.599   8.174  31.057  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 29.9090   0.7780  38.443 < 2e-16 ***  
## TRB         1.4692   0.2032   7.229 1.17e-12 ***  
## AST         1.2910   0.3248   3.975 7.71e-05 ***  
## PTS         0.7970   0.1114   7.154 1.95e-12 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 10.93 on 778 degrees of freedom  
## Multiple R-squared:  0.4041, Adjusted R-squared:  0.4018  
## F-statistic: 175.9 on 3 and 778 DF,  p-value: < 2.2e-16
```