Data preprocessing

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

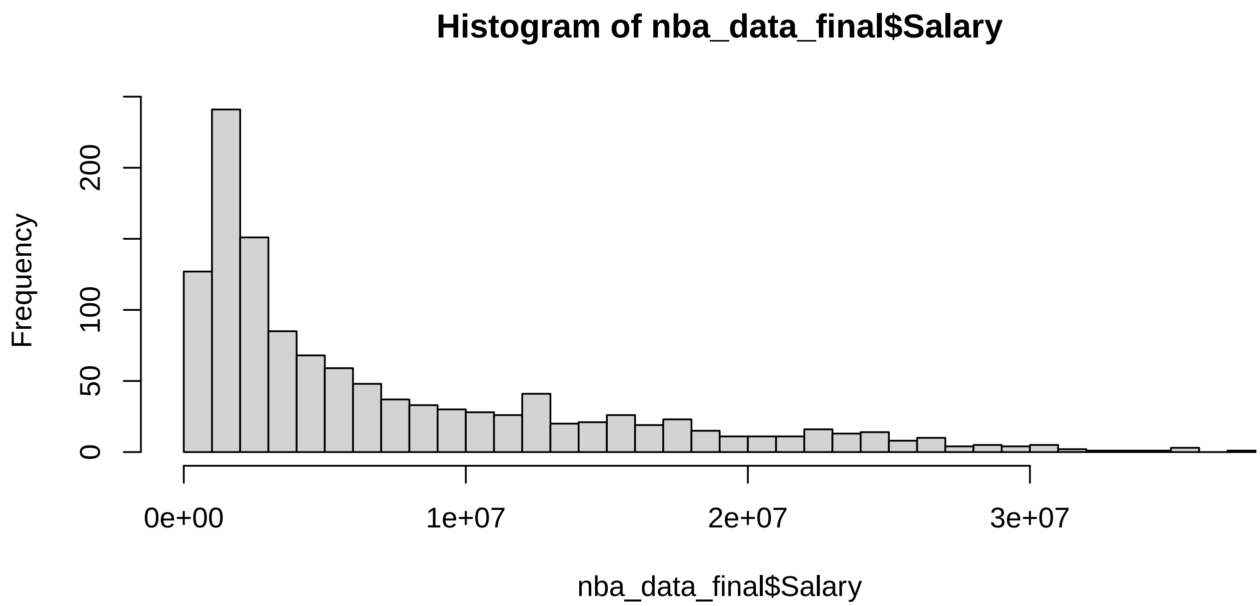
```r
library(readr)

# Read the data
nba_data <- read_csv("nba_final.csv")
```

```
## Rows: 1408 Columns: 45

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (9): Player.x, Player_ID, Pos1, Pos2, Tm, Season, Conference, Role, Play
## dbl (36): Rk, Age, G, GS, MP, FG, FGA, FG., X3P, X3PA, X3P., X2P, X2PA, X2P....
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

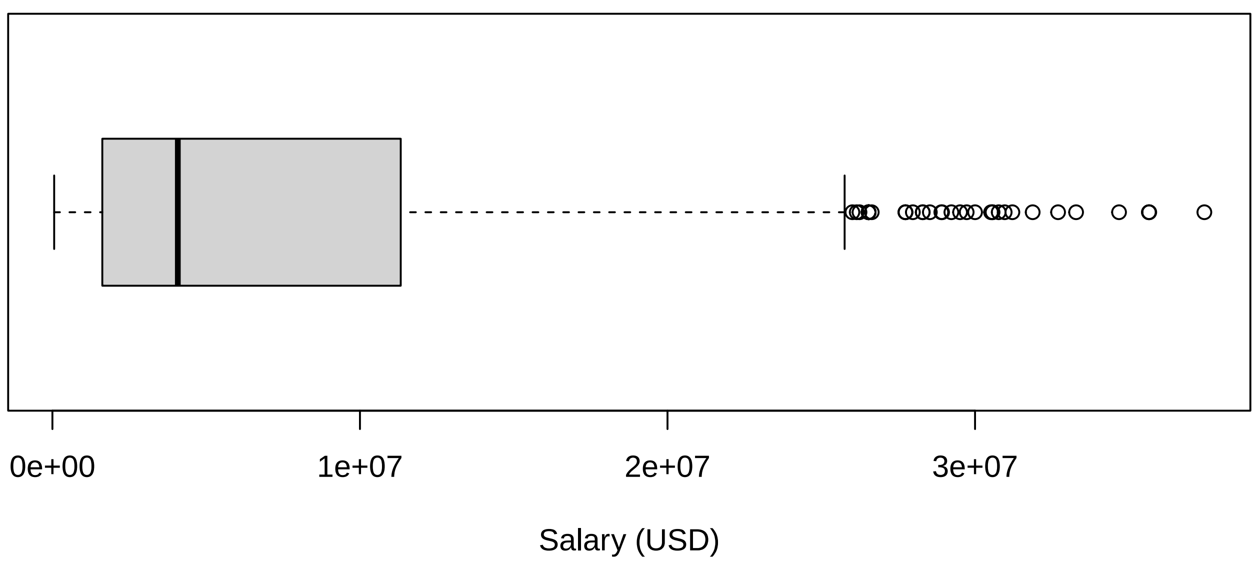
```r
# Clean the data by removing observations with NA entries
nba_data_final <- nba_data %>%
  filter(!is.na(Age),
         !is.na(PTS),
         !is.na(AST),
         !is.na(TRB),
         !is.na(Pos1),
         !is.na(mean_views),
         Salary >0)

nba_data_final <- nba_data_final %>%
  mutate(Pos1 = factor(Pos1))
```

EDA

```r
# histogram of Salary
hist(nba_data_final$Salary, breaks = 30)
```




```r
# boxplot of Salary
boxplot(nba_data_final$Salary, xlab = "Salary (USD)", horizontal = TRUE)
```




```r
# Salary vs each predictor
par(mfrow = c(2, 3),
mar = c(4, 4, 2, 1))

plot(nba_data_final$Age, nba_data_final$Salary,
     main = "Salary vs Age",
     xlab = "Age",
     ylab = "Salary (USD)",
     pch = 16,            # Solid circle
     col = "steelblue")

# Add linear regression line
abline(lm(Salary ~ Age, data = nba_data_final), col = "red", lwd = 2)
```

```r
# Scatter plot of Salary vs PTS
plot(nba_data_final$PTS, nba_data_final$Salary,
     main = "Salary vs Points (PTS)",
     xlab = "Points (PTS)",
     ylab = "Salary (USD)",
     pch = 16,
     col = "darkgreen")

# Add linear regression line
abline(lm(Salary ~ PTS, data = nba_data_final), col = "red", lwd = 2)

# Scatter plot of Salary vs AST
plot(nba_data_final$AST, nba_data_final$Salary,
     main = "Salary vs Assists (AST)",
     xlab = "Assists (AST)",
     ylab = "Salary (USD)",
     pch = 16,
     col = "purple")

# Add linear regression line
abline(lm(Salary ~ AST, data = nba_data_final), col = "red", lwd = 2)

# Scatter plot of Salary vs TRB
plot(nba_data_final$TRB, nba_data_final$Salary,
     main = "Salary vs Rebounds (TRB)",
     xlab = "Rebounds (TRB)",
     ylab = "Salary (USD)",
     pch = 16,
     col = "orange")

# Add linear regression line
abline(lm(Salary ~ TRB, data = nba_data_final), col = "red", lwd = 2)

# Scatter plot of Salary vs Mean Views
plot(nba_data_final$mean_views, nba_data_final$Salary,
     main = "Salary vs Mean Views",
     xlab = "Mean Views",
     ylab = "Salary (USD)",
     pch = 16,
     col = "brown")

# Add linear regression line
abline(lm(Salary ~ mean_views, data = nba_data_final), col = "red", lwd = 2)

# Boxplot of Salary by Position
boxplot(Salary ~ Pos1, data = nba_data_final,
        main = "Salary by Position",
        xlab = "Position",
        ylab = "Salary (USD)",
        col = "lightcoral",
        border = "darkred")
```

Checking multicolinearity

```r
predictor_fit <- lm(PTS ~ Age + AST + TRB + Pos1 + mean_views, data = nba_data_final)
summary(predictor_fit)
```

```
##
## Call:
## lm(formula = PTS ~ Age + AST + TRB + Pos1 + mean_views, data = nba_data_final)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.777  -1.854  -0.288   1.541  13.743
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2941110  0.6926241   0.425   0.6712
## Age         -0.0568211  0.0223575  -2.541   0.0112 *
## AST          1.6352396  0.0811068  20.162  < 2e-16 ***
## TRB          1.2538118  0.0537110  23.344  < 2e-16 ***
## Pos1PF       1.7681028  0.3023840   5.847 6.42e-09 ***
## Pos1PG       1.1330273  0.4480789   2.529   0.0116 *
## Pos1SF       2.8422595  0.3486863   8.151 8.91e-16 ***
## Pos1SG       4.0517314  0.3554429  11.399  < 2e-16 ***
## mean_views   0.0003406  0.0000495   6.881 9.53e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.308 on 1210 degrees of freedom
## Multiple R-squared:  0.7045, Adjusted R-squared:  0.7026
## F-statistic: 360.6 on 8 and 1210 DF,  p-value: < 2.2e-16
```

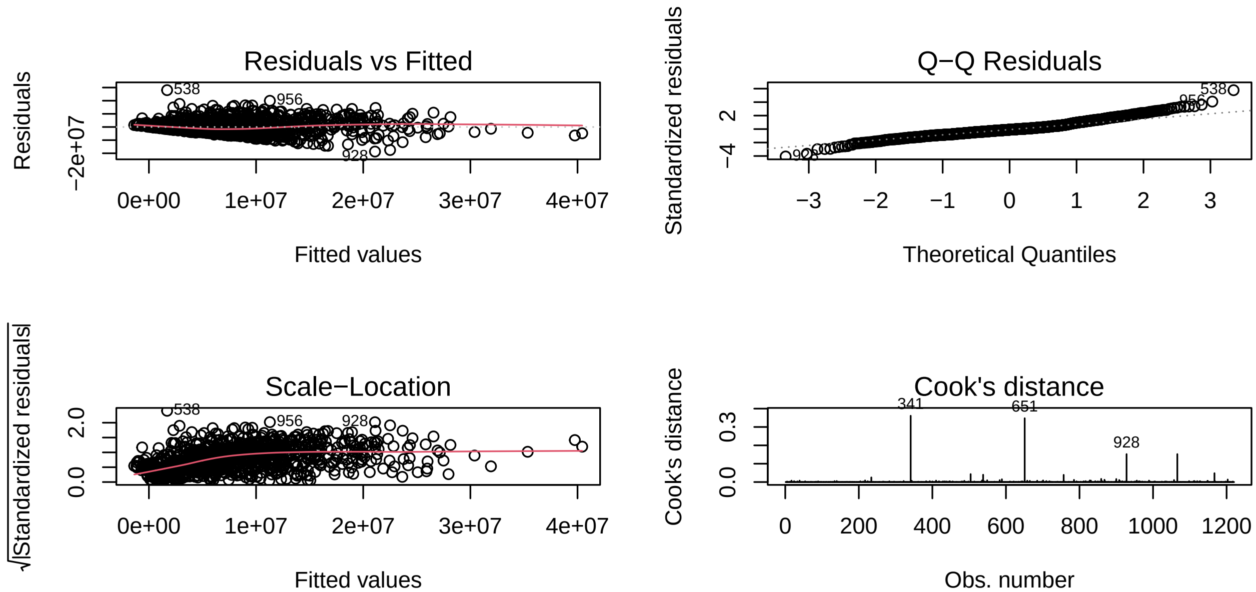
Fit the multiple regression model

```
nba_model <- lm(Salary ~ Age * PTS + AST + TRB + Pos1 * mean_views, data = nba_data_final)

# View the summary of the model
summary(nba_model)
```

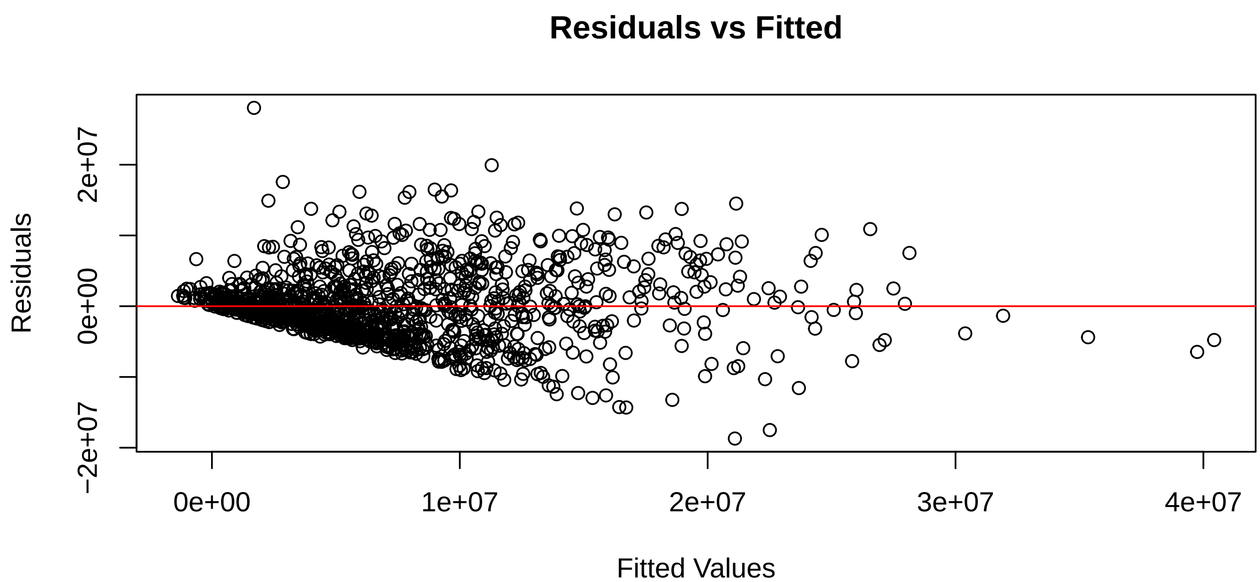
```
##
## Call:
## lm(formula = Salary ~ Age * PTS + AST + TRB + Pos1 * mean_views,
##     data = nba_data_final)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -18703871  -2962849   -439232   2199038  28031207
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           648686.9  1665198.7   0.390  0.69693
## Age                   -26045.9    60542.9  -0.430  0.66712
## PTS                 -1328992.3   168174.5  -7.902 6.13e-15 ***
## AST                   705474.5   140025.7   5.038 5.42e-07 ***
## TRB                   424777.2    99265.2   4.279 2.02e-05 ***
## Pos1PF               -537215.7   543403.3  -0.989  0.32305
## Pos1PG              -2310005.5   706790.3  -3.268  0.00111 **
## Pos1SF               -165652.1   596387.6  -0.278  0.78124
## Pos1SG               -685450.3   617559.6  -1.110  0.26725
## mean_views               677.6      337.4   2.008  0.04486 *
## Age:PTS                70212.5     6328.9  11.094  < 2e-16 ***
## Pos1PF:mean_views       -536.5      370.3  -1.449  0.14767
## Pos1PG:mean_views       -558.5      352.2  -1.586  0.11308
## Pos1SF:mean_views       -608.0      351.4  -1.730  0.08385 .
## Pos1SG:mean_views       -894.5      414.8  -2.156  0.03124 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4900000 on 1204 degrees of freedom
## Multiple R-squared:  0.586,  Adjusted R-squared:  0.5812
## F-statistic: 121.7 on 14 and 1204 DF,  p-value: < 2.2e-16
```

Diagnostic plots

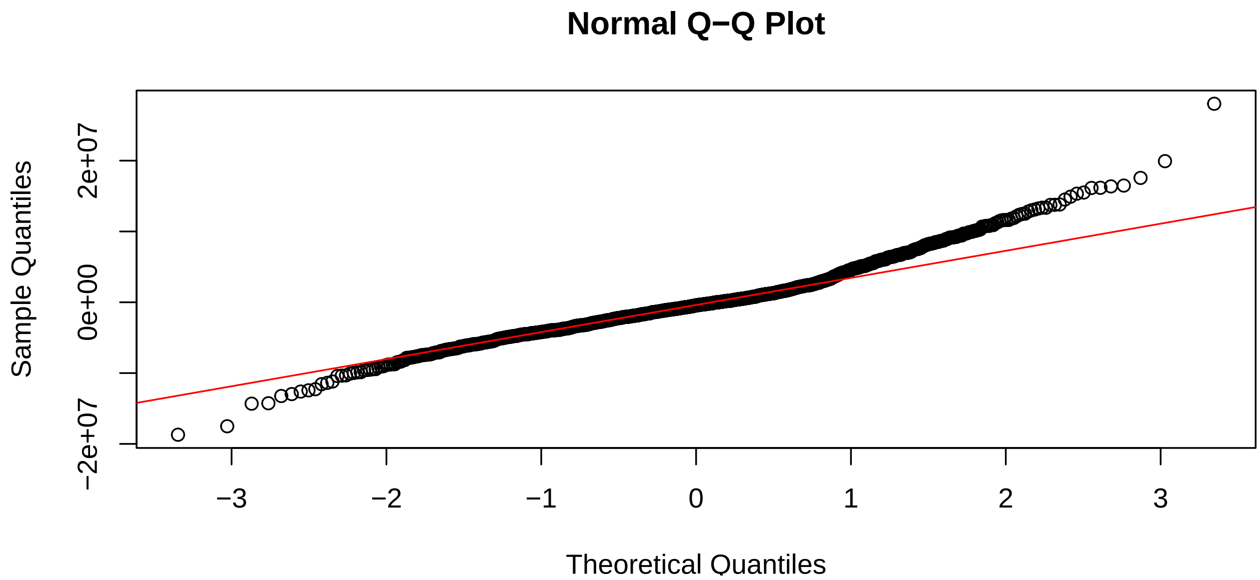
```
par(mfrow = c(2, 2))
plot(nba_model, which = c(1, 2, 3, 4))
```



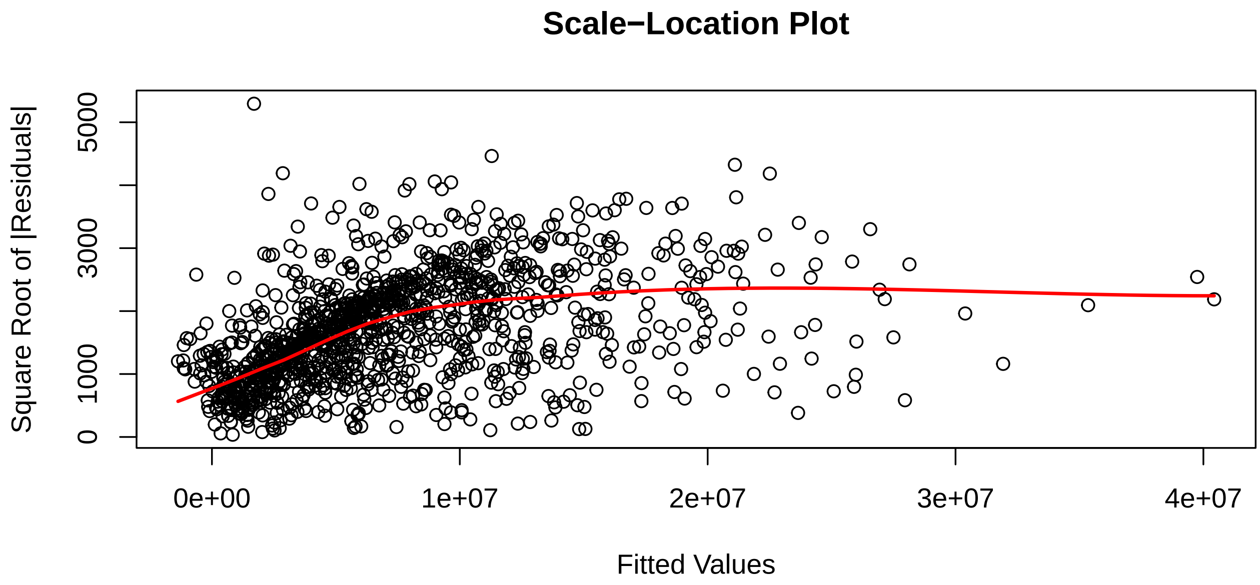

```r
# Closer look at diagnostic plots
par(mfrow = c(1,1))
residuals <- resid(nba_model)
fitted_values <- fitted(nba_model)

# Residuals vs Fitted
plot(fitted_values, residuals,
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, col = "red")  # Horizontal line at zero
```




```r
# Normal Q-Q Plot
qqnorm(residuals, main = "Normal Q-Q Plot")
qqline(residuals, col = "red")
```



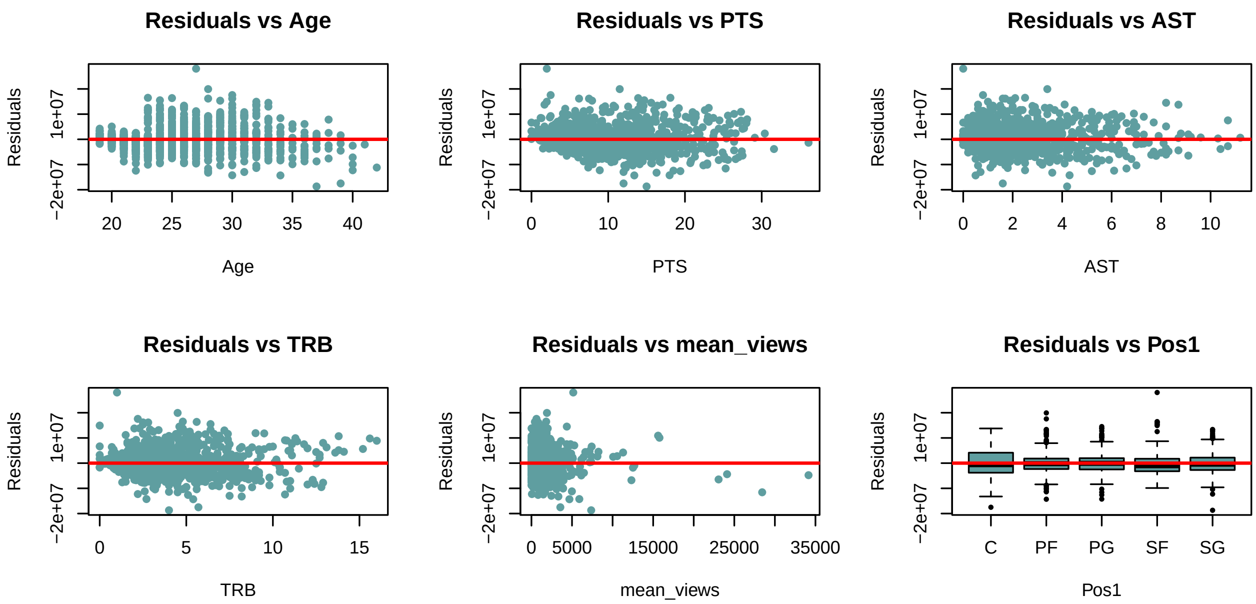

```
# Scale-Location Plot
sqrt_std_resid <- sqrt(abs(residuals))  # Square root of standardized residuals
plot(fitted_values, sqrt_std_resid,
     xlab = "Fitted Values",
     ylab = "Square Root of |Residuals|",
     main = "Scale-Location Plot")
lines(loess.smooth(fitted_values, sqrt_std_resid), col = "red", lwd = 2)
```



residuals vs each predictor

```
predictors <- nba_data_final[, c("Age", "PTS", "AST", "TRB", "mean_views", "Pos1")]  # Select predictor
par(mfrow = c(2, 3))
residuals <- resid(nba_model)
for (predictor in names(predictors)) {
  plot(predictors[[predictor]], residuals,
       xlab = predictor,
       ylab = "Residuals",
       main = paste("Residuals vs", predictor),
```

```
         col = "cadetblue", pch = 16)
  abline(h = 0, col = "red", lwd = 2)
}
```

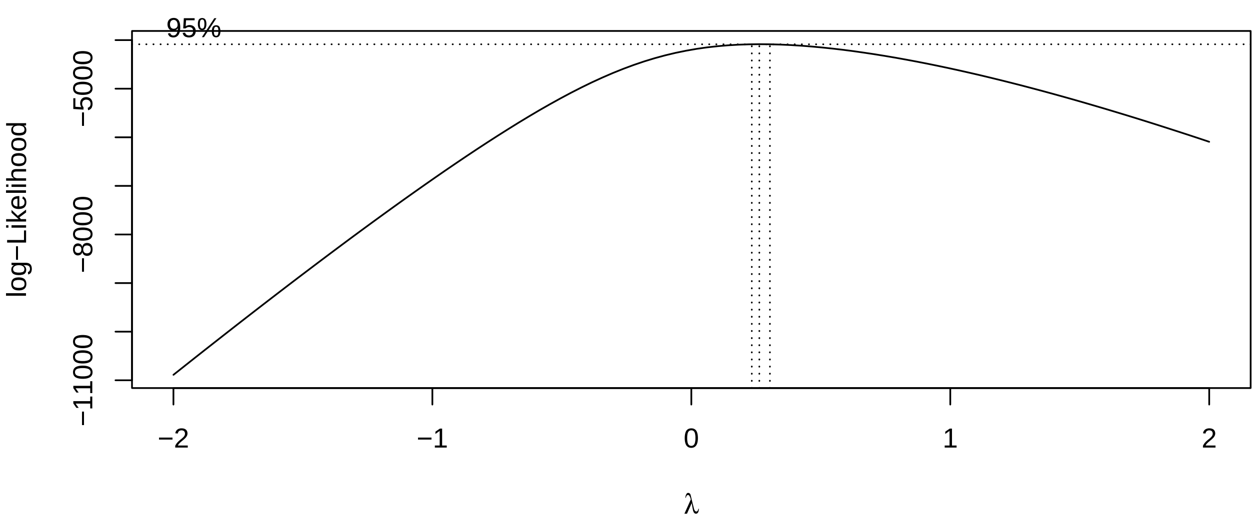


Transformations

```
# Perform Box-Cox transformation
library(MASS)
```

```
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##     select
```

```
# Apply Box-Cox transformation to find optimal lambda
boxcox_fit <- boxcox(nba_model, lambda = seq(-2, 2, 0.1))
```




```
lambda_optimal <- boxcox_fit$x[which.max(boxcox_fit$y)]  # Extract the optimal lambda
lambda_optimal
```

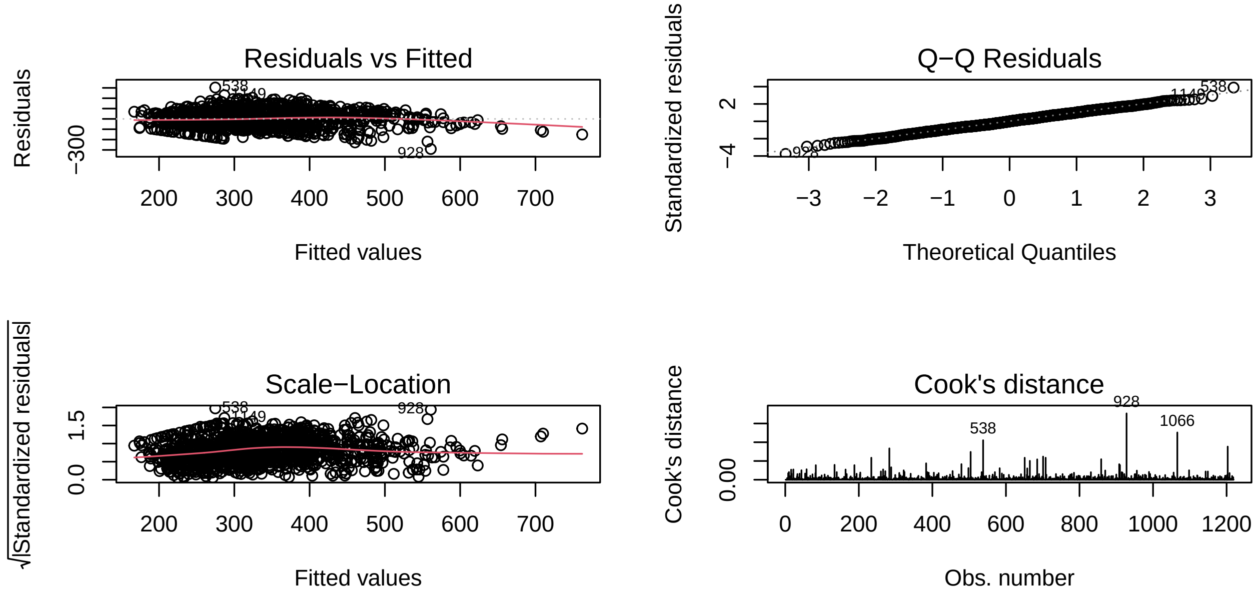
```
## [1] 0.2626263
```

```r
# Apply the transformation with the optimal lambda
lambda <- 0.3
if (lambda == 0) {
  nba_data_final$Salary_transformed <- log(nba_data_final$Salary)
} else {
  nba_data_final$Salary_transformed <- (nba_data_final$Salary^lambda - 1) / lambda
}

# transform AST
nba_data_final$Age_sq <- nba_data_final$Age^2

# transform mean_views
nba_data_final$log_mean_views <- log(nba_data_final$mean_views)

# Fit the model with the transformed variables
transformed_fit <- lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 * log_mean_views,
                      data = nba_data_final)

# Summary of the transformed model
summary(transformed_fit)
```

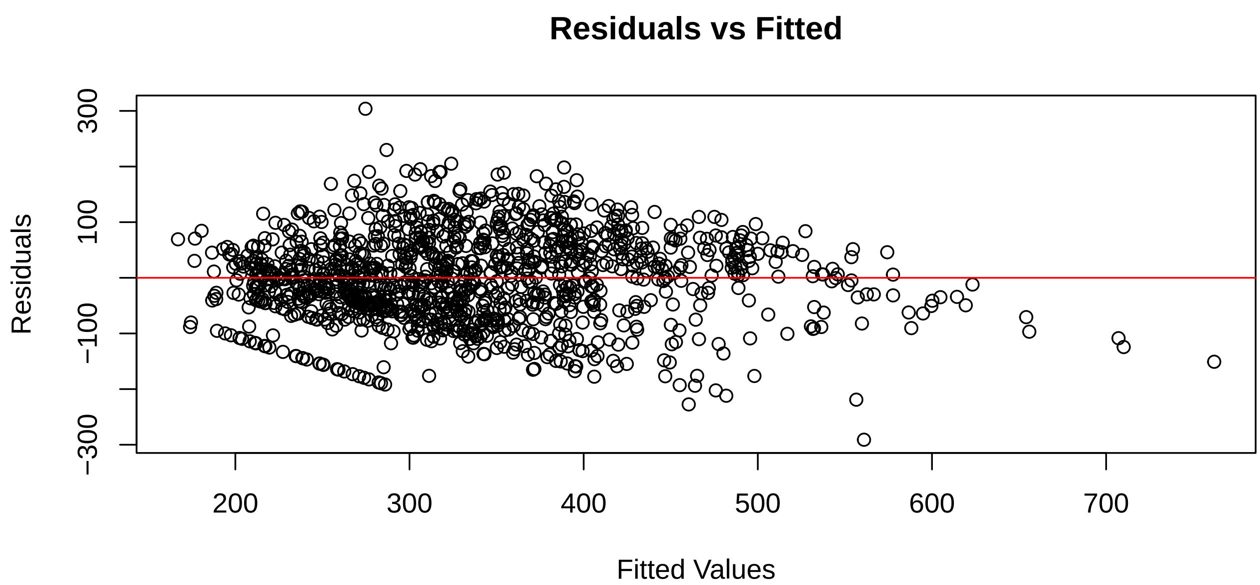
```
##
## Call:
## lm(formula = Salary_transformed ~ Age_sq * PTS + AST + TRB +
##     Pos1 * log_mean_views, data = nba_data_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -291.014  -51.107   -0.975   54.702  303.840
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           126.696045  26.874071   4.714 2.71e-06 ***
## Age_sq                  0.103881   0.017436   5.958 3.35e-09 ***
## PTS                     1.563073   1.422574   1.099   0.2721
## AST                     9.592956   2.252890   4.258 2.22e-05 ***
## TRB                     7.784957   1.567034   4.968 7.74e-07 ***
## Pos1PF                  8.949916  30.833570   0.290   0.7717
## Pos1PG                -53.954931  30.606064  -1.763   0.0782 .
## Pos1SF                -34.251729  29.971995  -1.143   0.2534
## Pos1SG                 -3.415460  30.296690  -0.113   0.9103
## log_mean_views          5.417214   3.835489   1.412   0.1581
## Age_sq:PTS              0.008167   0.001816   4.498 7.52e-06 ***
## Pos1PF:log_mean_views  -3.770668   5.011034  -0.752   0.4519
## Pos1PG:log_mean_views   2.207409   4.887976   0.452   0.6516
## Pos1SF:log_mean_views   4.376313   4.844863   0.903   0.3666
## Pos1SG:log_mean_views  -1.723499   4.940541  -0.349   0.7273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 78.88 on 1204 degrees of freedom
## Multiple R-squared:  0.5562, Adjusted R-squared:  0.551
## F-statistic: 107.8 on 14 and 1204 DF,  p-value: < 2.2e-16
```

Verify assumptions for transformed fit

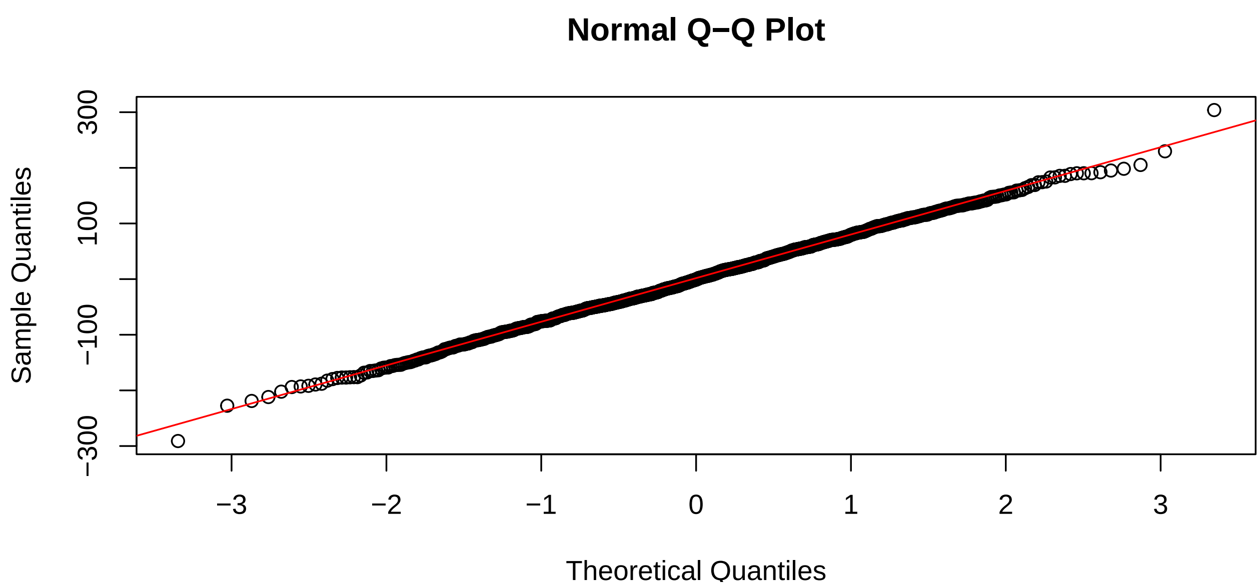
```r
# Diagnostic plots
par(mfrow = c(2, 2))
plot(transformed_fit, which = c(1, 2, 3, 4))
```



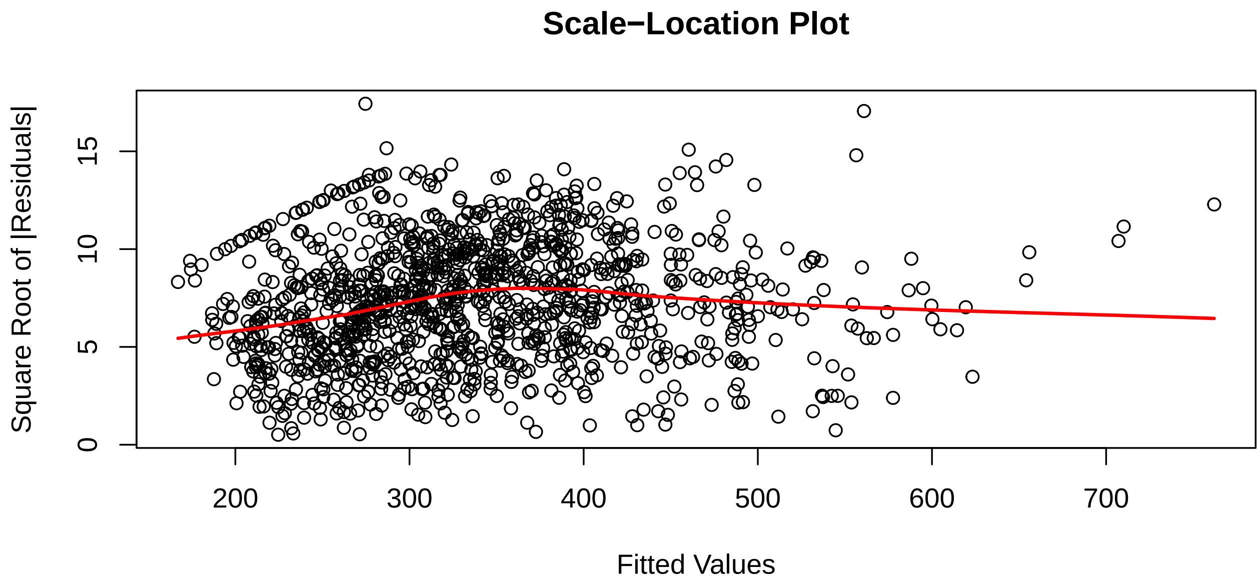

```r
# Closer look at diagnostic plots
par(mfrow = c(1,1))
residuals <- resid(transformed_fit)
fitted_values <- fitted(transformed_fit)
sresidual_values <- rstandard(transformed_fit)

# Residuals vs Fitted
plot(fitted_values, residuals,
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, col = "red")  # Horizontal line at zero
```



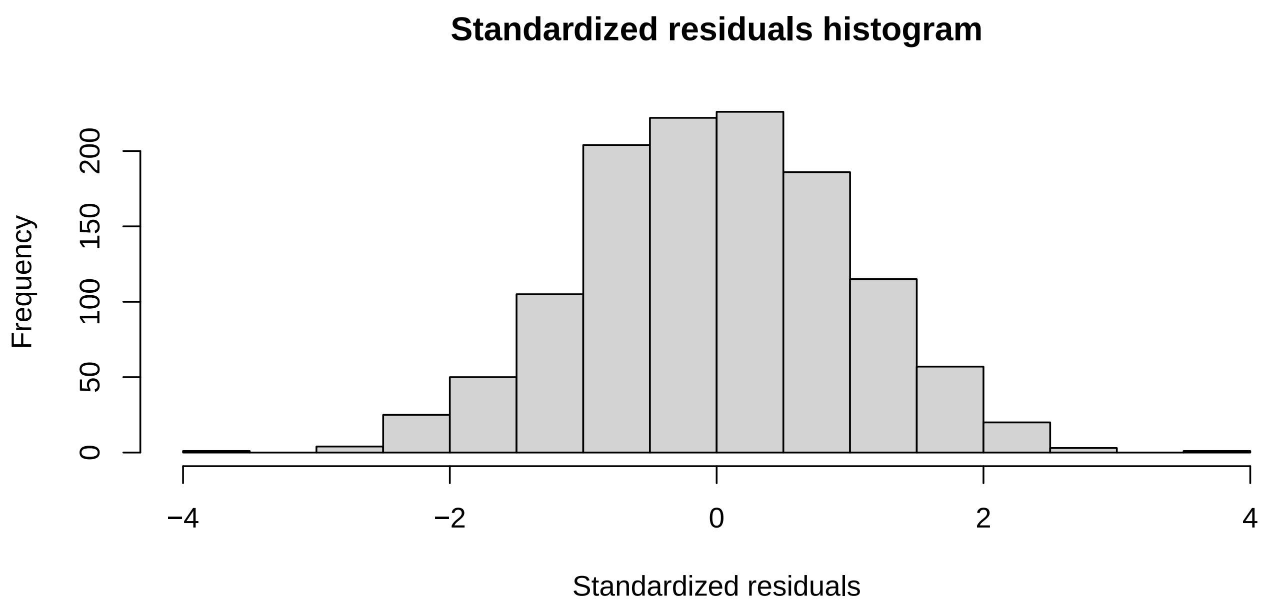

```
# Normal Q-Q Plot
qqnorm(residuals, main = "Normal Q-Q Plot")
qqline(residuals, col = "red")
```



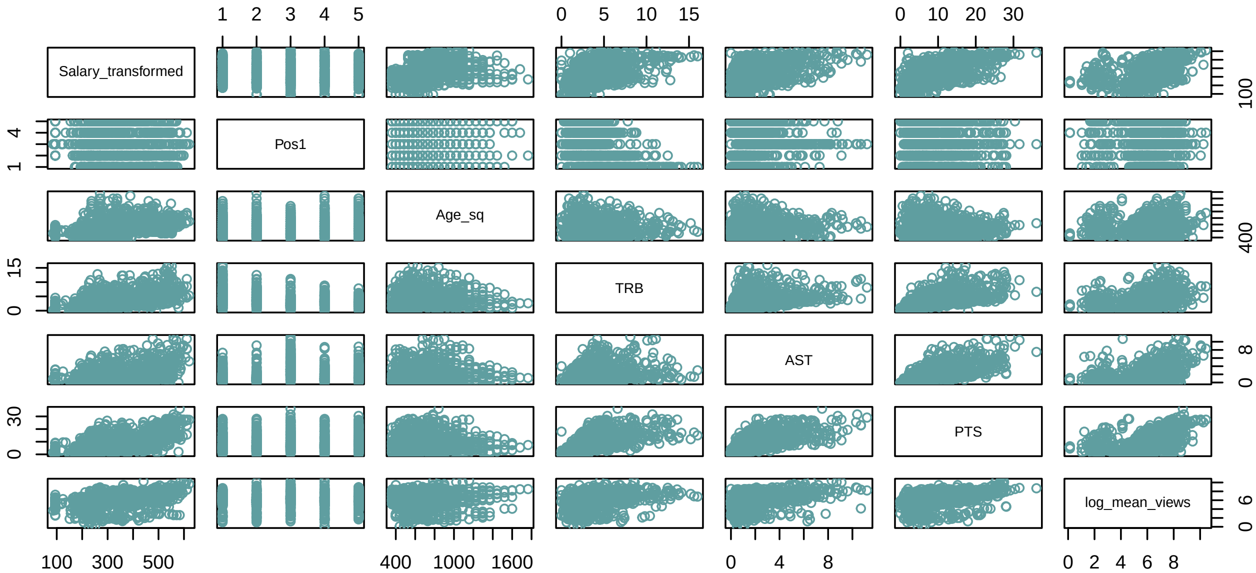

```
# Scale-Location Plot
sqrt_std_resid <- sqrt(abs(residuals))  # Square root of standardized residuals
plot(fitted_values, sqrt_std_resid,
     xlab = "Fitted Values",
     ylab = "Square Root of |Residuals|",
     main = "Scale-Location Plot")
lines(loess.smooth(fitted_values, sqrt_std_resid), col = "red", lwd = 2)
```




```r
hist(sresidual_values,
     main = "Standardized residuals histogram",
     xlab = "Standardized residuals")
```

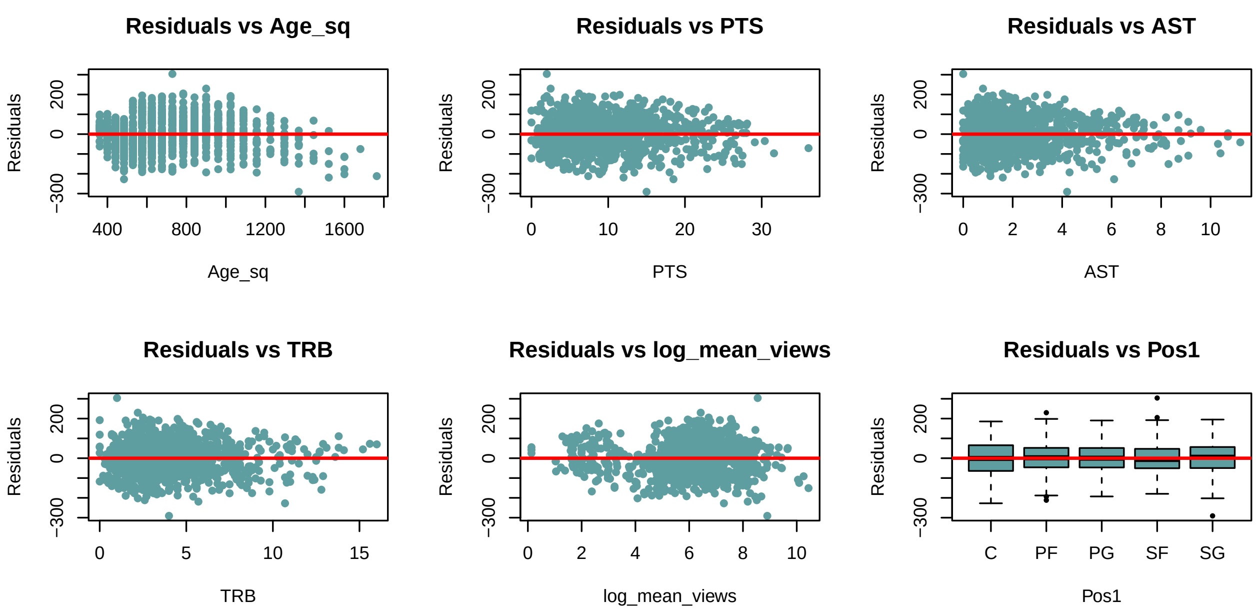



```r
plot(nba_data_final[, c(46, 4, 47, 26, 27, 32, 48)], col="cadetblue")
```



residuals vs each predictor

```
predictors <- nba_data_final[, c("Age_sq", "PTS", "AST", "TRB", "log_mean_views", "Pos1")]  # Select pr
par(mfrow = c(2, 3))

residuals <- resid(transformed_fit)

for (predictor in names(predictors)) {
  plot(predictors[[predictor]], residuals,
       xlab = predictor,
       ylab = "Residuals",
       main = paste("Residuals vs", predictor),
       col = "cadetblue", pch = 16)
  abline(h = 0, col = "red", lwd = 2)
}
```



full F-test

```r
# Perform ANOVA to compare with the null model
fit_null <- lm(Salary_transformed ~ 1, data = nba_data_final)  # Null model with only intercept
anova(fit_null, transformed_fit)
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ 1
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 * log_mean_views
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1   1218 16879008
## 2   1204  7490792 14   9388216 107.78 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Partial F-tests

```r
# dropping Age_Sq:PTS
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 * log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ Age_sq + PTS + AST + TRB + Pos1 * log_mean_views,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # is significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ Age_sq + PTS + AST + TRB + Pos1 * log_mean_views
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 * log_mean_views
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   1205 7616678
## 2   1204 7490792  1    125886 20.234 7.515e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# dropping Pos1:log_mean_views
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 * log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # not significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 * log_mean_views
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1   1208 7516863
## 2   1204 7490792  4     26071 1.0476 0.3813
```

```r
# dropping Pos1
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + log_mean_views,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # is significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ Age_sq * PTS + AST + TRB + log_mean_views
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views
##   Res.Df     RSS Df Sum of Sq      F   Pr(>F)
## 1   1212 7626125
## 2   1208 7516863  4    109262 4.3898 0.001582 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# dropping log_mean_views
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # is significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   1209 7600259
## 2   1208 7516863  1     83396 13.402 0.0002622 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# dropping TRB
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + Pos1 + log_mean_views,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # is significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ Age_sq * PTS + AST + Pos1 + log_mean_views
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   1209 7671910
## 2   1208 7516863  1    155047 24.917 6.863e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# dropping AST
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ Age_sq * PTS + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # is significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ Age_sq * PTS + TRB + Pos1 + log_mean_views
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1   1209 7634853
## 2   1208 7516863  1    117990 18.962 1.447e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# dropping Age_sq
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # is significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ PTS + AST + TRB + Pos1 + log_mean_views
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   1210 9566390
## 2   1208 7516863  2   2049527 164.69 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# dropping PTS
full_fit = lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
reduced_fit = lm(Salary_transformed ~ PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
anova(reduced_fit, full_fit) # is significant
```

```
## Analysis of Variance Table
##
## Model 1: Salary_transformed ~ PTS + AST + TRB + Pos1 + log_mean_views
## Model 2: Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   1210 9566390
## 2   1208 7516863  2   2049527 164.69 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Final model

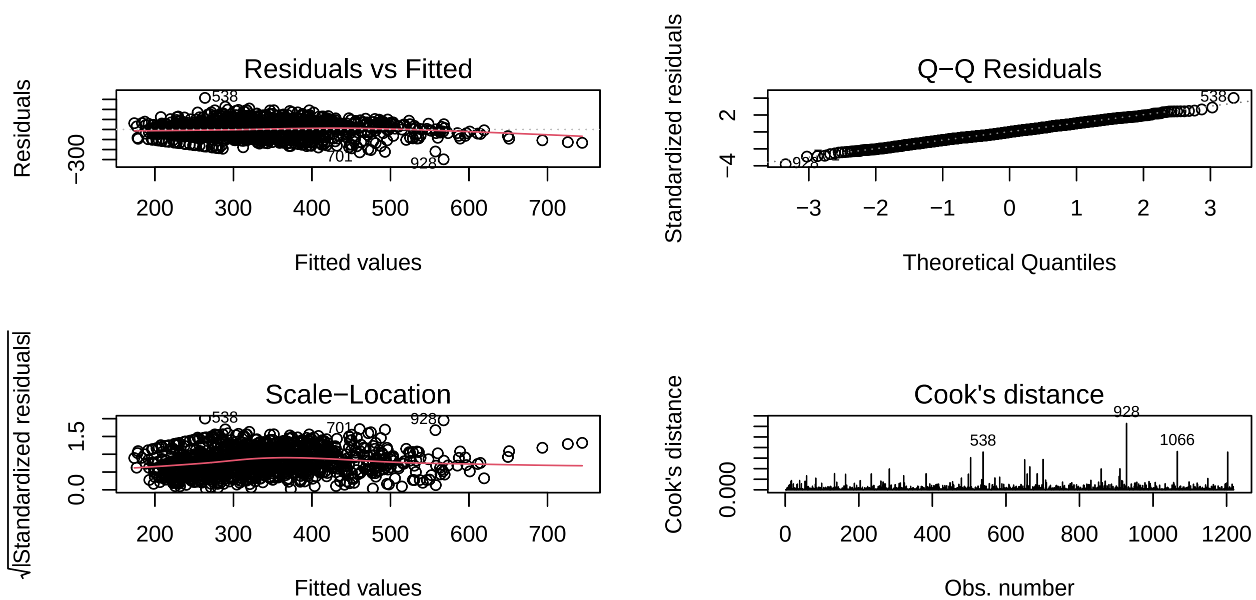
```r
final_model <- lm(Salary_transformed ~ Age_sq * PTS + AST + TRB + Pos1 + log_mean_views,
                      data = nba_data_final)
summary(final_model)
```

```
##
## Call:
## lm(formula = Salary_transformed ~ Age_sq * PTS + AST + TRB +
##     Pos1 + log_mean_views, data = nba_data_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -297.720  -50.894   -0.838   55.624  314.776
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)     125.278572  16.828335   7.445 1.84e-13 ***
## Age_sq            0.102745   0.017398   5.905 4.57e-09 ***
## PTS               1.505158   1.414674   1.064 0.287559
## AST               9.694878   2.226404   4.355 1.45e-05 ***
## TRB               7.700196   1.542606   4.992 6.86e-07 ***
## Pos1PF          -12.698336   7.349789  -1.728 0.084295 .
## Pos1PG          -40.944888  10.733928  -3.815 0.000143 ***
## Pos1SF           -9.080352   8.578649  -1.058 0.290047
## Pos1SG          -13.662659   8.934200  -1.529 0.126463
## log_mean_views    5.842177   1.595834   3.661 0.000262 ***
## Age_sq:PTS        0.008243   0.001810   4.553 5.82e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 78.88 on 1208 degrees of freedom
## Multiple R-squared:  0.5547, Adjusted R-squared:  0.551
## F-statistic: 150.5 on 10 and 1208 DF,  p-value: < 2.2e-16
```

Verifying assumptions for final model

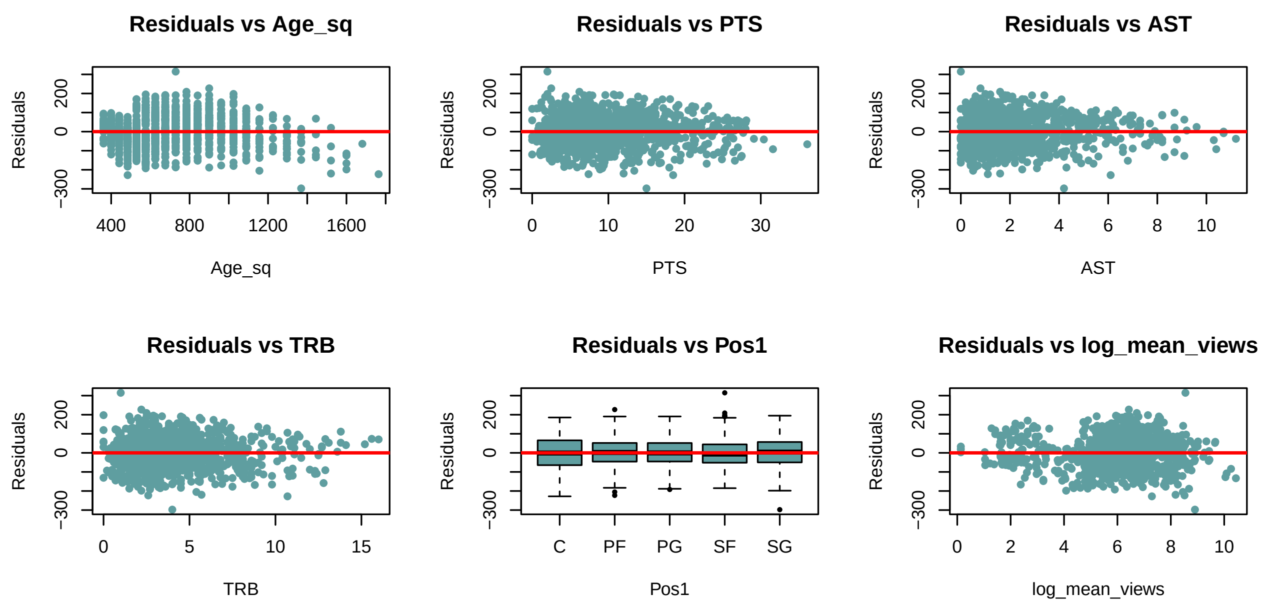
```r
par(mfrow = c(2, 2))
plot(final_model, which = c(1, 2, 3, 4))
```



residuals vs each predictor

```r
predictors <- nba_data_final[, c("Age_sq", "PTS", "AST", "TRB", "Pos1", "log_mean_views")]  # Select pr
par(mfrow = c(2, 3))

residuals <- resid(final_model)

for (predictor in names(predictors)) {
  plot(predictors[[predictor]], residuals,
       xlab = predictor,
       ylab = "Residuals",
```

```r
       main = paste("Residuals vs", predictor),
       col = "cadetblue", pch = 16)
  abline(h = 0, col = "red", lwd = 2)
}
```



VIF

```r
library(car)
```

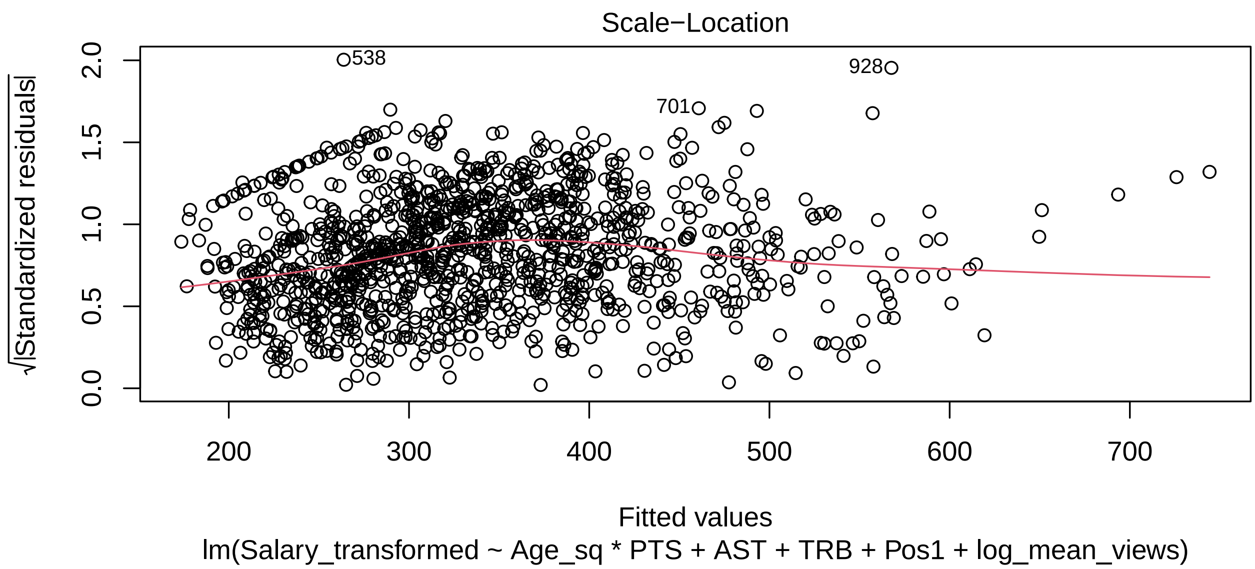
```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
vif(final_model)
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif

##                      GVIF Df GVIF^(1/(2*Df))
## Age_sq           3.410453  1        1.846741
## PTS             14.412524  1        3.796383
## AST              3.217689  1        1.793792
## TRB              2.961150  1        1.720799
## Pos1             3.118050  4        1.152751
## log_mean_views   1.267110  1        1.125660
## Age_sq:PTS      15.174677  1        3.895469
```

Outliers, influential points, leverage points

```r
# outliers
plot(final_model, which = c(3))
```



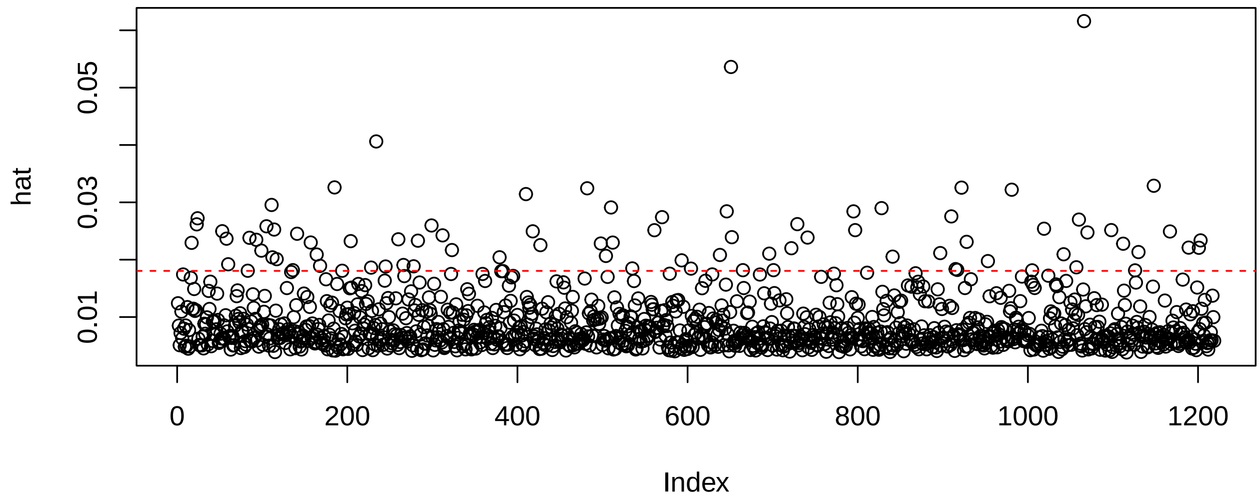

```
std_residuals <- rstandard(final_model)
outliers <- which(abs(std_residuals) > 4)
outliers
```

```
## 538
## 538
```

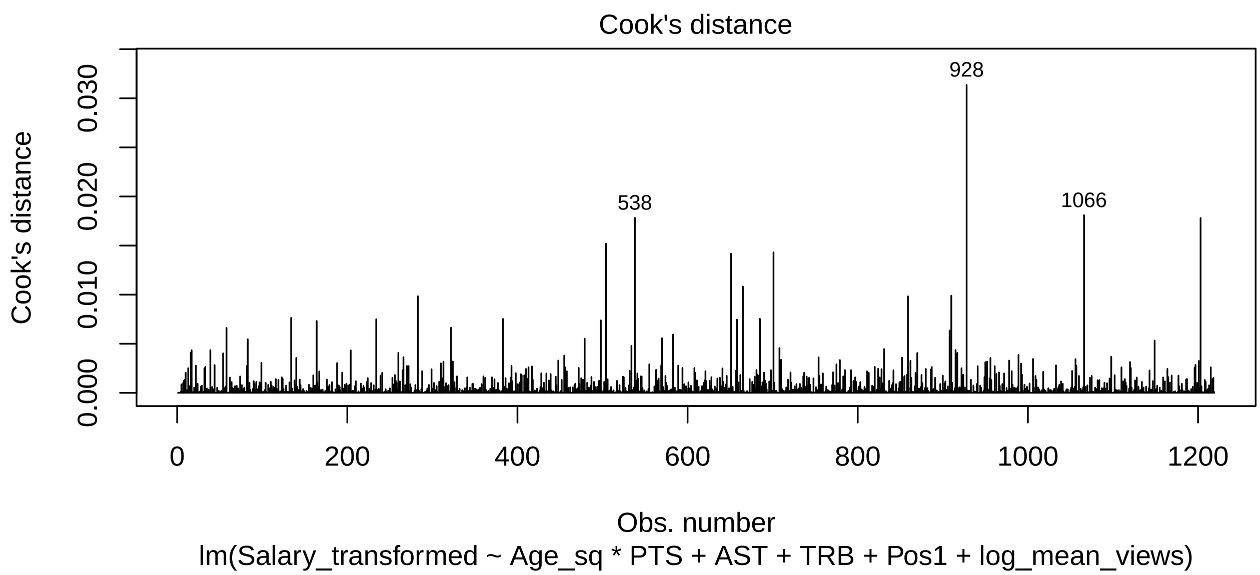
```
nba_data_final[538,]
```

```
## # A tibble: 1 x 48
##      Rk Player.x Player_ID Pos1  Pos2    Age Tm        G    GS    MP    FG   FGA
##   <dbl> <chr>    <chr>     <fct> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   207 Gordon ~ haywago01 SF    <NA>     27 BOS       1     1     5     1     2
## # i 36 more variables: FG. <dbl>, X3P <dbl>, X3PA <dbl>, X3P. <dbl>, X2P <dbl>,
## #   X2PA <dbl>, X2P. <dbl>, eFG. <dbl>, FT <dbl>, FTA <dbl>, FT. <dbl>,
## #   ORB <dbl>, DRB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>,
## #   TOV <dbl>, PF <dbl>, PTS <dbl>, Salary <dbl>, mean_views <dbl>,
## #   Season <chr>, Conference <chr>, Role <chr>, Fvot <dbl>, FRank <dbl>,
## #   Pvot <dbl>, PRank <dbl>, Mvot <dbl>, MRank <dbl>, Score <dbl>, Play <chr>,
## #   Salary_transformed <dbl>, Age_sq <dbl>, log_mean_views <dbl>
```

```
# high leverage points
hat <- hatvalues(final_model)
threshold <- 2 * mean(hat)
plot(hat)
abline(h = threshold, col = "red", lty = 2)
```




```r
# influential points
plot(final_model, which = c(4))
```




```r
nba_data_final[538,]
```

```
## # A tibble: 1 x 48
##      Rk Player.x Player_ID Pos1  Pos2    Age Tm        G    GS    MP    FG   FGA
##   <dbl> <chr>    <chr>     <fct> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   207 Gordon ~ haywago01 SF    <NA>     27 BOS       1     1     5     1     2
## # i 36 more variables: FG. <dbl>, X3P <dbl>, X3PA <dbl>, X3P. <dbl>, X2P <dbl>,
## #   X2PA <dbl>, X2P. <dbl>, eFG. <dbl>, FT <dbl>, FTA <dbl>, FT. <dbl>,
## #   ORB <dbl>, DRB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>,
## #   TOV <dbl>, PF <dbl>, PTS <dbl>, Salary <dbl>, mean_views <dbl>,
## #   Season <chr>, Conference <chr>, Role <chr>, Fvot <dbl>, FRank <dbl>,
## #   Pvot <dbl>, PRank <dbl>, Mvot <dbl>, MRank <dbl>, Score <dbl>, Play <chr>,
## #   Salary_transformed <dbl>, Age_sq <dbl>, log_mean_views <dbl>
```

```r
nba_data_final[928,]
```

```
## # A tibble: 1 x 48
##      Rk Player.x Player_ID Pos1  Pos2    Age Tm        G    GS    MP    FG   FGA
```

```
##   <dbl> <chr>    <chr>     <fct> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   493 Dwyane ~ wadedw01  SG    <NA>     37 MIA      72     2  26.2   5.8  13.3
## # i 36 more variables: FG. <dbl>, X3P <dbl>, X3PA <dbl>, X3P. <dbl>, X2P <dbl>,
## #   X2PA <dbl>, X2P. <dbl>, eFG. <dbl>, FT <dbl>, FTA <dbl>, FT. <dbl>,
## #   ORB <dbl>, DRB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>,
## #   TOV <dbl>, PF <dbl>, PTS <dbl>, Salary <dbl>, mean_views <dbl>,
## #   Season <chr>, Conference <chr>, Role <chr>, Fvot <dbl>, FRank <dbl>,
## #   Pvot <dbl>, PRank <dbl>, Mvot <dbl>, MRank <dbl>, Score <dbl>, Play <chr>,
## #   Salary_transformed <dbl>, Age_sq <dbl>, log_mean_views <dbl>
```

```
nba_data_final[1066,]
```

```
## # A tibble: 1 x 48
##      Rk Player.x Player_ID Pos1  Pos2    Age Tm        G    GS    MP    FG   FGA
##   <dbl> <chr>    <chr>     <fct> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   258 LeBron ~ jamesle01 SF    <NA>     34 LAL      55    55  35.2  10.1  19.9
## # i 36 more variables: FG. <dbl>, X3P <dbl>, X3PA <dbl>, X3P. <dbl>, X2P <dbl>,
## #   X2PA <dbl>, X2P. <dbl>, eFG. <dbl>, FT <dbl>, FTA <dbl>, FT. <dbl>,
## #   ORB <dbl>, DRB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>,
## #   TOV <dbl>, PF <dbl>, PTS <dbl>, Salary <dbl>, mean_views <dbl>,
## #   Season <chr>, Conference <chr>, Role <chr>, Fvot <dbl>, FRank <dbl>,
## #   Pvot <dbl>, PRank <dbl>, Mvot <dbl>, MRank <dbl>, Score <dbl>, Play <chr>,
## #   Salary_transformed <dbl>, Age_sq <dbl>, log_mean_views <dbl>
```

Inference

```
confidence_intervals <- confint(final_model)
print(confidence_intervals)
```

```
##                        2.5 %       97.5 %
## (Intercept)     92.262560759 158.29458310
## Age_sq           0.068610167   0.13687929
## PTS             -1.270333213   4.28064918
## AST              5.326829731  14.06292627
## TRB              4.673711440  10.72668139
## Pos1PF         -27.118106281   1.72143370
## Pos1PG         -62.004100958 -19.88567527
## Pos1SF         -25.911057565   7.75035351
## Pos1SG         -31.190931993   3.86561440
## log_mean_views   2.711263090   8.97309121
## Age_sq:PTS       0.004691395   0.01179549
```