Sports Analytics Project

Ian Keller

2023-10-19

Reading in Each CSV

```
advanced_stats = read.csv('Player Season Totals - 2022-23 Advanced Stats.csv')

basic_stats = read.csv('Player Season Totals - 2022-23 Basic Stats.csv')

player_bio = read.csv('Player Season Totals - 2022-23 Player Bios.csv')
```

Merging Data Sets Based on Name

```
#Combining the player bio and basic stats data sets
combined_data <- merge(player_bio, basic_stats, by = "Player")

#Combining the advanced stats with the previously merged data
combined_data2 <- merge(advanced_stats, combined_data, by = "Player")
```

Adding in the salary dataset

```
player_salary = read.csv("2022-2023 Player Salaries.csv")
#Indexing only the player name and salary columns
player_and_salary <- player_salary[, c("Name", "AAV")]

#Renaming the "Name" Column to "Player" for a proper merge
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
```

```
player_and_salary <- player_and_salary %>%
  rename(Player = Name)
```

```
#Renaming the player columns to match case sensitivity
combined_data2$Player <- tolower(combined_data2$Player)
player_and_salary$Player <- tolower(player_and_salary$Player)

#Finally merging all the data into the final dataset
final_data <- merge(combined_data2, player_and_salary, by = "Player")
```

Next, I will have to clean the data. Remove irrelevant and duplicate columns as we move closer towards analysis. -For starters, I want to remove some count variables like Corsi For and Corsi Against and leave Corsi %

```
library(dplyr)
cleaned_data <- final_data %>% select(-CF,-CA,-FF,-FA,-SF,-SA,-GF,-GA,-xGF,-xGA,-SCF,-SCA,-HDCF,-HDCA,-
```

I now want to convert some count statistics to rate variables. Additionally, I want to create a variable called "Years_since_draft" as I believe the longer they have been in the league can correlate to higher salaries. I expect this variable to be normally distributed as well, meaning that the middle of the career is the highest with two tails at a young and old age.

```
library(dplyr)

#Converting Draft.Year collumn to a numeric value
cleaned_data$Draft.Year <- as.numeric(cleaned_data$Draft.Year)
```

```
## Warning: NAs introduced by coercion
```

```
cleaned_data2 <- cleaned_data %>%
  mutate(
    GPG = Goals / GP.x,
    TOIpg = TOI.x / GP.x,
    AsPG = Total.Assists / GP.x,
    Primary_AsPG = First.Assists / GP.x,
    Secondary_AsPG = Second.Assists / GP.x,
    Years_since_draft = abs(Draft.Year - 2022),
    PPG = Total.Points / GP.x)
```

Testing if there are any duplicate names

```
library(dplyr)

non_unique_players <- cleaned_data2 %>%
  group_by(Player) %>%
  filter(n() > 1) %>%
  arrange(Player)

print(unique(non_unique_players$Player))
```

```
## [1] "sebastian aho"
```

There were 16 duplicate sebastian aho rows, but only two players in the league with that name. This will filter it out to show the two proper rows.

```
library(dplyr)

non_unique_players <- non_unique_players[c(3, 14), ]
```

Merging the two datasets

```
other_players_data <- cleaned_data2 %>%
  filter(Player != "sebastian aho")

# Combine the two datasets
all_data_filtered <- bind_rows(non_unique_players, other_players_data)
```

Removing all NA values for salary and renaming AAV to Salary

```
all_data_filtered <- na.omit(all_data_filtered, cols = "AAV")
#Renaming AAV
all_data_filtered <- all_data_filtered %>%
  rename(Salary = AAV)
```

Furthermore, I need to transform Salary from a string to an integer

```
library(readr)

all_data_filtered$Salary <- parse_number(all_data_filtered$Salary)
```

Multiplying Salary by 1,000,000

```
all_data_filtered$Salary <- all_data_filtered$Salary * 1000000
```

Removing salaries of 0

```
library(dplyr)

all_data_filtered <- all_data_filtered %>%
  filter(Salary != 0 & GP.x >= 41)
```

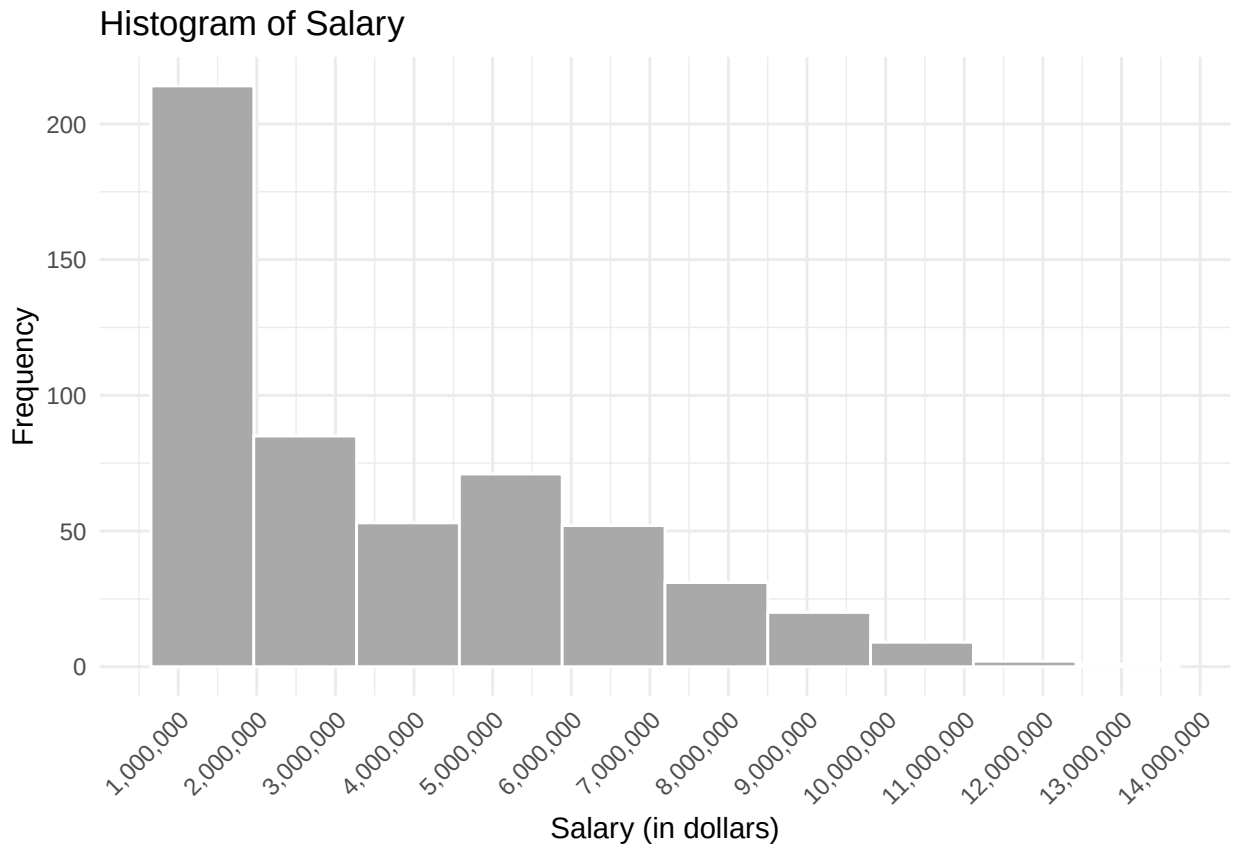
Looking at some descriptive statistics. First let's inspect the dependent variable of salary (AAV).

```
library(ggplot2)
library(scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:readr':
##
##     col_factor
```

```
ggplot(all_data_filtered, aes(x = Salary)) +
  geom_histogram(bins = 10, fill = "darkgrey", color = "white") +
  theme_minimal() +
  scale_x_continuous(labels = scales::comma, breaks = scales::pretty_breaks(n = 12)) +
  labs(title = "Histogram of Salary", x = "Salary (in dollars)", y = "Frequency") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
```



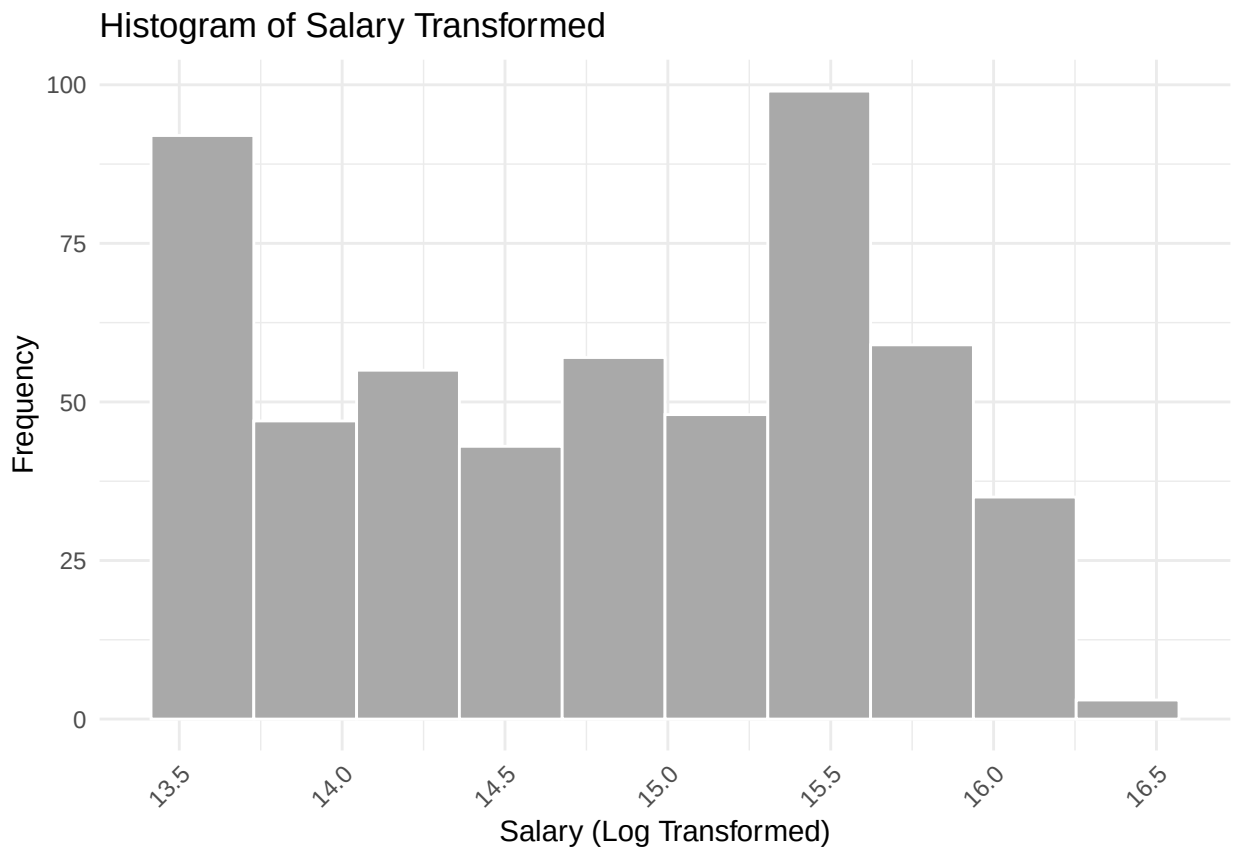
Performing a log transformation

```
all_data_filtered$Log_salary <- log(all_data_filtered$Salary)
```

Histogram of Log Salaries

```
library(ggplot2)
library(scales)

ggplot(all_data_filtered, aes(x = Log_salary)) +
  geom_histogram(bins = 10, fill = "darkgrey", color = "white") +
  theme_minimal() +
  scale_x_continuous(labels = scales::comma, breaks = scales::pretty_breaks(n = 12)) +
  labs(title = "Histogram of Salary Transformed", x = "Salary (Log Transformed)", y = "Frequency") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
```



5 number summary

```
# Calculate the 5-number summary for the Salary variable
five_num_summary <- summary(all_data_filtered$Salary)
print(five_num_summary)
```

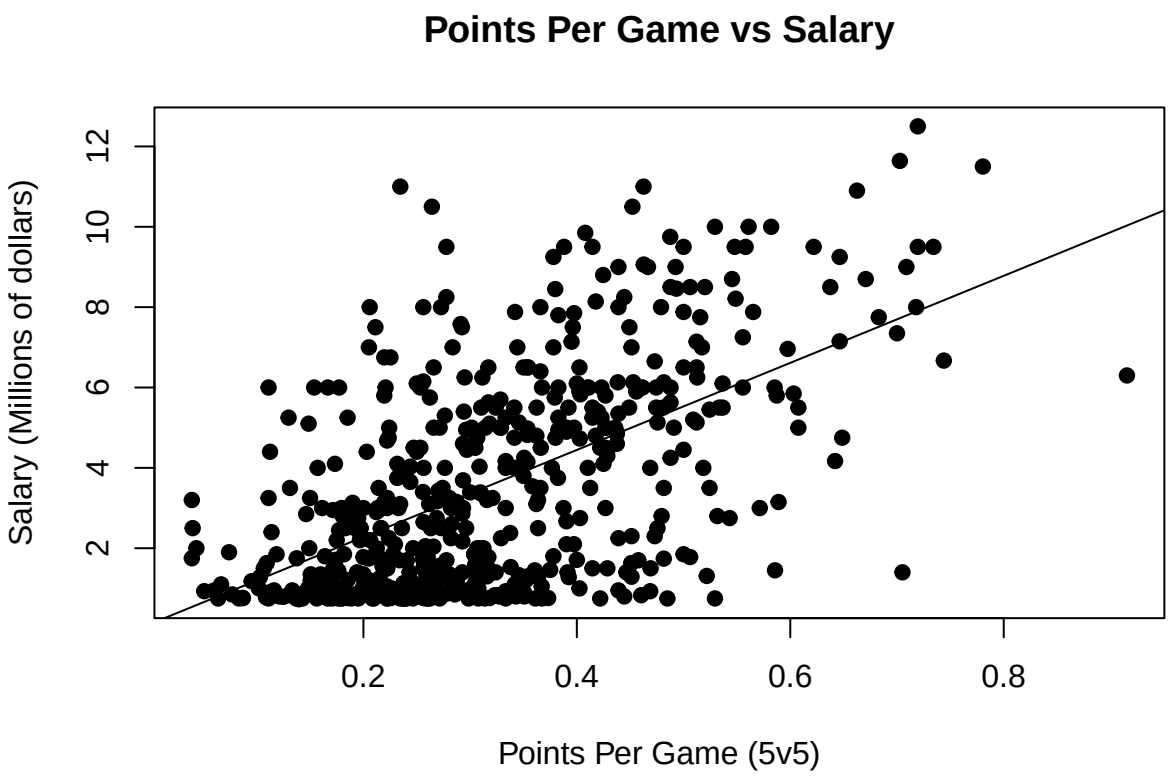
```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##   730000  1200000  2975000  3602565  5500000 12500000
```

Creating a binary column for Offense vs Defense

```
all_data_filtered$Offense_defense <- ifelse(all_data_filtered$Position == "D",1,0)
```

Plotting Points Per Game against Salary

```
model <- lm(Salary / 1e6 ~ PPG, data = all_data_filtered)
plot(all_data_filtered$PPG, all_data_filtered$Salary / 1e6, pch = 19, xlab = "Points Per Game (5v5)", yl
abline(model, col = "black")
```



Creating a subset of data for a correlation plot

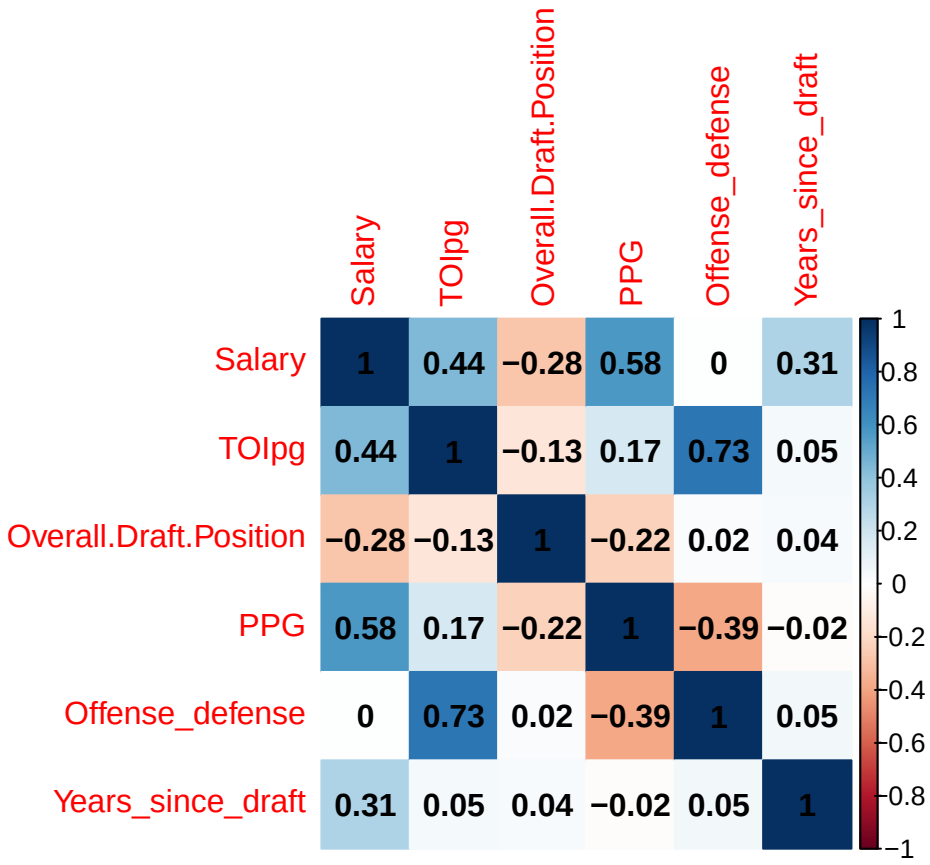
```
library(dplyr)
subset <- all_data_filtered %>%
  ungroup() %>%
  select(Salary, TOIpg, Overall.Draft.Position, PPG, Offense_defense, Years_since_draft)
```

Creating the correlation plot

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
subset2 <- data.frame(lapply(subset, as.numeric))
#Transforming Draft Position to a numeric value before creating the model
all_data_filtered$Overall.Draft.Position <- as.numeric(all_data_filtered$Overall.Draft.Position)
# Calculate the correlation matrix
correlation_matrix = cor(subset2)
corrplot(correlation_matrix, method = "color", addCoef.col = "black")
```



Creating a multiple regression model for our hypothesized significant variables

```
hypothesized_best_model <- lm(Log_salary ~ TOIpg + Overall.Draft.Position + PPG + Offense_defense + Yea
summary(hypothesized_best_model)
```

```
##
## Call:
## lm(formula = Log_salary ~ TOIpg + Overall.Draft.Position + PPG +
##     Offense_defense + Years_since_draft, data = all_data_filtered)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7840 -0.3461  0.0587  0.3825  1.1882
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             11.0346950  0.1890314  58.375  < 2e-16 ***
## TOIpg                    0.2233585  0.0184569  12.102  < 2e-16 ***
## Overall.Draft.Position -0.0018245  0.0004562  -3.999 7.26e-05 ***
## PPG                      1.4375815  0.2508653   5.730 1.68e-08 ***
## Offense_defense         -0.7037527  0.1088269  -6.467 2.27e-10 ***
## Years_since_draft        0.0646090  0.0056914  11.352  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5386 on 532 degrees of freedom
```

```
## Multiple R-squared:  0.5883, Adjusted R-squared:  0.5844
## F-statistic:   152 on 5 and 532 DF,  p-value: < 2.2e-16
```

Using all subsets regression to find the best possible model using all of our predictors

```
#First I will need to create a subset of the data frame
full_predictors <- all_data_filtered[c(
  "Log_salary", "CF.", "FF.", "SF.", "xGF.", "SCF.", "PDO", "Draft.Round",
  "Overall.Draft.Position", "IPP", "Shots", "Rush.Attempts", "Rebounds.Created", "PIM",
  "Penalties.Drawn", "Giveaways", "Takeaways", "Hits", "Shots.Blocked", "GPG", "TOIpg",
  "AsPG", "Years_since_draft", "Offense_defense", "GF.", "HDGF.", "MDGF.", "LDGF.", "PPG"
)]

#Making the entire df numeric
library(dplyr)

full_predictors <- full_predictors %>%
  mutate_all(as.numeric)
```

Running the all subsets regression based on adjr2

```
require(MASS)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

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

```
min_model = lm(Log_salary ~ 1, data = full_predictors)
max_model = lm(Log_salary ~ CF. + FF. + SF. + xGF. + SCF. + PDO + Draft.Round + Overall.Draft.Position +

best_model = step(min_model, scope = formula(max_model), direction = "both", k = log(538))
```

```
## Start:  AIC=-188.12
## Log_salary ~ 1
##
##                          Df Sum of Sq    RSS     AIC
## + AsPG                    1   120.681 254.17 -390.86
## + Shots                   1   114.319 260.53 -377.56
## + PPG                     1   108.617 266.23 -365.91
## + Giveaways               1    92.030 282.82 -333.39
## + TOIpg                   1    86.940 287.91 -323.79
## + Takeaways               1    85.768 289.08 -321.61
## + Rebounds.Created        1    69.274 305.57 -291.76
## + Years_since_draft       1    38.166 336.68 -239.60
## + GPG                     1    37.495 337.35 -238.53
## + SCF.                    1    28.314 346.53 -224.08
```

```
## + CF.                       1    26.698 348.15 -221.58
## + Overall.Draft.Position    1    25.195 349.65 -219.26
## + FF.                       1    22.224 352.62 -214.71
## + SF.                       1    20.425 354.42 -211.97
## + Draft.Round               1    18.676 356.17 -209.32
## + GF.                       1    17.922 356.92 -208.19
## + xGF.                      1    17.354 357.49 -207.33
## + Rush.Attempts             1    17.198 357.65 -207.10
## + MDGF.                     1    16.255 358.59 -205.68
## + Penalties.Drawn           1    13.686 361.16 -201.84
## + Shots.Blocked             1     8.552 366.29 -194.24
## + IPP                       1     7.021 367.83 -192.00
## + Hits                      1     6.878 367.97 -191.79
## + PDO                       1     5.387 369.46 -189.62
## + LDGF.                     1     5.303 369.54 -189.49
## + HDGF.                     1     4.864 369.98 -188.85
## <none>                                  374.85 -188.12
## + Offense_defense           1     0.291 374.56 -182.24
## + PIM                       1     0.221 374.62 -182.14
##
## Step:  AIC=-390.86
## Log_salary ~ AsPG
##
##                           Df Sum of Sq    RSS     AIC
## + Years_since_draft        1    37.429 216.74 -470.28
## + Shots                    1    33.672 220.49 -461.03
## + TOIpg                    1    32.867 221.30 -459.07
## + Giveaways                1    24.637 229.53 -439.42
## + Rebounds.Created         1    17.182 236.98 -422.23
## + Takeaways                1    12.640 241.53 -412.01
## + Shots.Blocked            1    10.514 243.65 -407.30
## + Overall.Draft.Position   1     8.050 246.12 -401.88
## + Draft.Round              1     5.245 248.92 -395.79
## + IPP                      1     3.960 250.20 -393.02
## + PPG                      1     3.954 250.21 -393.01
## + GPG                      1     3.954 250.21 -393.01
## + PDO                      1     3.338 250.83 -391.68
## + Offense_defense          1     3.044 251.12 -391.05
## <none>                                  254.17 -390.86
## + Rush.Attempts            1     1.598 252.57 -387.96
## + CF.                      1     1.572 252.59 -387.91
## + HDGF.                    1     1.429 252.74 -387.60
## + FF.                      1     1.403 252.76 -387.55
## + SCF.                     1     1.340 252.82 -387.41
## + SF.                      1     1.183 252.98 -387.08
## + PIM                      1     0.623 253.54 -385.89
## + GF.                      1     0.615 253.55 -385.87
## + xGF.                     1     0.499 253.67 -385.63
## + LDGF.                    1     0.120 254.05 -384.82
## + Penalties.Drawn          1     0.094 254.07 -384.77
## + MDGF.                    1     0.037 254.13 -384.65
## + Hits                     1     0.002 254.16 -384.57
## - AsPG                     1   120.681 374.85 -188.12
##
```

```
## Step:  AIC=-470.28
## Log_salary ~ AsPG + Years_since_draft
##
##                           Df Sum of Sq    RSS     AIC
## + Shots                    1    29.746 186.99 -543.41
## + TOIpg                    1    29.626 187.11 -543.07
## + Giveaways                1    20.325 196.41 -516.97
## + Rebounds.Created         1    12.447 204.29 -495.81
## + Takeaways                1    11.476 205.26 -493.26
## + Overall.Draft.Position   1     9.447 207.29 -487.96
## + Shots.Blocked            1     6.878 209.86 -481.34
## + Draft.Round              1     6.441 210.29 -480.22
## + PPG                      1     5.185 211.55 -477.01
## + GPG                      1     5.185 211.55 -477.01
## + PDO                      1     4.056 212.68 -474.15
## + HDGF.                    1     2.732 214.00 -470.81
## + IPP                      1     2.609 214.13 -470.50
## <none>                                 216.74 -470.28
## + Offense_defense          1     1.997 214.74 -468.97
## + Rush.Attempts            1     1.303 215.43 -467.23
## + GF.                      1     1.248 215.49 -467.09
## + CF.                      1     0.966 215.77 -466.39
## + SCF.                     1     0.844 215.89 -466.09
## + FF.                      1     0.704 216.03 -465.74
## + SF.                      1     0.589 216.15 -465.45
## + Penalties.Drawn          1     0.545 216.19 -465.34
## + Hits                     1     0.240 216.50 -464.58
## + xGF.                     1     0.122 216.61 -464.29
## + LDGF.                    1     0.042 216.69 -464.09
## + PIM                      1     0.017 216.72 -464.03
## + MDGF.                    1     0.010 216.73 -464.01
## - Years_since_draft        1    37.429 254.17 -390.86
## - AsPG                     1   119.944 336.68 -239.60
##
## Step:  AIC=-543.41
## Log_salary ~ AsPG + Years_since_draft + Shots
##
##                           Df Sum of Sq    RSS     AIC
## + TOIpg                    1    22.766 164.22 -606.97
## + Giveaways                1     8.787 178.20 -563.01
## + Shots.Blocked            1     7.232 179.76 -558.34
## + IPP                      1     6.415 180.57 -555.90
## + Overall.Draft.Position   1     6.026 180.96 -554.75
## + HDGF.                    1     5.354 181.64 -552.75
## + Offense_defense          1     5.353 181.64 -552.75
## + Draft.Round              1     4.029 182.96 -548.84
## + GF.                      1     3.589 183.40 -547.55
## + PDO                      1     2.680 184.31 -544.89
## + xGF.                     1     2.406 184.58 -544.09
## <none>                                 186.99 -543.41
## + Penalties.Drawn          1     1.685 185.31 -541.99
## + SF.                      1     1.294 185.70 -540.86
## + Takeaways                1     1.278 185.71 -540.81
## + FF.                      1     1.270 185.72 -540.79
```

```
## + Hits                     1     1.187 185.80 -540.55
## + CF.                      1     1.046 185.94 -540.14
## + GPG                      1     1.043 185.95 -540.13
## + PPG                      1     1.043 185.95 -540.13
## + SCF.                     1     1.038 185.95 -540.12
## + Rush.Attempts            1     1.026 185.96 -540.08
## + PIM                      1     0.518 186.47 -538.61
## + Rebounds.Created         1     0.300 186.69 -537.98
## + LDGF.                    1     0.155 186.84 -537.57
## + MDGF.                    1     0.138 186.85 -537.52
## - Shots                    1    29.746 216.74 -470.28
## - Years_since_draft        1    33.503 220.49 -461.03
## - AsPG                     1    41.933 228.92 -440.84
##
## Step:  AIC=-606.97
## Log_salary ~ AsPG + Years_since_draft + Shots + TOIpg
##
##                           Df Sum of Sq    RSS     AIC
## + Offense_defense          1   12.9426 151.28 -644.84
## + Overall.Draft.Position   1    4.8780 159.35 -616.90
## + HDGF.                    1    4.0380 160.19 -614.07
## + IPP                      1    3.9117 160.31 -613.65
## + Draft.Round              1    3.3815 160.84 -611.87
## + Shots.Blocked            1    3.3305 160.89 -611.70
## + GF.                      1    2.8552 161.37 -610.12
## + PDO                      1    2.2487 161.97 -608.10
## <none>                                 164.22 -606.97
## + Hits                     1    1.7814 162.44 -606.55
## + xGF.                     1    1.7270 162.50 -606.37
## + GPG                      1    1.4720 162.75 -605.52
## + PPG                      1    1.4720 162.75 -605.52
## + PIM                      1    1.1639 163.06 -604.51
## + Takeaways                1    1.0498 163.17 -604.13
## + SCF.                     1    0.8970 163.33 -603.63
## + SF.                      1    0.8426 163.38 -603.45
## + FF.                      1    0.8352 163.39 -603.42
## + CF.                      1    0.7131 163.51 -603.02
## + Rush.Attempts            1    0.4350 163.79 -602.11
## + Rebounds.Created         1    0.3985 163.82 -601.99
## + Giveaways                1    0.2104 164.01 -601.37
## + MDGF.                    1    0.1794 164.04 -601.27
## + Penalties.Drawn          1    0.0755 164.15 -600.93
## + LDGF.                    1    0.0411 164.18 -600.81
## - TOIpg                    1   22.7663 186.99 -543.41
## - Shots                    1   22.8857 187.11 -543.07
## - AsPG                     1   25.3574 189.58 -536.01
## - Years_since_draft        1   31.2174 195.44 -519.63
##
## Step:  AIC=-644.84
## Log_salary ~ AsPG + Years_since_draft + Shots + TOIpg + Offense_defense
##
##                           Df Sum of Sq    RSS     AIC
## + Overall.Draft.Position   1     4.143 147.14 -653.50
## + Draft.Round              1     2.791 148.49 -648.57
```

```
## <none>                                151.28 -644.84
## + HDGF.                    1     1.623 149.66 -644.36
## + PDO                      1     1.289 149.99 -643.16
## + Hits                     1     1.045 150.24 -642.28
## + Giveaways                1     0.887 150.39 -641.72
## + GF.                      1     0.785 150.50 -641.35
## + Penalties.Drawn          1     0.611 150.67 -640.73
## + PIM                      1     0.350 150.93 -639.80
## + Rebounds.Created         1     0.321 150.96 -639.70
## + Rush.Attempts            1     0.302 150.98 -639.63
## + GPG                      1     0.222 151.06 -639.35
## + PPG                      1     0.222 151.06 -639.35
## + xGF.                     1     0.193 151.09 -639.24
## + Shots.Blocked            1     0.093 151.19 -638.89
## + IPP                      1     0.061 151.22 -638.77
## + LDGF.                    1     0.038 151.24 -638.69
## + MDGF.                    1     0.017 151.26 -638.62
## + SCF.                     1     0.016 151.26 -638.61
## + Takeaways                1     0.012 151.27 -638.60
## + CF.                      1     0.000 151.28 -638.56
## + SF.                      1     0.000 151.28 -638.56
## + FF.                      1     0.000 151.28 -638.56
## - Shots                    1     6.185 157.47 -629.57
## - AsPG                     1     9.432 160.71 -618.59
## - Offense_defense          1    12.943 164.22 -606.97
## - TOIpg                    1    30.356 181.64 -552.75
## - Years_since_draft        1    33.478 184.76 -543.58
##
## Step:  AIC=-653.5
## Log_salary ~ AsPG + Years_since_draft + Shots + TOIpg + Offense_defense +
##     Overall.Draft.Position
##
##                           Df Sum of Sq    RSS     AIC
## + Draft.Round              1     3.736 143.40 -661.04
## <none>                                 147.14 -653.50
## + HDGF.                    1     1.359 145.78 -652.20
## + PDO                      1     1.128 146.01 -651.35
## + Hits                     1     0.746 146.39 -649.95
## + Giveaways                1     0.623 146.51 -649.49
## + GF.                      1     0.614 146.52 -649.46
## + Penalties.Drawn          1     0.434 146.70 -648.80
## + PIM                      1     0.304 146.83 -648.32
## + GPG                      1     0.300 146.84 -648.31
## + PPG                      1     0.300 146.84 -648.31
## + Rebounds.Created         1     0.213 146.93 -647.99
## + Rush.Attempts            1     0.143 146.99 -647.73
## + IPP                      1     0.091 147.05 -647.54
## + xGF.                     1     0.069 147.07 -647.46
## + Shots.Blocked            1     0.067 147.07 -647.46
## + LDGF.                    1     0.014 147.12 -647.26
## + Takeaways                1     0.008 147.13 -647.24
## + SF.                      1     0.006 147.13 -647.23
## + FF.                      1     0.004 147.13 -647.22
## + MDGF.                    1     0.002 147.14 -647.21
```

```
## + SCF.                     1     0.001 147.14 -647.21
## + CF.                      1     0.001 147.14 -647.21
## - Overall.Draft.Position   1     4.143 151.28 -644.84
## - Shots                    1     5.437 152.57 -640.26
## - AsPG                     1     8.522 155.66 -629.49
## - Offense_defense          1    12.208 159.35 -616.90
## - TOIpg                    1    28.703 175.84 -563.91
## - Years_since_draft        1    34.469 181.61 -546.55
##
## Step:  AIC=-661.04
## Log_salary ~ AsPG + Years_since_draft + Shots + TOIpg + Offense_defense +
##     Overall.Draft.Position + Draft.Round
##
##                           Df Sum of Sq    RSS     AIC
## <none>                                 143.40 -661.04
## + HDGF.                    1     1.273 142.13 -659.55
## + PDO                      1     0.939 142.46 -658.29
## + Hits                     1     0.891 142.51 -658.11
## + Penalties.Drawn          1     0.524 142.88 -656.73
## + GF.                      1     0.470 142.93 -656.52
## + Giveaways                1     0.412 142.99 -656.30
## + PIM                      1     0.411 142.99 -656.30
## + GPG                      1     0.186 143.22 -655.45
## + PPG                      1     0.186 143.22 -655.45
## + Rebounds.Created         1     0.167 143.24 -655.38
## + Rush.Attempts            1     0.141 143.26 -655.29
## + Shots.Blocked            1     0.098 143.30 -655.13
## + LDGF.                    1     0.066 143.34 -655.01
## + xGF.                     1     0.053 143.35 -654.95
## + IPP                      1     0.052 143.35 -654.95
## + SF.                      1     0.015 143.39 -654.81
## + FF.                      1     0.011 143.39 -654.80
## + CF.                      1     0.005 143.40 -654.78
## + Takeaways                1     0.001 143.40 -654.76
## + SCF.                     1     0.000 143.40 -654.76
## + MDGF.                    1     0.000 143.40 -654.76
## - Draft.Round              1     3.736 147.14 -653.50
## - Shots                    1     4.820 148.22 -649.55
## - Overall.Draft.Position   1     5.088 148.49 -648.57
## - AsPG                     1     8.589 151.99 -636.04
## - Offense_defense          1    12.151 155.55 -623.57
## - TOIpg                    1    27.607 171.01 -572.61
## - Years_since_draft        1    34.359 177.76 -551.78
```

Summary Statistics and ANOVA table of the best model

```
best_BIC_model = lm(Log_salary ~ AsPG + Years_since_draft + Shots + TOIpg + Offense_defense +
    Overall.Draft.Position + Draft.Round, data = full_predictors)

summary(best_BIC_model)
```

```
##
## Call:
```

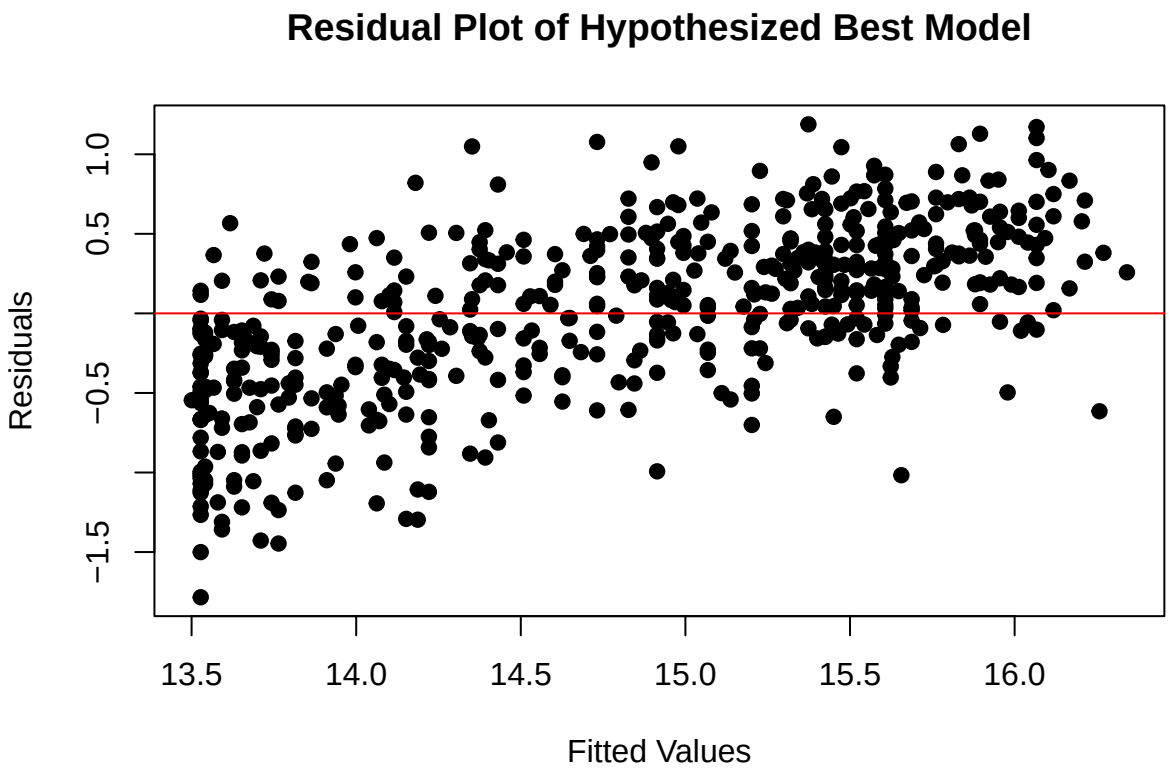
```
## lm(formula = Log_salary ~ AsPG + Years_since_draft + Shots +
##     TOIpg + Offense_defense + Overall.Draft.Position + Draft.Round,
##     data = full_predictors)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.80969 -0.33085  0.05368  0.40409  1.15455
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)            11.1148798  0.1882572  59.041  < 2e-16 ***
## AsPG                    1.8614540  0.3303907   5.634 2.86e-08 ***
## Years_since_draft       0.0621244  0.0055129  11.269  < 2e-16 ***
## Shots                   0.0031676  0.0007505   4.221 2.87e-05 ***
## TOIpg                   0.1879796  0.0186097  10.101  < 2e-16 ***
## Offense_defense        -0.6486806  0.0967979  -6.701 5.29e-11 ***
## Overall.Draft.Position -0.0110490  0.0025478  -4.337 1.73e-05 ***
## Draft.Round             0.2941449  0.0791619   3.716 0.000224 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5202 on 530 degrees of freedom
## Multiple R-squared:  0.6174, Adjusted R-squared:  0.6124
## F-statistic: 122.2 on 7 and 530 DF,  p-value: < 2.2e-16
```

```
anova(best_BIC_model)
```

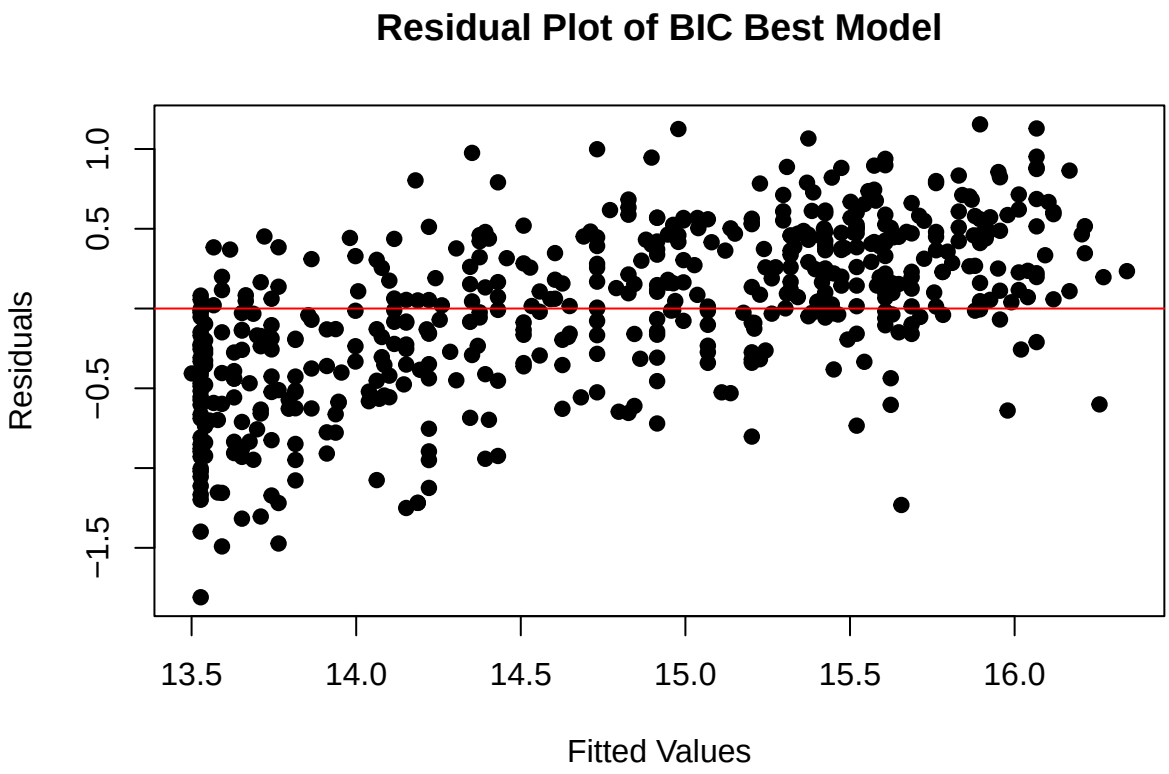
```
## Analysis of Variance Table
##
## Response: Log_salary
##                          Df  Sum Sq Mean Sq F value    Pr(>F)
## AsPG                      1 120.681 120.681 446.025 < 2.2e-16 ***
## Years_since_draft         1  37.429  37.429 138.336 < 2.2e-16 ***
## Shots                     1  29.746  29.746 109.937 < 2.2e-16 ***
## TOIpg                     1  22.766  22.766  84.142 < 2.2e-16 ***
## Offense_defense           1  12.943  12.943  47.834  1.34e-11 ***
## Overall.Draft.Position    1   4.143   4.143  15.314 0.0001029 ***
## Draft.Round               1   3.736   3.736  13.807 0.0002241 ***
## Residuals               530 143.402   0.271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plotting the hypothesized and best BIC model residuals

```
plot(full_predictors$Log_salary, residuals(hypothesized_best_model), pch =19, main = "Residual Plot of
abline(h=0,col="red")
```




```
plot(full_predictors$Log_salary, residuals(best_BIC_model), pch =19, main = "Residual Plot of BIC Best M
abline(h=0,col="red")
```



Running the full model

```
full_model = lm(Log_salary ~ PPG + CF. + FF. + SF. + xGF. + SCF. + PDO + Draft.Round + Overall.Draft.Pos
summary(full_model)
```

```
##
## Call:
## lm(formula = Log_salary ~ PPG + CF. + FF. + SF. + xGF. + SCF. +
##     PDO + Draft.Round + Overall.Draft.Position + IPP + Shots +
##     Rush.Attempts + Rebounds.Created + PIM + Penalties.Drawn +
##     Giveaways + Takeaways + Hits + Shots.Blocked + GPG + TOIpg +
##     Years_since_draft + Offense_defense + GF. + HDGF. + MDGF. +
##     LDGF., data = full_predictors)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8498 -0.3175  0.0511  0.3842  1.2125
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            19.1960736  8.0710044   2.378 0.017755 *
## PPG                     2.1626549  0.5118167   4.225 2.83e-05 ***
## CF.                    -0.0480720  0.0330589  -1.454 0.146525
## FF.                     0.0452693  0.0397783   1.138 0.255639
## SF.                    -0.0246659  0.0359223  -0.687 0.492619
## xGF.                   -0.0157515  0.0179546  -0.877 0.380740
```

```
## SCF.                     0.0253065  0.0193030   1.311 0.190443
## PDO                     -7.6952564  8.0129418  -0.960 0.337332
## Draft.Round              0.2788991  0.0803055   3.473 0.000558 ***
## Overall.Draft.Position -0.0103793  0.0025870  -4.012 6.92e-05 ***
## IPP                     -0.0027764  0.0032821  -0.846 0.398002
## Shots                    0.0042071  0.0012495   3.367 0.000817 ***
## Rush.Attempts           -0.0052881  0.0039461  -1.340 0.180816
## Rebounds.Created        -0.0026353  0.0051311  -0.514 0.607763
## PIM                     -0.0004852  0.0015147  -0.320 0.748847
## Penalties.Drawn         -0.0021967  0.0047550  -0.462 0.644289
## Giveaways                0.0039110  0.0024618   1.589 0.112750
## Takeaways               -0.0010196  0.0026029  -0.392 0.695442
## Hits                    -0.0006090  0.0005481  -1.111 0.267034
## Shots.Blocked           -0.0004256  0.0015566  -0.273 0.784646
## GPG                     -2.2346904  0.5906065  -3.784 0.000173 ***
## TOIpg                    0.1726732  0.0231585   7.456 3.85e-13 ***
## Years_since_draft        0.0631072  0.0057030  11.066  < 2e-16 ***
## Offense_defense         -0.7079339  0.1438934  -4.920 1.17e-06 ***
## GF.                      0.0297098  0.0357962   0.830 0.406943
## HDGF.                   -0.0092486  0.0115856  -0.798 0.425077
## MDGF.                   -0.0029591  0.0064721  -0.457 0.647710
## LDGF.                   -0.0007670  0.0029895  -0.257 0.797620
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5218 on 510 degrees of freedom
## Multiple R-squared:  0.6295, Adjusted R-squared:  0.6099
## F-statistic:  32.1 on 27 and 510 DF,  p-value: < 2.2e-16
```