

Load, Clean, and Explore Data

Basketball Salaries Team

Load NBA 2K Data

Note: Primary dataset is directly downloaded from Kaggle. This video-game rankings dataset is scraped from <http://mtddb.com/20>

```
library(stringr)
library(rvest)
library(tidyr)
if (!file.exists('../data/raw/nba2k/nba2k_16.csv')){ # only run if data is not already scraped
# constants
root <- 'data/raw/nba2k'
years <- c(16,17,18,19,20)
pages = c(84,68,72,68,46)
url_f <- 'http://mtddb.com/%d?page=%d&sortedBy=overall&sortOrder=Descending&'
for (i in 1:length(years)){
  year_df <- vector('list',12)
  names(year_df) <- c('name','position','ovr','out','ins','pla','ath','def','reb','xbox','ps4','pc')
  year <- years[i]
  page <- pages[i]
  for (page in 1:page){
    # load webpage
    url <- sprintf(url_f,year,page)
    webpage <- read_html(url)
    # load salary table
    player_tables <- html_nodes(webpage, css = 'table')
    player_df_page <- html_table(player_tables[[1]])#[-(1),]
    names(player_df_page) <- c('name','position','ovr','out','ins','pla','ath','def','reb','xbox','ps4','pc')
    year_df <- rbind(year_df,player_df_page)}
write.csv(year_df,sprintf('%s/nba2k_%d.csv',root,year))
cat(sprintf('%d nrow: %d\n',year,nrow(year_df)))}}
```

Clean Primary Dataset

```
library("readxl")
df_primary <- read_excel('../data/raw/primary_dataset_raw.xlsx')

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in D24626 / R24626C4: got 'z'

df_primary <- df_primary[!(names(df_primary)%in%c('#','blanl','blank2'))] # drop empty/non-stat columns
colnames(df_primary)[1:3] <- c('year','name_p','salary')
df_primary <- df_primary[!is.na(df_primary[['salary']]),] # drop rows with no salaries
df_primary[is.na(df_primary)] <- 0
df_primary <- df_primary[df_primary$year%in%c(2016:2020),] # take 2016-2017 player data
head(df_primary)
```

```
## # A tibble: 6 x 51
##   year name_p salary Pos   Age Tm      G   GS   MP   PER `TS%` `3PAR`
##   <dbl> <chr>   <dbl> <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2017 A.J. ~ 1.31e6 C      24 DAL    22    0   163   8.4 0.472 0.238
## 2  2016 Aaron~ 2.70e6 PG      31 CHI    69    0  1108  11.8 0.494 0.394
## 3  2017 Aaron~ 2.12e6 PG      32 IND    65    0   894   9.5 0.507 0.427
## 4  2016 Aaron~ 4.35e6 PF      20 ORL    78   37  1863  17   0.541 0.245
## 5  2017 Aaron~ 5.50e6 SF      21 ORL    80   72  2298  14.4 0.53  0.309
```

```
## 6 2016 Aaron~ 3.76e5 SG      21 CHO      21      0      93  4.3 0.371  0.526
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

```
summary(df_primary)
```

```
##      year      name_p      salary      Pos
## Min.   :2016   Length:965   Min.    : 11534   Length:965
## 1st Qu.:2016   Class :character 1st Qu.: 1551659   Class :character
## Median :2017   Mode  :character  Median : 4000000   Mode  :character
## Mean    :2017                                     Mean    : 6789399
## 3rd Qu.:2017                                     3rd Qu.:10500000
## Max.    :2017                                     Max.    :34682550
##      Age      Tm      G      GS
## Min.   :19.00   Length:965   Min.    : 1.00   Min.    : 0.00
## 1st Qu.:23.00   Class :character 1st Qu.:32.00   1st Qu.: 1.00
## Median :26.00   Mode  :character  Median :61.00   Median :12.00
## Mean    :26.48                                     Mean    :53.41   Mean    :25.99
## 3rd Qu.:29.00                                     3rd Qu.:75.00   3rd Qu.:52.00
## Max.    :40.00                                     Max.    :82.00   Max.    :82.00
##      MP      PER      TS%      3PAr
## Min.   : 1    Min.   : -35.30   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:496   1st Qu.: 10.50   1st Qu.:0.5040   1st Qu.:0.1360
## Median :1197   Median : 13.30   Median :0.5380   Median :0.3110
## Mean    :1247   Mean    : 13.61   Mean    :0.5324   Mean    :0.3045
## 3rd Qu.:1954   3rd Qu.: 16.30   3rd Qu.:0.5710   3rd Qu.:0.4470
## Max.    :3125   Max.    : 39.30   Max.    :1.0000   Max.    :1.0000
##      FTr      ORB%      DRB%      TRB%
## Min.   :0.0000   Min.   : 0.000   Min.   : 0.00   Min.   : 0.000
## 1st Qu.:0.1670   1st Qu.: 1.900   1st Qu.:10.30   1st Qu.: 6.200
## Median :0.2400   Median : 3.300   Median :14.00   Median : 8.800
## Mean    :0.2682   Mean    : 4.868   Mean    :15.13   Mean    : 9.992
## 3rd Qu.:0.3380   3rd Qu.: 7.100   3rd Qu.:19.20   3rd Qu.:13.100
## Max.    :2.0000   Max.    :27.300   Max.    :39.20   Max.    :30.300
##      AST%      STL%      BLK%      TOV%
## Min.   : 0.00   Min.   : 0.000   Min.   : 0.000   Min.   : 0.00
## 1st Qu.: 7.00   1st Qu.: 1.100   1st Qu.: 0.500   1st Qu.: 9.90
## Median :10.40   Median : 1.500   Median : 1.200   Median :12.50
## Mean    :13.38   Mean    : 1.583   Mean    : 1.652   Mean    :12.82
## 3rd Qu.:17.80   3rd Qu.: 1.900   3rd Qu.: 2.300   3rd Qu.:15.20
## Max.    :72.30   Max.    :11.100   Max.    :15.100   Max.    :43.60
##      USG%      OWS      DWS      WS
## Min.   : 0.00   Min.   : -3.300   Min.   :0.000   Min.   : -2.10
## 1st Qu.:15.30   1st Qu.: 0.100   1st Qu.:0.400   1st Qu.: 0.50
## Median :18.40   Median : 0.800   Median :1.000   Median : 1.80
## Mean    :18.85   Mean    : 1.387   Mean    :1.272   Mean    : 2.66
## 3rd Qu.:21.80   3rd Qu.: 2.100   3rd Qu.:1.900   3rd Qu.: 3.80
## Max.    :41.70   Max.    :13.800   Max.    :6.000   Max.    :17.90
##      WS/48      OBPM      DBPM      BPM
## Min.   : -0.28300   Min.   : -17.3000   Min.   : -8.5000   Min.   : -24.100
## 1st Qu.: 0.05000   1st Qu.: -2.4000   1st Qu.: -1.5000   1st Qu.: -3.100
## Median : 0.08700   Median : -0.9000   Median : -0.3000   Median : -1.200
## Mean    : 0.08683   Mean    : -0.9566   Mean    : -0.2671   Mean    : -1.225
## 3rd Qu.: 0.12100   3rd Qu.: 0.4000   3rd Qu.: 1.0000   3rd Qu.: 0.700
## Max.    : 0.63400   Max.    : 15.3000   Max.    :12.0000   Max.    : 15.600
##      VORP      FG      FGA      FG%
## Min.   : -1.4000   Min.   : 0.0      Min.   : 0.0      Min.   :0.0000
## 1st Qu.: -0.1000   1st Qu.: 62.0     1st Qu.:146.0     1st Qu.:0.4050
## Median : 0.2000   Median :166.0     Median : 368.0     Median :0.4410
```

| | | | | | | | | |
|----|----------|----------|----------|---------|----------|---------|----------|---------|
| ## | Mean | : 0.6493 | Mean | :200.8 | Mean | : 441.5 | Mean | :0.4463 |
| ## | 3rd Qu.: | 1.0000 | 3rd Qu.: | 294.0 | 3rd Qu.: | 644.0 | 3rd Qu.: | 0.4810 |
| ## | Max. | :12.4000 | Max. | :824.0 | Max. | :1941.0 | Max. | :1.0000 |
| ## | 3P | | 3PA | | 3P% | | 2P | |
| ## | Min. | : 0.00 | Min. | : 0.0 | Min. | :0.0000 | Min. | : 0 |
| ## | 1st Qu.: | 3.00 | 1st Qu.: | 12.0 | 1st Qu.: | 0.2450 | 1st Qu.: | 43 |
| ## | Median | : 30.00 | Median | : 92.0 | Median | :0.3330 | Median | :113 |
| ## | Mean | : 47.83 | Mean | :133.8 | Mean | :0.2846 | Mean | :153 |
| ## | 3rd Qu.: | 77.00 | 3rd Qu.: | 215.0 | 3rd Qu.: | 0.3750 | 3rd Qu.: | 219 |
| ## | Max. | :402.00 | Max. | :886.0 | Max. | :1.0000 | Max. | :730 |
| ## | 2PA | | 2P% | | eFG% | | FT | |
| ## | Min. | : 0.0 | Min. | :0.0000 | Min. | :0.0000 | Min. | : 0.00 |
| ## | 1st Qu.: | 93.0 | 1st Qu.: | 0.4460 | 1st Qu.: | 0.4670 | 1st Qu.: | 23.00 |
| ## | Median | : 235.0 | Median | :0.4830 | Median | :0.5010 | Median | : 59.00 |
| ## | Mean | : 307.8 | Mean | :0.4837 | Mean | :0.4986 | Mean | : 92.23 |
| ## | 3rd Qu.: | 444.0 | 3rd Qu.: | 0.5290 | 3rd Qu.: | 0.5360 | 3rd Qu.: | 120.00 |
| ## | Max. | :1421.0 | Max. | :1.0000 | Max. | :1.0000 | Max. | :746.00 |
| ## | FTA | | FT% | | ORB | | DRB | |
| ## | Min. | : 0.0 | Min. | :0.0000 | Min. | : 0.00 | Min. | : 0 |
| ## | 1st Qu.: | 33.0 | 1st Qu.: | 0.6740 | 1st Qu.: | 13.00 | 1st Qu.: | 62 |
| ## | Median | : 78.0 | Median | :0.7640 | Median | : 33.00 | Median | :143 |
| ## | Mean | :120.3 | Mean | :0.7305 | Mean | : 52.69 | Mean | :173 |
| ## | 3rd Qu.: | 161.0 | 3rd Qu.: | 0.8310 | 3rd Qu.: | 70.00 | 3rd Qu.: | 243 |
| ## | Max. | :881.0 | Max. | :1.0000 | Max. | :395.00 | Max. | :817 |
| ## | TRB | | AST | | STL | | BLK | |
| ## | Min. | : 0.0 | Min. | : 0.0 | Min. | : 0.00 | Min. | : 0.00 |
| ## | 1st Qu.: | 79.0 | 1st Qu.: | 30.0 | 1st Qu.: | 14.00 | 1st Qu.: | 5.00 |
| ## | Median | : 178.0 | Median | : 74.0 | Median | : 33.00 | Median | : 15.00 |
| ## | Mean | : 225.7 | Mean | :115.5 | Mean | : 40.02 | Mean | : 25.03 |
| ## | 3rd Qu.: | 307.0 | 3rd Qu.: | 151.0 | 3rd Qu.: | 58.00 | 3rd Qu.: | 33.00 |
| ## | Max. | :1198.0 | Max. | :906.0 | Max. | :169.00 | Max. | :269.00 |
| ## | TOV | | PF | | PTS | | | |
| ## | Min. | : 0.00 | Min. | : 0.0 | Min. | : 0.0 | | |
| ## | 1st Qu.: | 25.00 | 1st Qu.: | 47.0 | 1st Qu.: | 166.0 | | |
| ## | Median | : 57.00 | Median | :102.0 | Median | : 437.0 | | |
| ## | Mean | : 70.13 | Mean | :103.4 | Mean | : 541.8 | | |
| ## | 3rd Qu.: | 99.00 | 3rd Qu.: | 152.0 | 3rd Qu.: | 780.0 | | |
| ## | Max. | :464.00 | Max. | :278.0 | Max. | :2558.0 | | |

Numeric / Factor Variables

```
df_primary$Tm <- as.factor(df_primary$Tm) # TOT means they played for multiple teams
# will be useful later when multiple records for a single player in a single year
df_primary$year <- as.factor(df_primary$year) # make year a factor variable
df_primary[df_primary$Pos=='PF-C',] # only 2 Power-Forwards / Centers
```

```
## # A tibble: 2 x 51
##   year name_p salary Pos      Age Tm      G      GS      MP      PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <chr> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Chann~ 7.81e6 PF-C    32 TOT     70    32 1200 12.9 0.586 0.677
## 2 2017 Joffr~ 1.52e6 PF-C    25 TOT     70     1  980 12.6 0.509 0.292
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

```
# each player should only have 1 position
# both Channing Frye and Joffrey Lauvergne are classified as Forwards (PF)
# https://www.espn.com/nba/player/stats/_/id/2754/channing-frye
# https://www.espn.com/nba/player/stats/_/id/2959753/joffrey-lauvergne
```

```
df_primary$Pos <- gsub('PF-C','PF',df_primary$Pos)
df_primary$Pos <- as.factor(df_primary$Pos) # make Pos a factor variable
table(df_primary$Pos)
```

```
##
##    C  PF  PG  SF  SG
## 185 192 200 194 194
```

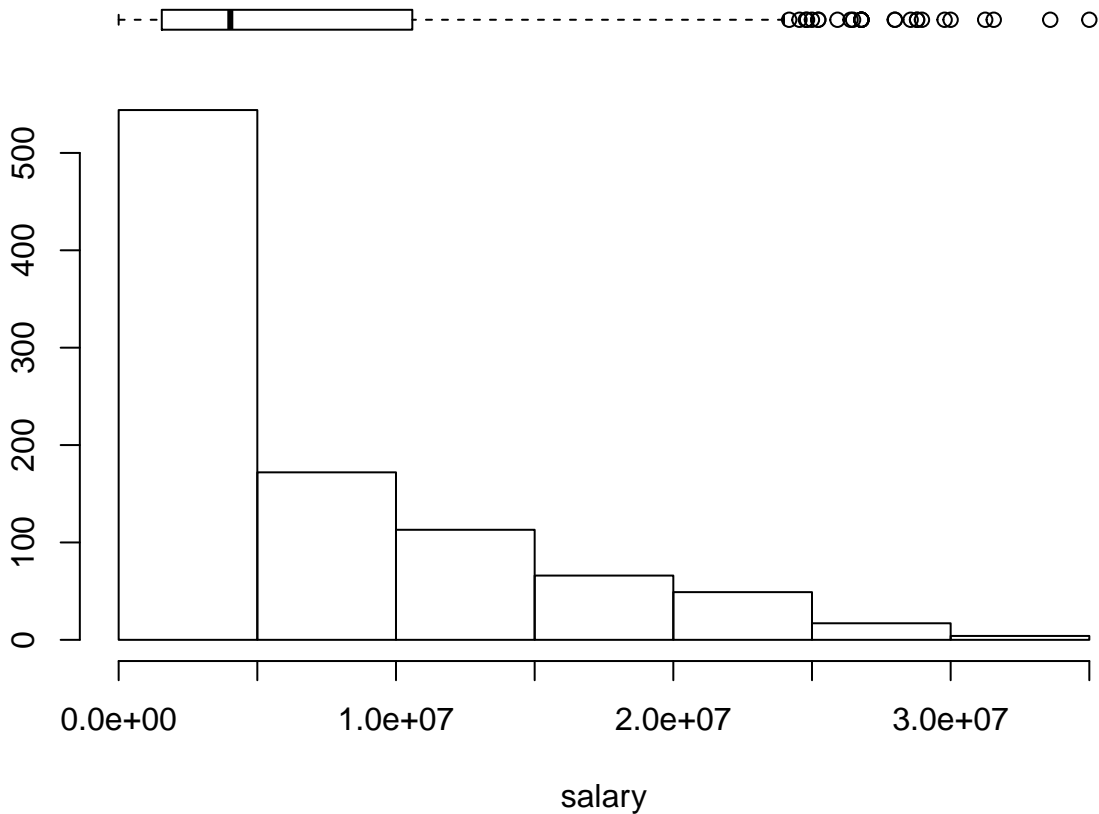
```
str(df_primary)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    965 obs. of  51 variables:
## $ year   : Factor w/ 2 levels "2016","2017": 2 1 2 1 2 1 1 1 2 1 ...
## $ name_p: chr  "A.J. Hammons" "Aaron Brooks" "Aaron Brooks" "Aaron Gordon" ...
## $ salary: num  1312611 2700000 2116955 4351320 5504420 ...
## $ Pos    : Factor w/ 5 levels "C","PF","PG",...: 1 3 3 2 4 5 2 1 1 1 ...
## $ Age    : num  24 31 32 20 21 21 24 29 30 31 ...
## $ Tm     : Factor w/ 31 levels "ATL","BOS","BRK",...: 7 4 12 22 22 5 18 1 2 5 ...
## $ G      : num  22 69 65 78 80 21 52 82 68 47 ...
## $ GS     : num  0 0 0 37 72 0 2 82 68 18 ...
## $ MP     : num  163 1108 894 1863 2298 ...
## $ PER    : num  8.4 11.8 9.5 17 14.4 4.3 5.6 19.4 17.7 18.2 ...
## $ TS%    : num  0.472 0.494 0.507 0.541 0.53 0.371 0.422 0.565 0.553 0.507 ...
## $ 3PAr   : num  0.238 0.394 0.427 0.245 0.309 0.526 0.221 0.244 0.302 0 ...
## $ FTr    : num  0.476 0.136 0.133 0.333 0.251 0.632 0.179 0.123 0.169 0.22 ...
## $ ORB%   : num  5.4 2 2.3 9 5.3 4.7 4.8 6.3 4.9 5.6 ...
## $ DRB%   : num  20.9 7.5 6.3 21.3 14.1 13.1 21.5 18.2 18.6 24.6 ...
## $ TRB%   : num  12.8 4.8 4.3 15.1 9.6 8.8 13.3 12.4 11.8 15 ...
## $ AST%   : num  3.8 26 20.7 10.3 10.5 3 8.9 16.7 24.4 11.8 ...
## $ STL%   : num  0.3 1.4 1.4 1.6 1.4 3.2 1.7 1.3 1.2 1.4 ...
## $ BLK%   : num  7.2 0.7 0.9 2.4 1.4 0 1.8 3.6 3.3 3 ...
## $ TOV%   : num  16.4 14.2 17.2 9 8.5 14.1 18.7 8.8 11.9 5.8 ...
## $ USG%   : num  17.6 22.9 19.2 17.3 20.1 13.7 17.7 20.6 19.8 24.2 ...
## $ OWS    : num  -0.2 0.2 -0.2 3.2 2 -0.2 -0.9 4.9 3.6 1 ...
## $ DWS    : num  0.2 0.7 0.5 2.2 1.7 0.1 0.4 4.5 2.7 1.8 ...
## $ WS     : num  0 0.9 0.3 5.4 3.7 0 -0.5 9.4 6.3 2.8 ...
## $ WS/48  : num  -0.001 0.04 0.016 0.139 0.076 -0.014 -0.047 0.172 0.137 0.123 ...
## $ OBPM   : num  -7.5 -0.5 -2.1 0.6 -0.2 -5.6 -5.9 1.5 1 -2.3 ...
## $ DBPM   : num  1.9 -2.8 -2.6 1.2 -0.4 0.1 -0.2 2.6 2.1 1.2 ...
## $ BPM    : num  -5.6 -3.3 -4.6 1.8 -0.7 -5.5 -6.1 4.1 3.1 -1.1 ...
## $ VORP   : num  -0.1 -0.4 -0.6 1.8 0.8 -0.1 -0.5 4.1 2.8 0.2 ...
## $ FG     : num  17 188 121 274 393 5 53 529 379 245 ...
## $ FGA    : num  42 469 300 579 865 ...
## $ FG%    : num  0.405 0.401 0.403 0.473 0.454 0.263 0.366 0.505 0.473 0.485 ...
## $ 3P     : num  5 66 48 42 77 3 9 88 86 0 ...
## $ 3PA    : num  10 185 128 142 267 10 32 256 242 0 ...
## $ 3P%    : num  0.5 0.357 0.375 0.296 0.288 0.3 0.281 0.344 0.355 0 ...
## $ 2P     : num  12 122 73 232 316 2 44 441 293 245 ...
## $ 2PA    : num  32 284 172 437 598 9 113 792 559 505 ...
## $ 2P%    : num  0.375 0.43 0.424 0.531 0.528 0.222 0.389 0.557 0.524 0.485 ...
## $ eFG%   : num  0.464 0.471 0.483 0.509 0.499 0.342 0.397 0.547 0.527 0.485 ...
## $ FT     : num  9 49 32 129 156 5 17 103 108 72 ...
## $ FTA    : num  20 64 40 193 217 12 26 129 135 111 ...
## $ FT%    : num  0.45 0.766 0.8 0.668 0.719 0.417 0.654 0.798 0.8 0.649 ...
## $ ORB    : num  8 21 18 154 116 4 20 148 95 57 ...
## $ DRB    : num  28 80 51 353 289 11 91 448 369 244 ...
## $ TRB    : num  36 101 69 507 405 15 111 596 464 301 ...
## $ AST    : num  4 180 125 128 150 2 29 263 337 70 ...
## $ STL    : num  1 30 25 59 64 6 16 68 52 30 ...
## $ BLK    : num  13 10 9 55 40 0 11 121 87 41 ...
## $ TOV    : num  10 82 66 66 89 4 36 107 116 34 ...
## $ PF     : num  21 132 93 153 172 10 77 163 138 117 ...
## $ PTS    : num  48 491 322 719 1019 ...
```

Histogram Barcharts for Numeric Variables

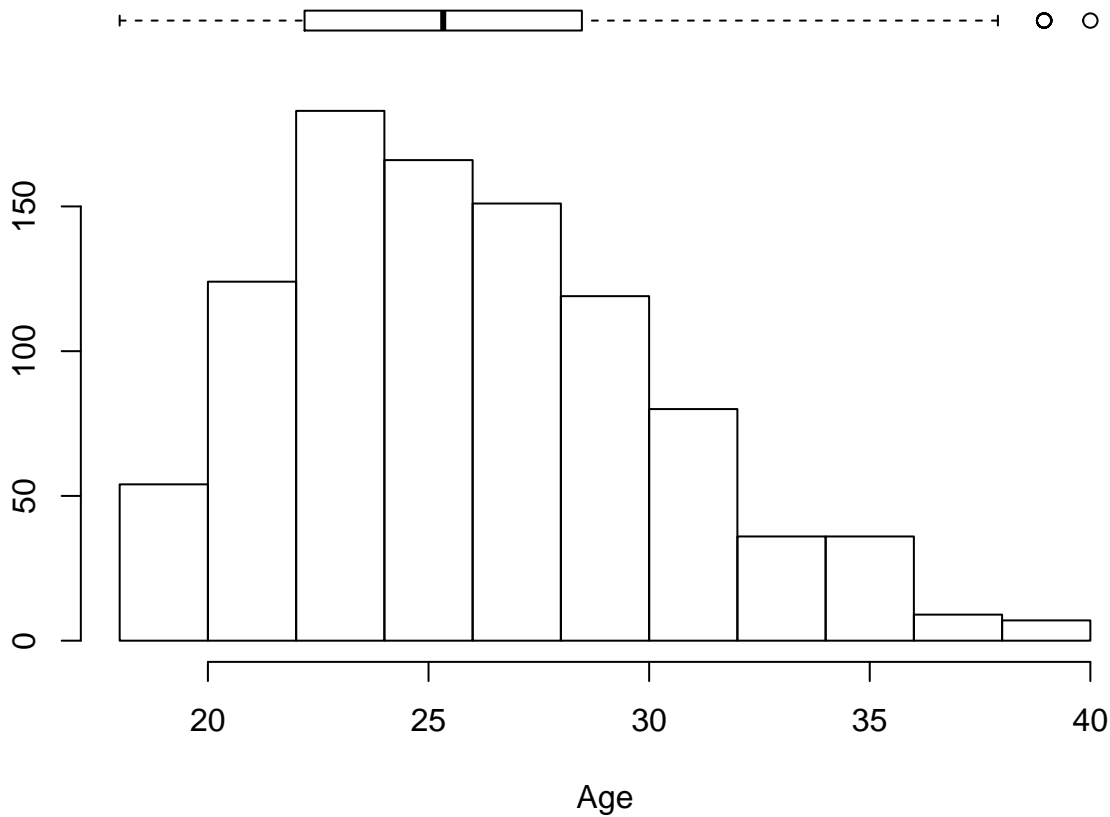
```
df_p_numeric <- Filter(is.numeric,df_primary) # numeric variables
for (col in names(df_p_numeric)){
  data <- df_p_numeric[[col]]
  layout(mat = matrix(c(1,2),2,1, byrow=TRUE), height = c(1,8))
  par(mar=c(0, 3.1, 1.1, 2.1))
  boxplot(data , horizontal=TRUE , xaxt="n", frame=F, main=sprintf('Histogram of %s',col))
  par(mar=c(4, 3.1, 1.1, 2.1))
  hist(data,xlab=col,main='')
  # print top players in this category
  cat(sprintf('Top 10 Players by %s\n',col))
  df_top <- df_primary[order(df_primary[[col]],decreasing=T),]
  print(df_top[1:10,])}
```

Histogram of salary



```
## Top 10 Players by salary
## # A tibble: 10 x 51
##   year name_p salary Pos Age Tm G GS MP PER `TS%` `3Par`
##   <fct> <chr> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Steph~ 3.47e7 PG 28 GSW 79 79 2638 24.6 0.624 0.547
## 2 2017 LeBro~ 3.33e7 SF 32 CLE 74 74 2794 27 0.619 0.254
## 3 2017 Paul ~ 3.13e7 PF 31 ATL 69 67 2343 17.8 0.542 0.248
## 4 2016 LeBro~ 3.10e7 SF 31 CLE 76 76 2709 27.5 0.588 0.199
## 5 2017 Gordo~ 2.97e7 SF 26 UTA 73 73 2516 22.2 0.595 0.324
## 6 2017 Blake~ 2.95e7 PF 27 LAC 61 61 2076 22.7 0.569 0.116
## 7 2017 Kyle ~ 2.87e7 PG 30 TOR 60 60 2244 22.9 0.623 0.51
## 8 2017 Mike ~ 2.85e7 PG 29 MEM 69 68 2292 23.2 0.604 0.415
## 9 2017 Russe~ 2.85e7 PG 28 OKC 81 81 2802 30.6 0.554 0.3
## 10 2017 James~ 2.83e7 PG 27 HOU 81 81 2947 27.3 0.613 0.493
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

Histogram of Age

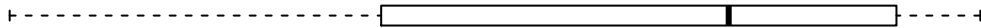


Top 10 Players by Age

A tibble: 10 x 51

```
##   year  name_p salary Pos   Age Tm    G    GS    MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Vince~ 8.00e6 SF    40 MEM    73   15  1799  11.7 0.542 0.604
## 2 2017 Jason~ 2.33e6 SG    39 MIL    74    0  1365   9   0.6   0.704
## 3 2016 Kevin~ 8.00e6 PF    39 MIN    38   38   556  12.3 0.491 0.009
## 4 2017 Manu ~ 2.50e6 SG    39 SAS    69    0  1291  13.9 0.532 0.517
## 5 2017 Paul ~ 1.10e6 SF    39 LAC    25    7   277   5.7 0.535 0.614
## 6 2016 Tim D~ 1.88e6 C     39 SAS    61   60  1536  16.9 0.523 0.005
## 7 2016 Vince~ 4.26e6 SG    39 MEM    60    3  1005  12.7 0.52   0.493
## 8 2017 Dirk ~ 5.00e6 PF    38 DAL    54   54  1424  17   0.529 0.308
## 9 2016 Jason~ 1.55e6 SG    38 HOU    72    7  1258  10.2 0.54   0.694
## 10 2016 Manu ~ 1.40e7 SG    38 SAS    58    0  1134  17.8 0.573 0.411
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

A horizontal dashed line represents a 1D lattice. A central segment of the line is enclosed in a red rectangular box, representing a unit cell. The box is divided into two equal halves by a vertical black line, representing the two atoms within the unit cell. The dashed line extends to the left and right of the box, with small vertical tick marks at the ends.

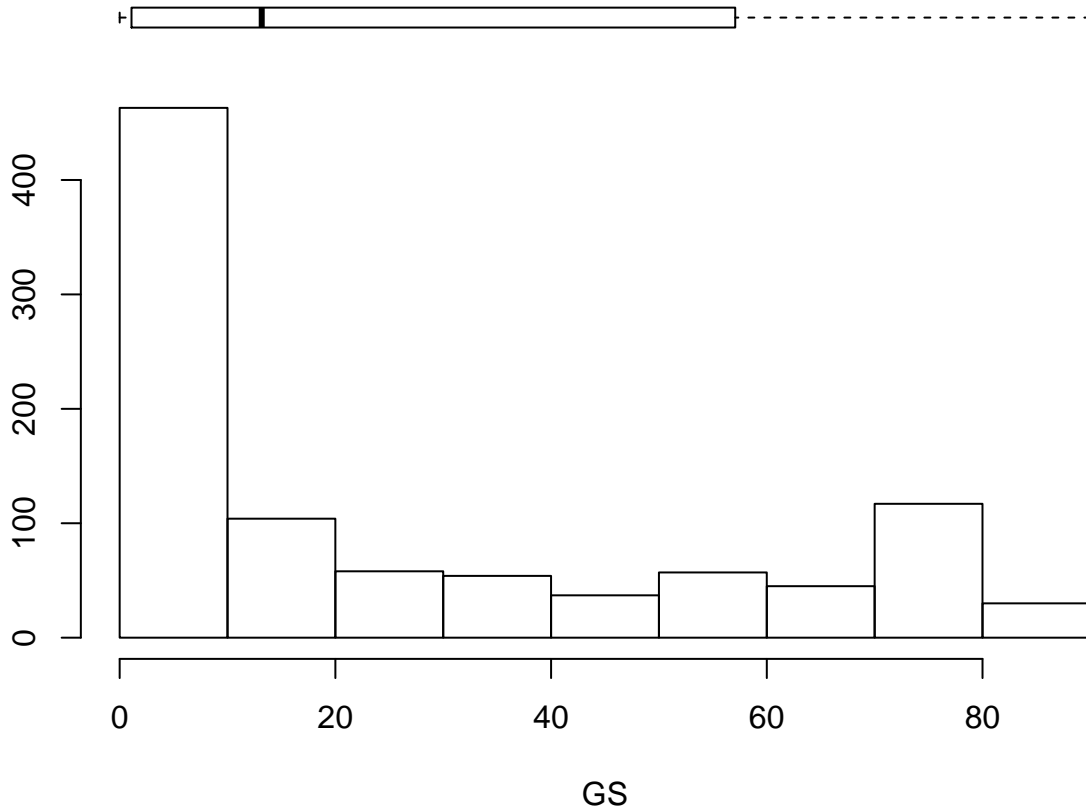


```
## # A tibble: 10 x 51
```

```
##      year  name_p  sala
```

```
##      <fct> <chr>      <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Al Ho~ 2.65e7 C      29 ATL      82      82 2631 19.4 0.565 0.244
## 2 2016 Al-Fa~ 7.68e6 SF     25 POR      82      82 2341 12.7 0.533 0.485
## 3 2017 Andre~ 7.57e6 SF     21 MIN      82      82 3048 16.5 0.534 0.184
## 4 2016 Bisma~ 1.70e7 C      23 TOR      82     22 1808 14.9 0.586 0.003
## 5 2017 Buddy~ 3.68e6 SG     23 TOT      82     55 1888 11.8 0.54  0.493
## 6 2016 Corey~ 7.60e6 SF     29 HOU      82     12 1669  9.9 0.481 0.406
## 7 2017 Corey~ 7.58e6 SF     30 TOT      82     11 1281  9.1 0.491 0.339
## 8 2017 Elfri~ 3.33e6 PG     22 ORL      82     58 2412 17.2 0.52  0.16
## 9 2016 Enes ~ 1.71e7 C      23 OKC      82      1 1721  24  0.626 0.029
## 10 2017 Ersan~ 6.00e6 PF     29 TOT      82     52 2142 14.6 0.546 0.452
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

Histogram of GS



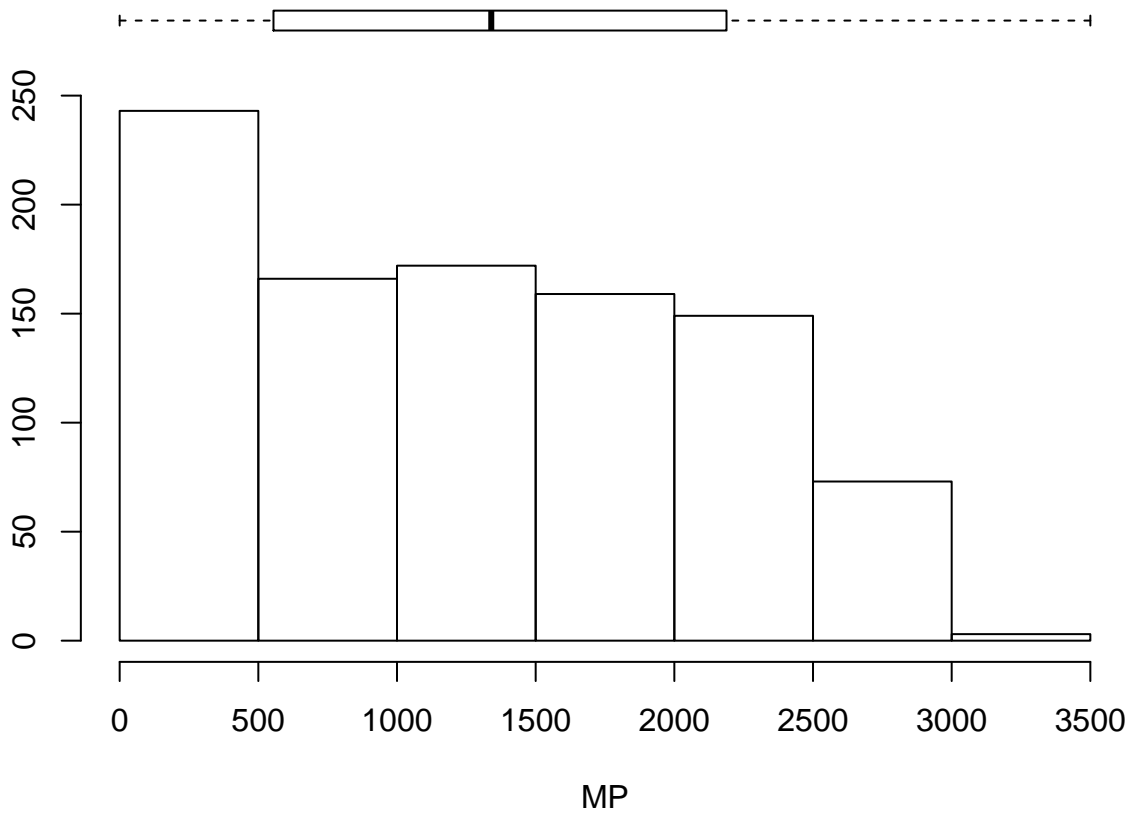
Top 10 Players by GS

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Al Ho~ | 2.65e7 | C | 29 | ATL | 82 | 82 | 2631 | 19.4 | 0.565 | 0.244 |
| ## | 2 | 2016 | Al-Fa~ | 7.68e6 | SF | 25 | POR | 82 | 82 | 2341 | 12.7 | 0.533 | 0.485 |
| ## | 3 | 2017 | Andre~ | 7.57e6 | SF | 21 | MIN | 82 | 82 | 3048 | 16.5 | 0.534 | 0.184 |
| ## | 4 | 2017 | Gorgu~ | 1.41e7 | PF | 27 | MIN | 82 | 82 | 2653 | 14.2 | 0.555 | 0.065 |
| ## | 5 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 | 0.406 |
| ## | 6 | 2017 | Jeff ~ | 1.90e7 | PG | 28 | IND | 82 | 82 | 2657 | 19.2 | 0.574 | 0.277 |
| ## | 7 | 2016 | Karl~ | 5.96e6 | C | 20 | MIN | 82 | 82 | 2627 | 22.5 | 0.59 | 0.076 |
| ## | 8 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 9 | 2017 | Marci~ | 1.28e7 | C | 32 | WAS | 82 | 82 | 2556 | 15.5 | 0.593 | 0.003 |
| ## | 10 | 2016 | Mason~ | 2.33e6 | C | 25 | POR | 82 | 82 | 2084 | 17.2 | 0.564 | 0.008 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of MP



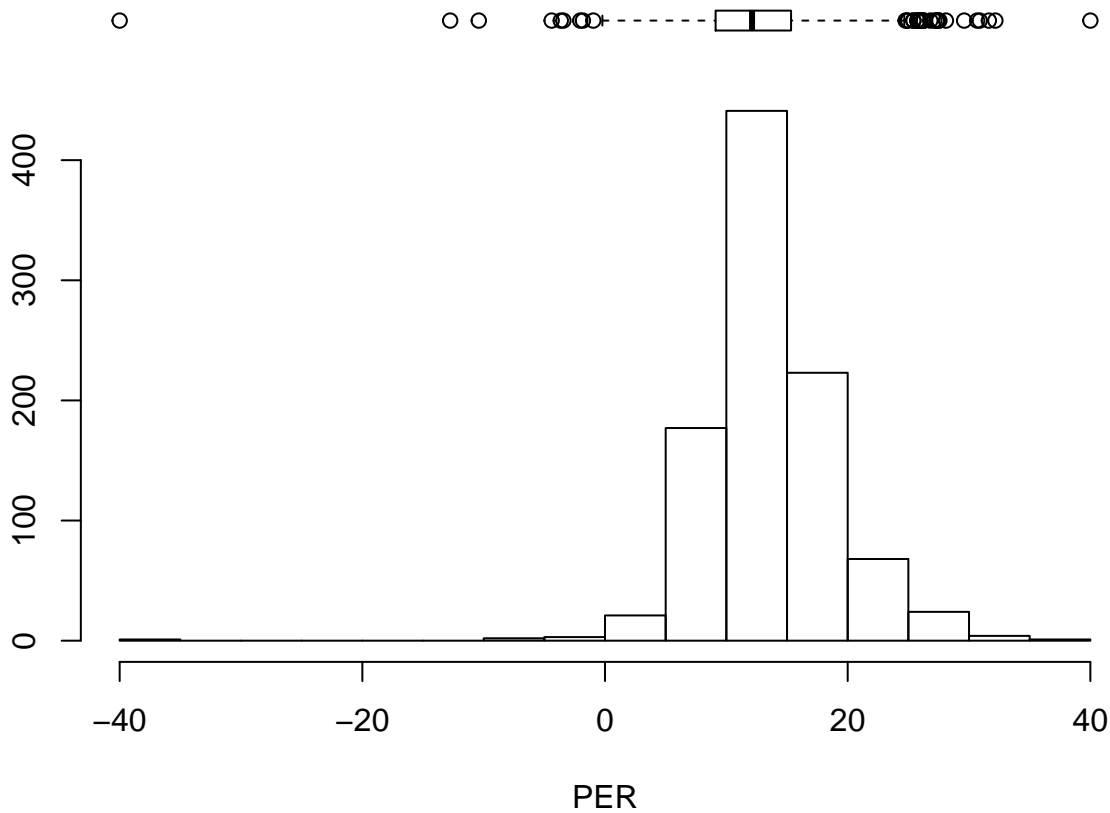
Top 10 Players by MP

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 | 0.406 |
| ## | 2 | 2017 | Andre~ | 7.57e6 | SF | 21 | MIN | 82 | 82 | 3048 | 16.5 | 0.534 | 0.184 |
| ## | 3 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 4 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 | 0.493 |
| ## | 5 | 2016 | Gordo~ | 1.61e7 | SF | 25 | UTA | 80 | 80 | 2893 | 18.3 | 0.559 | 0.341 |
| ## | 6 | 2016 | Kemba~ | 1.20e7 | PG | 25 | CHO | 81 | 81 | 2885 | 20.8 | 0.554 | 0.368 |
| ## | 7 | 2016 | Trevo~ | 7.81e6 | SF | 30 | HOU | 81 | 81 | 2859 | 12.9 | 0.551 | 0.581 |
| ## | 8 | 2016 | Marcu~ | 4.62e6 | SF | 26 | DET | 80 | 80 | 2856 | 12.7 | 0.531 | 0.315 |
| ## | 9 | 2016 | Khris~ | 1.52e7 | SG | 24 | MIL | 79 | 79 | 2852 | 16.8 | 0.56 | 0.316 |
| ## | 10 | 2016 | Kyle ~ | 1.20e7 | PG | 29 | TOR | 77 | 77 | 2851 | 22.2 | 0.578 | 0.457 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
`TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
`USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
`3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of PER



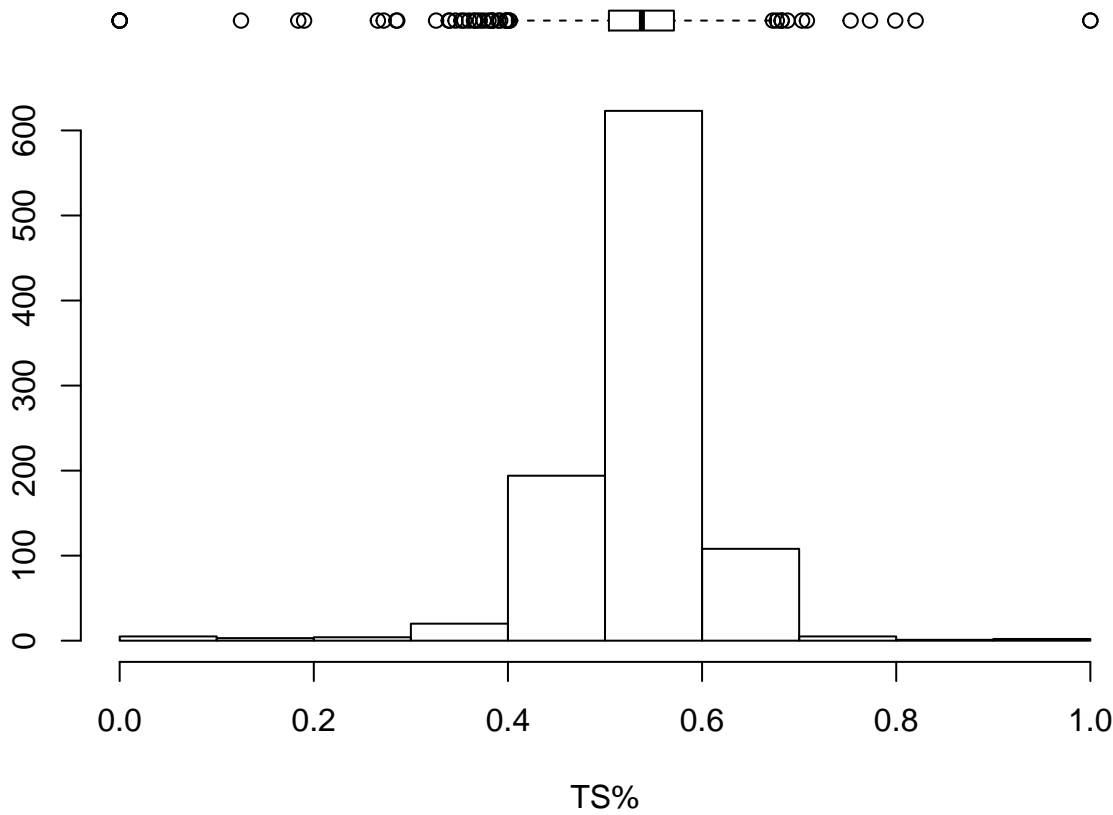
Top 10 Players by PER

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Brian~ | 3.28e5 | PG | 23 | MIA | 1 | 0 | 3 | 39.3 | 1 | 0 |
| ## | 2 | 2016 | Rakee~ | 1.05e6 | PF | 24 | IND | 1 | 0 | 6 | 32 | 1 | 0 |
| ## | 3 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 | 0.554 |
| ## | 4 | 2017 | Demet~ | 9.29e4 | PG | 22 | BOS | 5 | 0 | 17 | 30.8 | 0.753 | 0.25 |
| ## | 5 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 6 | 2017 | Boban~ | 7.00e6 | C | 28 | DET | 35 | 0 | 293 | 29.6 | 0.606 | 0 |
| ## | 7 | 2016 | Kevin~ | 2.65e7 | SF | 27 | OKC | 72 | 72 | 2578 | 28.2 | 0.634 | 0.348 |
| ## | 8 | 2016 | Boban~ | 7.00e6 | C | 27 | SAS | 54 | 4 | 508 | 27.7 | 0.662 | 0 |
| ## | 9 | 2017 | Kevin~ | 2.50e7 | SF | 28 | GSW | 62 | 62 | 2070 | 27.6 | 0.651 | 0.304 |
| ## | 10 | 2016 | Russe~ | 2.65e7 | PG | 27 | OKC | 80 | 80 | 2750 | 27.6 | 0.554 | 0.236 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of TS%



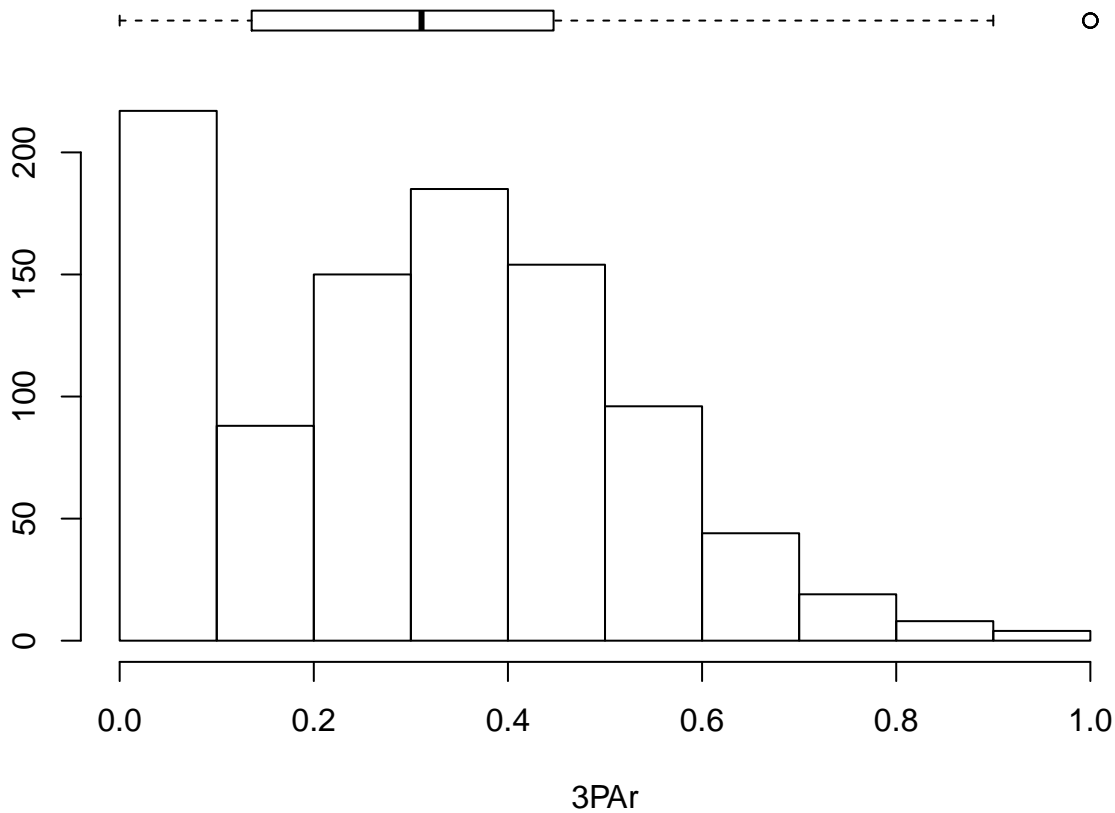
Top 10 Players by TS%

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | TS% | 3PAr |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## | 1 | 2016 | Brian~ | 3.28e5 | PG | 23 | MIA | 1 | 0 | 3 | 39.3 | 1 |
| ## | 2 | 2016 | Rakee~ | 1.05e6 | PF | 24 | IND | 1 | 0 | 6 | 32 | 1 |
| ## | 3 | 2017 | Wayne~ | 1.31e6 | SG | 22 | NOP | 3 | 3 | 47 | 10 | 0.82 |
| ## | 4 | 2017 | China~ | 1.31e6 | C | 20 | HOU | 5 | 1 | 52 | 12.3 | 0.799 |
| ## | 5 | 2017 | Jarre~ | 2.33e6 | PG | 33 | NOP | 2 | 0 | 33 | 7.7 | 0.773 |
| ## | 6 | 2017 | Demet~ | 9.29e4 | PG | 22 | BOS | 5 | 0 | 17 | 30.8 | 0.753 |
| ## | 7 | 2016 | Steve~ | 1.55e6 | PF | 32 | OKC | 7 | 0 | 24 | 20.8 | 0.708 |
| ## | 8 | 2017 | Tyson~ | 1.30e7 | C | 34 | PHO | 47 | 46 | 1298 | 16.6 | 0.703 |
| ## | 9 | 2017 | Axel ~ | 2.50e4 | SF | 24 | NOP | 2 | 0 | 41 | 8.6 | 0.688 |
| ## | 10 | 2017 | Lucas~ | 2.95e6 | C | 24 | TOR | 57 | 6 | 1088 | 15.5 | 0.682 |

... with 39 more variables: FTr <dbl>, ORB% <dbl>, DRB% <dbl>, TRB% <dbl>, AST% <dbl>, STL% <dbl>, BLK% <dbl>, TOV% <dbl>, USG% <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, WS/48 <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, FG% <dbl>, 3P <dbl>, 3PA <dbl>, 3P% <dbl>, 2P <dbl>, 2PA <dbl>, 2P% <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>

Histogram of 3PAr



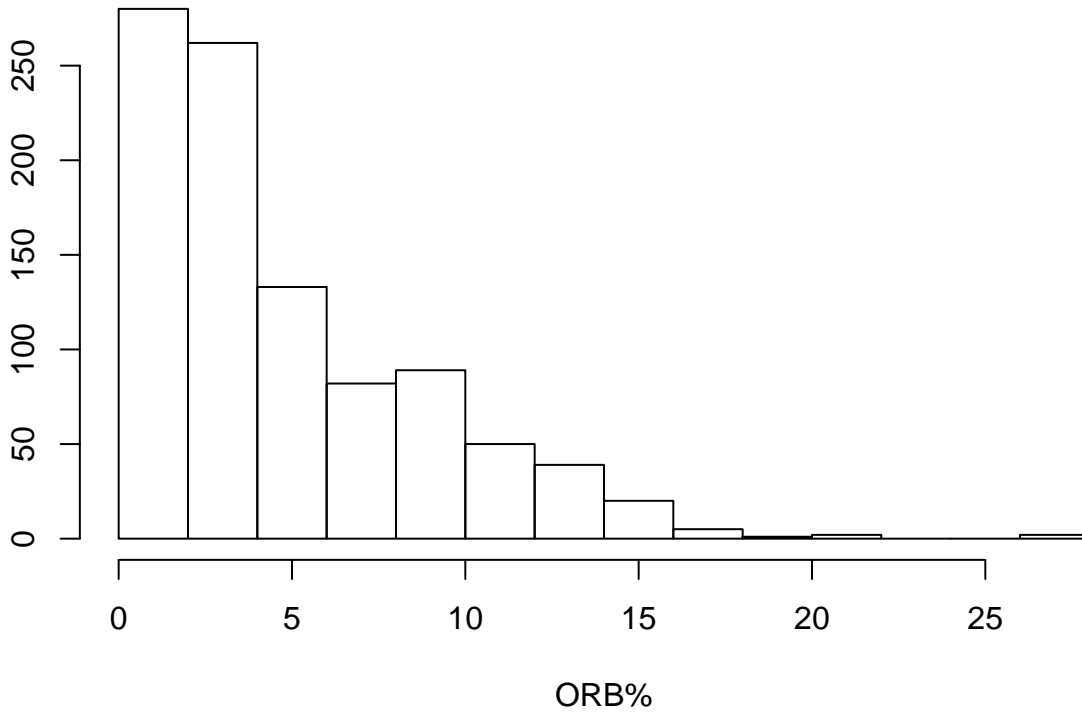
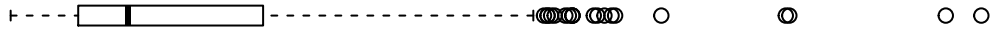
```
## Top 10 Players by 3PAr
## # A tibble: 10 x 51
##   year name_p salary Pos Age Tm G GS MP PER `TS%` `3PAr`
##   <fct> <chr> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Axel ~ 2.50e4 SF 24 MIL 2 0 6 -9.9 0 1
## 2 2017 Chris~ 1.47e6 PF 21 WAS 2 0 8 1.1 0.266 1
## 3 2016 Joe H~ 9.80e5 SG 24 CLE 5 0 15 3.4 0.375 1
## 4 2016 Steve~ 1.55e6 PF 32 MIL 3 0 20 6.7 0.543 1
## 5 2016 Justi~ 5.77e4 PF 26 DET 5 0 35 6.9 0.597 0.9
## 6 2017 Wayne~ 1.31e6 SG 22 NOP 3 3 47 10 0.82 0.875
## 7 2017 Jarel~ 1.72e4 SF 25 PHO 5 0 62 9.7 0.523 0.842
## 8 2016 Mike ~ 3.50e6 SF 35 DEN 47 2 373 6.5 0.508 0.839
## 9 2016 Steve~ 1.55e6 PF 32 TOT 10 0 44 14.4 0.651 0.833
## 10 2016 Antho~ 8.00e6 PF 30 DET 72 5 1341 10.2 0.543 0.819
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```



```
## # A tibble: 10 x 51
```

```
##      year name_p salary Pos      Age Tm      G      GS      MP      PER `TS%` `3PAr`
##      <fct> <chr>      <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Chris~ 1.47e6 PF      21 WAS      2      0      8      1.1 0.266 1
## 2 2017 Demet~ 9.29e4 PG      22 BOS      5      0     17    30.8 0.753 0.25
## 3 2017 Marcu~ 1.31e6 SG      22 ORL      5      0     48    10.2 0.614 0.286
## 4 2016 DeAnd~ 2.12e7 C       27 LAC     77     77   2598   20.6 0.628 0.002
## 5 2016 Jorda~ 1.47e6 SG      21 MEM      2      0     15    17.3 0.427 0.167
## 6 2016 Joel ~ 6.64e5 C       33 DET     19      0     96    14.1 0.666 0
## 7 2016 Rudy ~ 2.12e6 C       23 UTA     61     60   1932   17.5 0.582 0
## 8 2016 Dwigh~ 2.32e7 C       30 HOU     71     71   2280   18.9 0.604 0.01
## 9 2017 Ander~ 1.91e6 C       34 GSW     14      1     92     9.4 0.478 0
## 10 2016 Bisma~ 1.70e7 C       23 TOR     82     22   1808   14.9 0.586 0.003
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

Histogram of ORB%



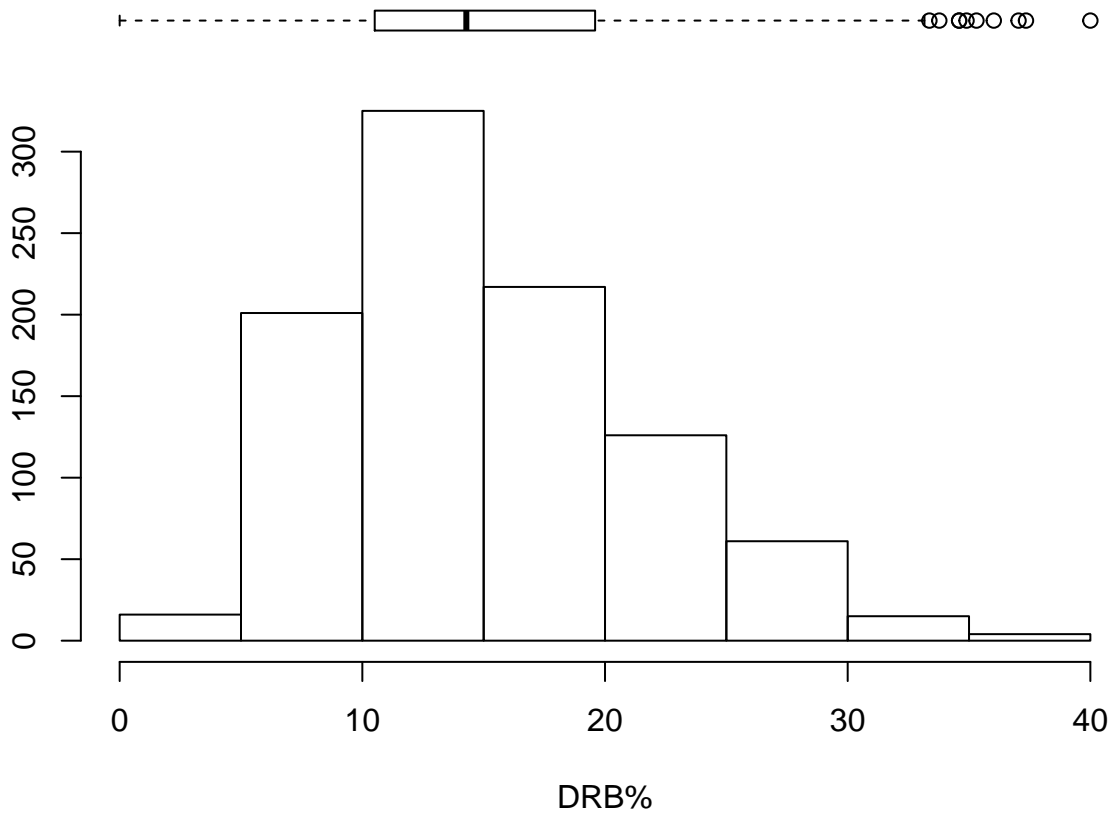
Top 10 Players by ORB%

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## | 1 | 2016 | Jarne~ | 1.50e5 | C | 22 | MEM | 2 | 0 | 4 | 13.6 | 0 |
| ## | 2 | 2017 | Larry~ | 1.87e6 | C | 28 | CLE | 5 | 0 | 13 | 6.5 | 0.41 |
| ## | 3 | 2016 | Alan ~ | 8.75e5 | PF | 23 | PHO | 10 | 0 | 68 | 21.1 | 0.481 |
| ## | 4 | 2016 | Kevon~ | 1.18e6 | PF | 19 | GSW | 5 | 0 | 21 | 18.6 | 0.643 |
| ## | 5 | 2016 | Rakee~ | 1.05e6 | PF | 24 | IND | 1 | 0 | 6 | 32 | 1 |
| ## | 6 | 2017 | Joaki~ | 1.78e7 | C | 31 | NYK | 46 | 46 | 1015 | 15.2 | 0.493 |
| ## | 7 | 2016 | Boban~ | 7.00e6 | C | 27 | SAS | 54 | 4 | 508 | 27.7 | 0.662 |
| ## | 8 | 2016 | Enes ~ | 1.71e7 | C | 23 | OKC | 82 | 1 | 1721 | 24 | 0.626 |
| ## | 9 | 2017 | Boban~ | 7.00e6 | C | 28 | DET | 35 | 0 | 293 | 29.6 | 0.606 |
| ## | 10 | 2016 | Thoma~ | 1.05e6 | PF | 24 | BRK | 71 | 7 | 917 | 14.5 | 0.453 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of DRB%



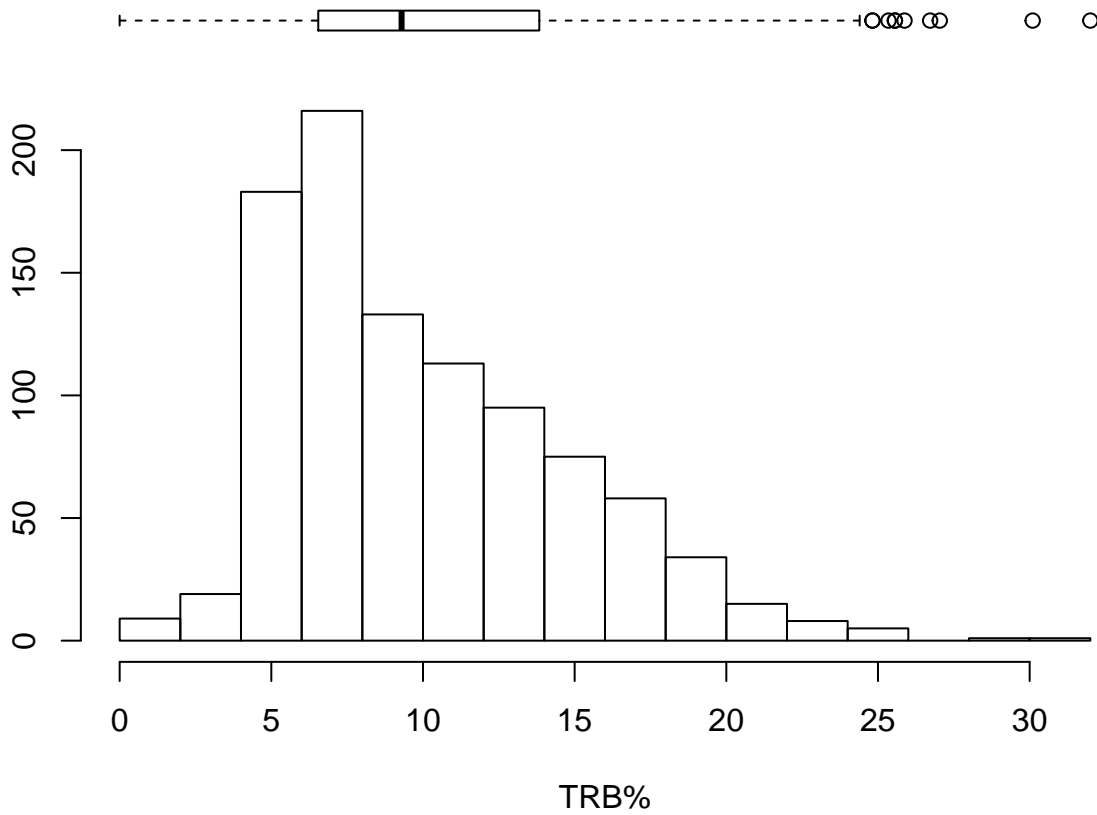
Top 10 Players by DRB%

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Alan ~ | 8.75e5 | PF | 23 | PHO | 10 | 0 | 68 | 21.1 | 0.481 | 0 |
| ## | 2 | 2016 | Brian~ | 3.28e5 | PG | 23 | MIA | 1 | 0 | 3 | 39.3 | 1 | 0 |
| ## | 3 | 2017 | Andre~ | 2.38e7 | C | 23 | DET | 81 | 81 | 2409 | 20.9 | 0.518 | 0.008 |
| ## | 4 | 2017 | Hassa~ | 2.38e7 | C | 27 | MIA | 77 | 77 | 2513 | 22.6 | 0.579 | 0 |
| ## | 5 | 2017 | DeAnd~ | 2.26e7 | C | 28 | LAC | 81 | 81 | 2570 | 21.8 | 0.673 | 0.003 |
| ## | 6 | 2016 | Andre~ | 2.21e7 | C | 22 | DET | 81 | 81 | 2666 | 21.2 | 0.499 | 0.006 |
| ## | 7 | 2017 | Andre~ | 2.33e6 | C | 32 | TOT | 27 | 21 | 583 | 9.3 | 0.46 | 0.012 |
| ## | 8 | 2017 | Andre~ | 2.33e6 | C | 32 | DAL | 26 | 21 | 582 | 9.4 | 0.46 | 0.012 |
| ## | 9 | 2017 | Tyson~ | 1.30e7 | C | 34 | PHO | 47 | 46 | 1298 | 16.6 | 0.703 | 0 |
| ## | 10 | 2016 | DeAnd~ | 2.12e7 | C | 27 | LAC | 77 | 77 | 2598 | 20.6 | 0.628 | 0.002 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of TRB%



Top 10 Players by TRB%

A tibble: 10 x 51

| | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` |
|-------|-------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## 1 | 2016 | Alan ~ | 8.75e5 | PF | 23 | PHO | 10 | 0 | 68 | 21.1 | 0.481 | 0 |
| ## 2 | 2016 | Jarne~ | 1.50e5 | C | 22 | MEM | 2 | 0 | 4 | 13.6 | 0 | 0 |
| ## 3 | 2016 | Kevon~ | 1.18e6 | PF | 19 | GSW | 5 | 0 | 21 | 18.6 | 0.643 | 0.286 |
| ## 4 | 2017 | Andre~ | 2.38e7 | C | 23 | DET | 81 | 81 | 2409 | 20.9 | 0.518 | 0.008 |
| ## 5 | 2016 | Andre~ | 2.21e7 | C | 22 | DET | 81 | 81 | 2666 | 21.2 | 0.499 | 0.006 |
| ## 6 | 2017 | Boban~ | 7.00e6 | C | 28 | DET | 35 | 0 | 293 | 29.6 | 0.606 | 0 |
| ## 7 | 2017 | DeAnd~ | 2.26e7 | C | 28 | LAC | 81 | 81 | 2570 | 21.8 | 0.673 | 0.003 |
| ## 8 | 2017 | Hassa~ | 2.38e7 | C | 27 | MIA | 77 | 77 | 2513 | 22.6 | 0.579 | 0 |
| ## 9 | 2017 | Dwigh~ | 2.35e7 | C | 31 | ATL | 74 | 74 | 2199 | 20.8 | 0.627 | 0.003 |
| ## 10 | 2016 | Kris ~ | 4.00e6 | PF | 30 | PHO | 4 | 3 | 74 | 13.5 | 0.367 | 0.278 |

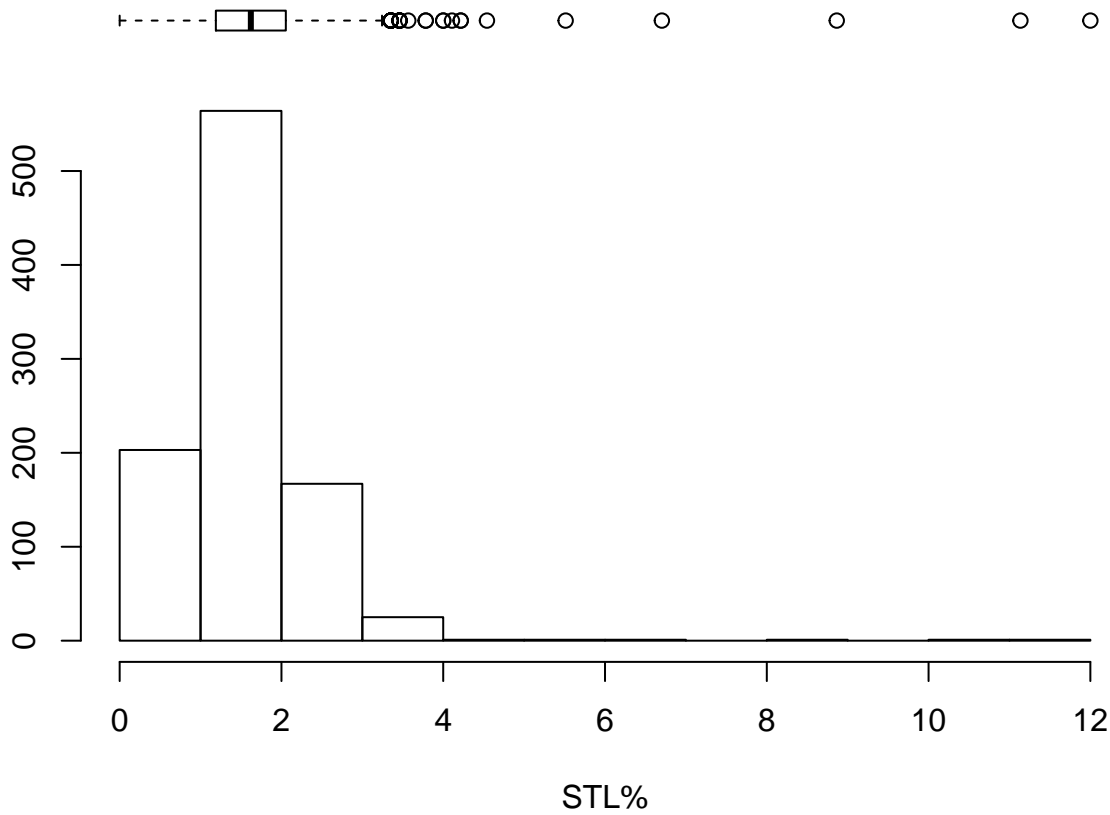
... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>



```
## # A tibble: 10 x 51
```

[illegible]

Histogram of STL%



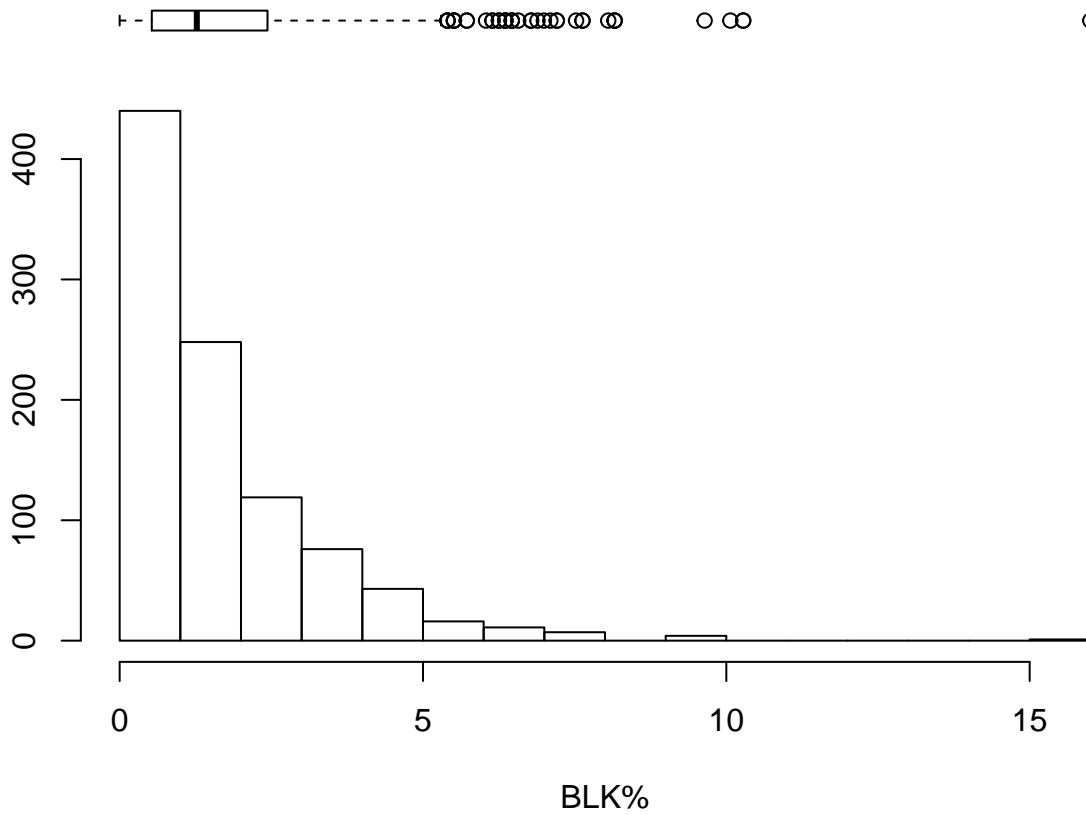
Top 10 Players by STL%

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | Brice~ | 1.33e6 | PF | 22 | LAC | 3 | 0 | 9 | 17.2 | 0.286 | 0 |
| ## | 2 | 2016 | Jorda~ | 1.47e6 | SG | 21 | MEM | 2 | 0 | 15 | 17.3 | 0.427 | 0.167 |
| ## | 3 | 2016 | Sam D~ | 1.72e6 | SF | 21 | HOU | 3 | 0 | 6 | 10.8 | 0 | 0 |
| ## | 4 | 2017 | Chris~ | 1.47e6 | PF | 21 | WAS | 2 | 0 | 8 | 1.1 | 0.266 | 1 |
| ## | 5 | 2016 | James~ | 2.90e6 | SF | 25 | MEM | 10 | 0 | 40 | 18.3 | 0.46 | 0.615 |
| ## | 6 | 2017 | DeAnd~ | 1.58e6 | SG | 28 | DAL | 1 | 0 | 25 | 17.6 | 0.546 | 0.167 |
| ## | 7 | 2016 | Chris~ | 1.19e6 | PF | 20 | BRK | 24 | 4 | 362 | 12.2 | 0.47 | 0.312 |
| ## | 8 | 2017 | Ronni~ | 2.44e6 | PG | 33 | PHO | 14 | 0 | 134 | 5.9 | 0.272 | 0.708 |
| ## | 9 | 2017 | Larry~ | 1.87e6 | C | 28 | CLE | 5 | 0 | 13 | 6.5 | 0.41 | 0 |
| ## | 10 | 2016 | Jarne~ | 1.50e5 | C | 22 | MIA | 5 | 0 | 14 | 21.8 | 0.595 | 0 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of BLK%

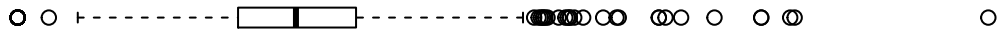


Top 10 Players by BLK%

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Jorda~ | 1.22e6 | PF | 21 | BOS | 16 | 0 | 57 | 15.3 | 0.398 | 0 |
| ## | 2 | 2016 | Hassa~ | 2.21e7 | C | 26 | MIA | 73 | 43 | 2125 | 25.7 | 0.629 | 0 |
| ## | 3 | 2016 | John ~ | 1.25e7 | C | 25 | MIL | 57 | 1 | 960 | 18.6 | 0.580 | 0.003 |
| ## | 4 | 2016 | Joel ~ | 6.64e5 | C | 33 | DET | 19 | 0 | 96 | 14.1 | 0.666 | 0 |
| ## | 5 | 2017 | Brice~ | 1.33e6 | PF | 22 | LAC | 3 | 0 | 9 | 17.2 | 0.286 | 0 |
| ## | 6 | 2017 | Jeram~ | 1.52e6 | SF | 22 | PHI | 2 | 0 | 41 | 3.3 | 0.39 | 0.118 |
| ## | 7 | 2017 | Joel ~ | 6.10e6 | C | 22 | PHI | 31 | 31 | 786 | 24.1 | 0.584 | 0.228 |
| ## | 8 | 2017 | Josh ~ | 1.47e6 | PF | 25 | OKC | 2 | 0 | 31 | 26.1 | 0.612 | 0.364 |
| ## | 9 | 2016 | Salah~ | 8.75e5 | C | 29 | DAL | 34 | 6 | 397 | 16.8 | 0.636 | 0.013 |
| ## | 10 | 2017 | A.J. ~ | 1.31e6 | C | 24 | DAL | 22 | 0 | 163 | 8.4 | 0.472 | 0.238 |

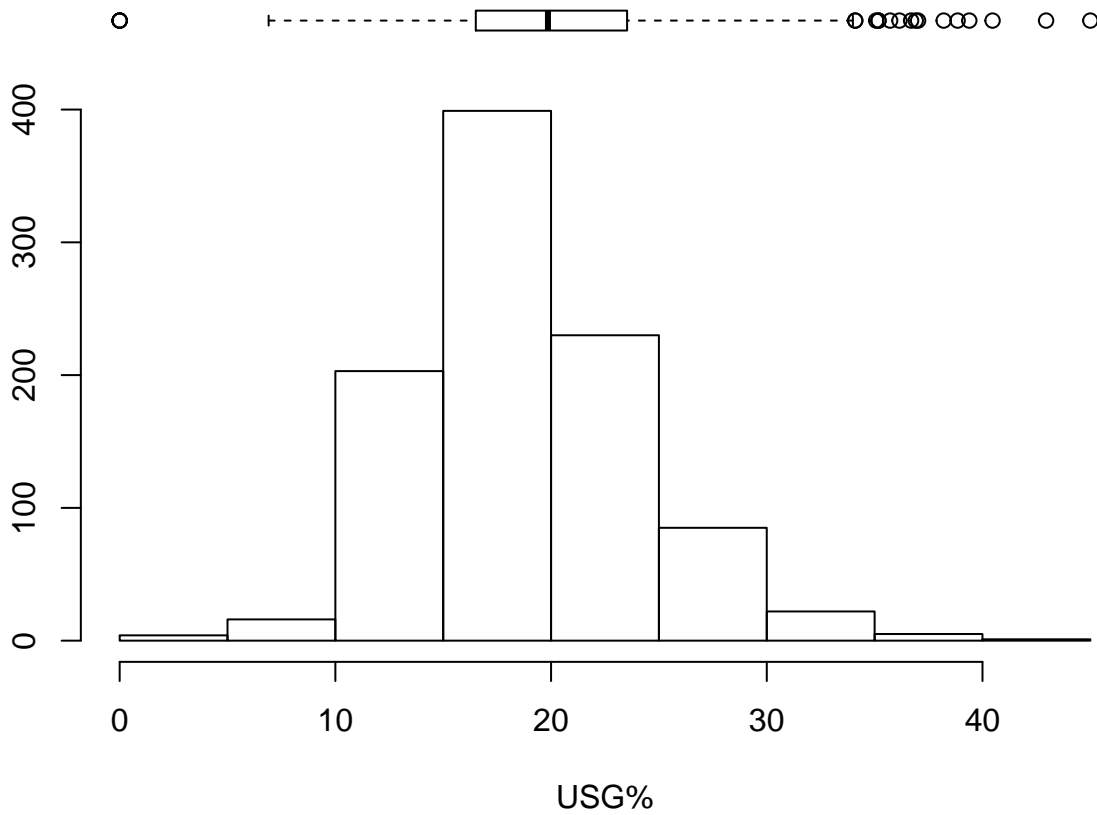
... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>



```
## # A tibble: 10 x 51
```

```
##      year name_p salary Pos       Age Tm           G        GS      MP    PER `TS`%   `3PAr`  
##      <fct> <chr>     <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2017 Jarre~ 2.33e6 PG         33 NOP          2         0     33    7.7 0.773 0.333  
## 2 2016 Phil ~ 3.50e4 PG         24 PHO          9         0    113    8.6 0.422 0.217  
## 3 2017 Chris~ 1.47e6 PF         21 WAS          2         0      8    1.1 0.266 1  
## 4 2017 Andre~ 2.33e6 C         32 TOT         27        21    583    9.3 0.46 0.012  
## 5 2017 Andre~ 2.33e6 C         32 DAL         26        21    582    9.4 0.46 0.012  
## 6 2017 China~ 1.31e6 C         20 HOU          5         1     52   12.3 0.799 0  
## 7 2017 Ander~ 1.91e6 C         34 GSW         14         1     92    9.4 0.478 0  
## 8 2017 Larry~ 1.87e6 C         28 CLE          5         0     13    6.5 0.41 0  
## 9 2016 Nick ~ 3.75e6 PF         35 OKC         59         4    699    7.7 0.498 0.018  
## 10 2016 Tim F~ 2.09e6 PG         25 POR         35         1    272    4.7 0.383 0.283  
## # ... with 39 more variables: FTTr <dbl>, `ORB`%<dbl>, `DRB`%<dbl>,  
## # `TRB`%<dbl>, `AST`%<dbl>, `STL`%<dbl>, `BLK`%<dbl>, `TOV`%<dbl>,  
## # `USG`%<dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48`<dbl>, OBPM <dbl>,  
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG`%<dbl>,  
## # `3P`<dbl>, `3PA`<dbl>, `3P`%<dbl>, `2P`<dbl>, `2PA`<dbl>, `2P`%<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>
```

Histogram of USG%



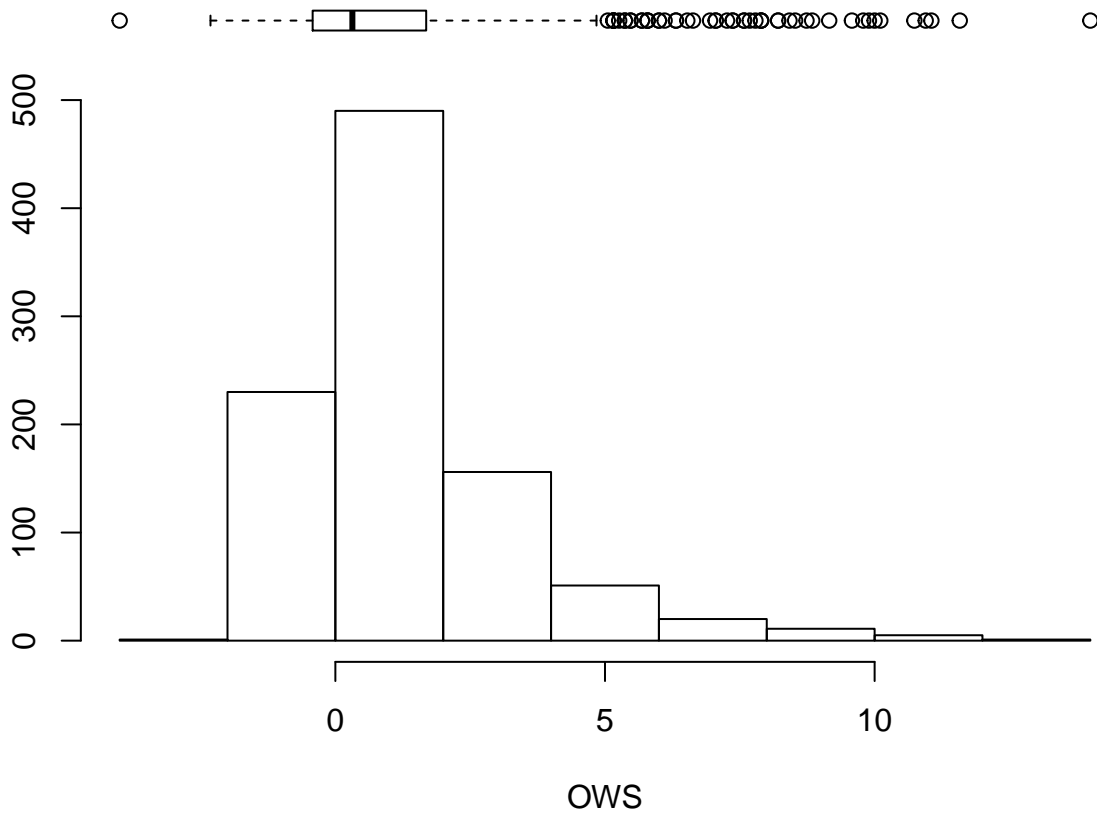
Top 10 Players by USG%

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 2 | 2017 | Brice~ | 1.33e6 | PF | 22 | LAC | 3 | 0 | 9 | 17.2 | 0.286 | 0 |
| ## | 3 | 2017 | DeMar~ | 1.81e7 | C | 26 | SAC | 55 | 55 | 1891 | 26.5 | 0.562 | 0.239 |
| ## | 4 | 2017 | DeMar~ | 1.81e7 | C | 26 | TOT | 72 | 72 | 2465 | 25.7 | 0.562 | 0.254 |
| ## | 5 | 2017 | Joel ~ | 6.10e6 | C | 22 | PHI | 31 | 31 | 786 | 24.1 | 0.584 | 0.228 |
| ## | 6 | 2016 | DeMar~ | 1.70e7 | C | 25 | SAC | 65 | 65 | 2246 | 23.6 | 0.538 | 0.158 |
| ## | 7 | 2017 | DeMar~ | 2.77e7 | SG | 27 | TOR | 74 | 74 | 2620 | 24 | 0.552 | 0.08 |
| ## | 8 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 | 0.493 |
| ## | 9 | 2017 | Isaia~ | 6.26e6 | PG | 27 | BOS | 76 | 76 | 2569 | 26.5 | 0.625 | 0.439 |
| ## | 10 | 2016 | Tony ~ | 2.50e4 | PG | 22 | PHI | 8 | 3 | 144 | 1.8 | 0.412 | 0.262 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of OWS

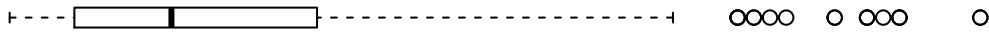


Top 10 Players by OWS

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## | 1 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 |
| ## | 2 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 |
| ## | 3 | 2016 | Kevin~ | 2.65e7 | SF | 27 | OKC | 72 | 72 | 2578 | 28.2 | 0.634 |
| ## | 4 | 2017 | Isaia~ | 6.26e6 | PG | 27 | BOS | 76 | 76 | 2569 | 26.5 | 0.625 |
| ## | 5 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 |
| ## | 6 | 2017 | Jimmy~ | 1.93e7 | SF | 27 | CHI | 76 | 75 | 2809 | 25.1 | 0.586 |
| ## | 7 | 2016 | Russe~ | 2.65e7 | PG | 27 | OKC | 80 | 80 | 2750 | 27.6 | 0.554 |
| ## | 8 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 |
| ## | 9 | 2017 | LeBro~ | 3.33e7 | SF | 32 | CLE | 74 | 74 | 2794 | 27 | 0.619 |
| ## | 10 | 2016 | LeBro~ | 3.10e7 | SF | 31 | CLE | 76 | 76 | 2709 | 27.5 | 0.588 |

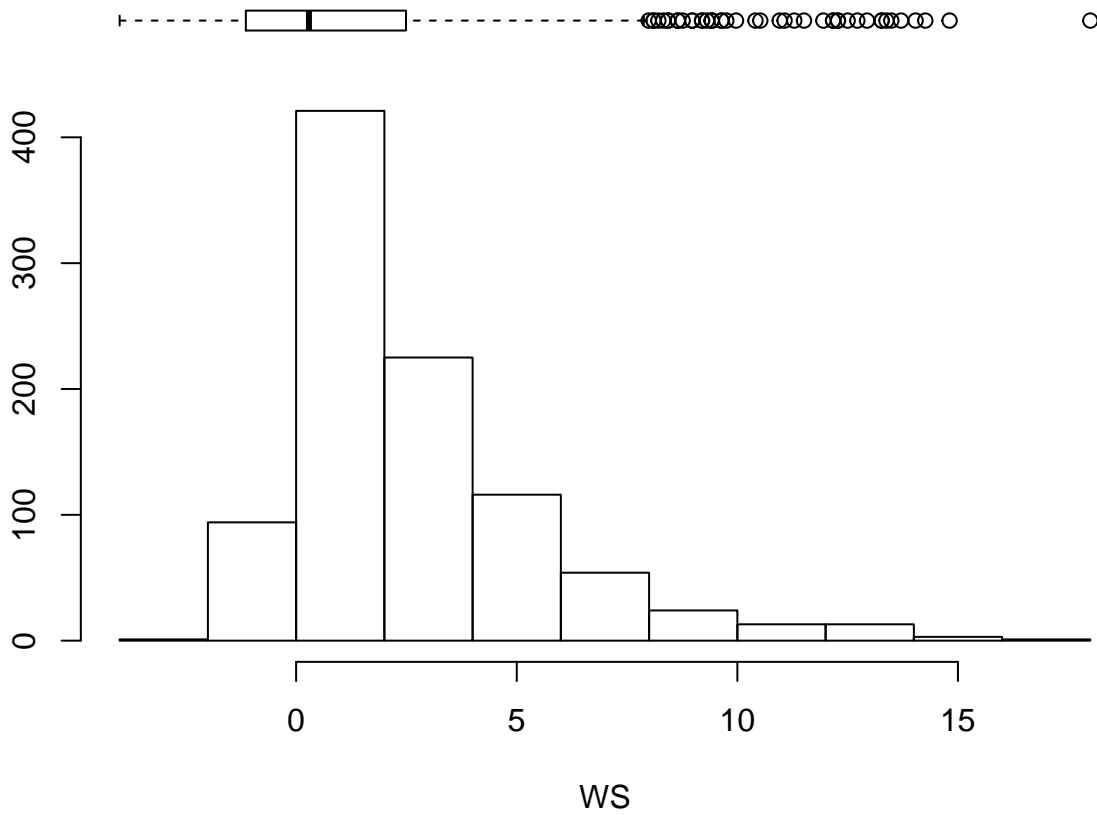
... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>



```
## # A tibble: 10 x 51
```

[illegible]

Histogram of WS



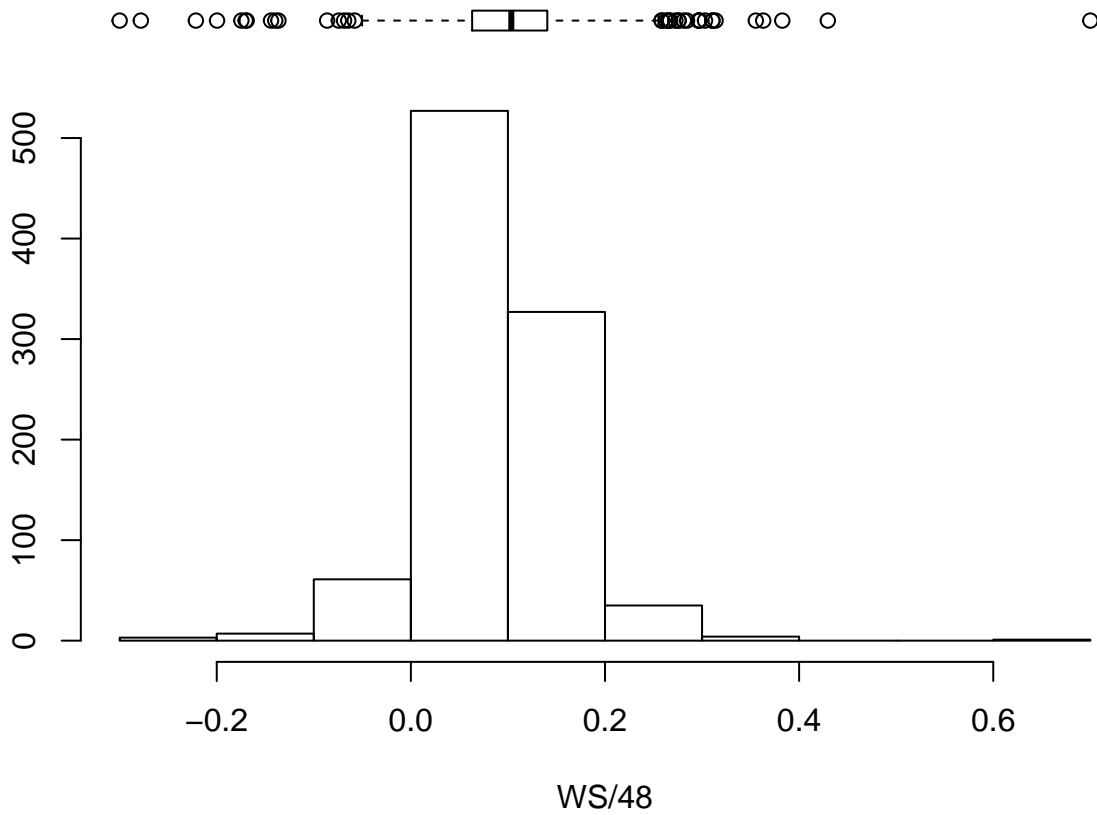
Top 10 Players by WS

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 | 0.554 |
| ## | 2 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 | 0.493 |
| ## | 3 | 2016 | Kevin~ | 2.65e7 | SF | 27 | OKC | 72 | 72 | 2578 | 28.2 | 0.634 | 0.348 |
| ## | 4 | 2017 | Rudy ~ | 2.20e7 | C | 24 | UTA | 81 | 81 | 2744 | 23.3 | 0.682 | 0.002 |
| ## | 5 | 2016 | Russe~ | 2.65e7 | PG | 27 | OKC | 80 | 80 | 2750 | 27.6 | 0.554 | 0.236 |
| ## | 6 | 2017 | Jimmy~ | 1.93e7 | SF | 27 | CHI | 76 | 75 | 2809 | 25.1 | 0.586 | 0.198 |
| ## | 7 | 2016 | Kawhi~ | 1.76e7 | SF | 24 | SAS | 72 | 72 | 2380 | 26 | 0.616 | 0.267 |
| ## | 8 | 2017 | Kawhi~ | 1.89e7 | SF | 25 | SAS | 74 | 74 | 2474 | 27.5 | 0.611 | 0.294 |
| ## | 9 | 2016 | LeBro~ | 3.10e7 | SF | 31 | CLE | 76 | 76 | 2709 | 27.5 | 0.588 | 0.199 |
| ## | 10 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 | 0.406 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of WS/48

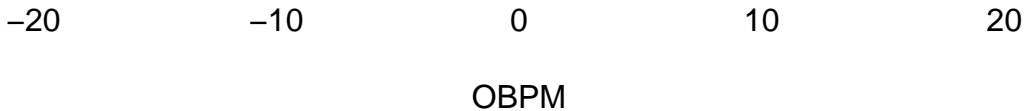


Top 10 Players by WS/48

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Brian~ | 3.28e5 | PG | 23 | MIA | 1 | 0 | 3 | 39.3 | 1 | 0 |
| ## | 2 | 2017 | Demet~ | 9.29e4 | PG | 22 | BOS | 5 | 0 | 17 | 30.8 | 0.753 | 0.25 |
| ## | 3 | 2016 | Rakee~ | 1.05e6 | PF | 24 | IND | 1 | 0 | 6 | 32 | 1 | 0 |
| ## | 4 | 2016 | Boban~ | 7.00e6 | C | 27 | SAS | 54 | 4 | 508 | 27.7 | 0.662 | 0 |
| ## | 5 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 | 0.554 |
| ## | 6 | 2017 | Boban~ | 7.00e6 | C | 28 | DET | 35 | 0 | 293 | 29.6 | 0.606 | 0 |
| ## | 7 | 2016 | Kawhi~ | 1.76e7 | SF | 24 | SAS | 72 | 72 | 2380 | 26 | 0.616 | 0.267 |
| ## | 8 | 2017 | Kevin~ | 2.50e7 | SF | 28 | GSW | 62 | 62 | 2070 | 27.6 | 0.651 | 0.304 |
| ## | 9 | 2016 | Kevin~ | 2.65e7 | SF | 27 | OKC | 72 | 72 | 2578 | 28.2 | 0.634 | 0.348 |
| ## | 10 | 2017 | Josh ~ | 1.47e6 | PF | 25 | OKC | 2 | 0 | 31 | 26.1 | 0.612 | 0.364 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

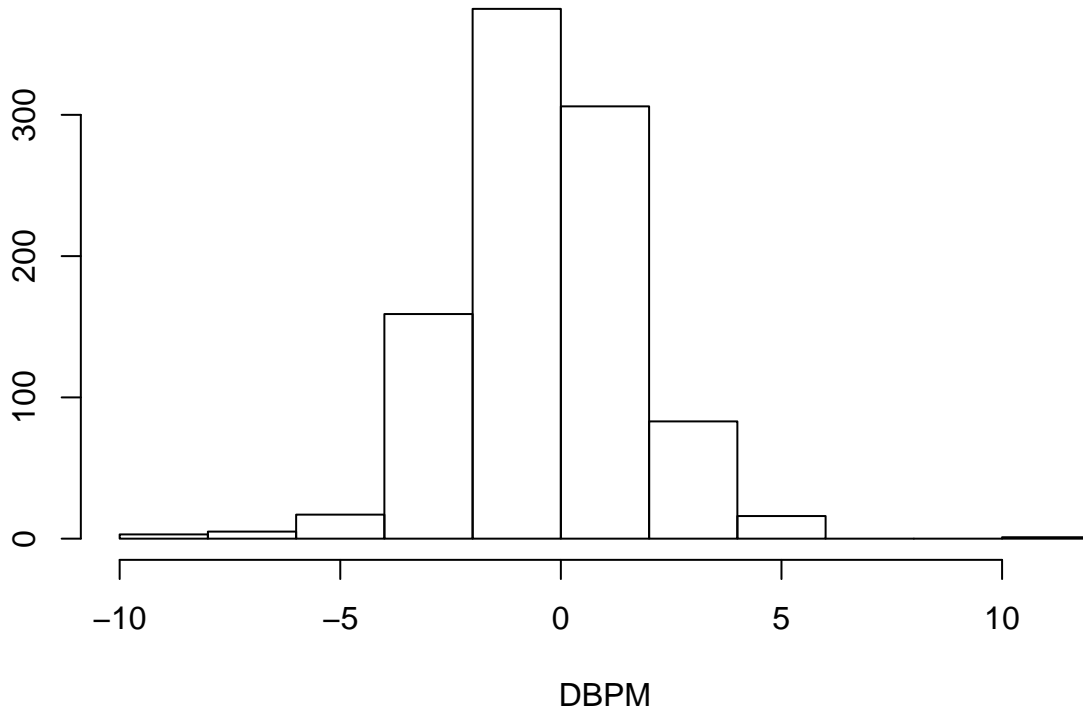


```
## # A tibble: 10 x 51
```

[illegible]

Histogram of DBPM

OO OO OO |-----|-----|OOOO O



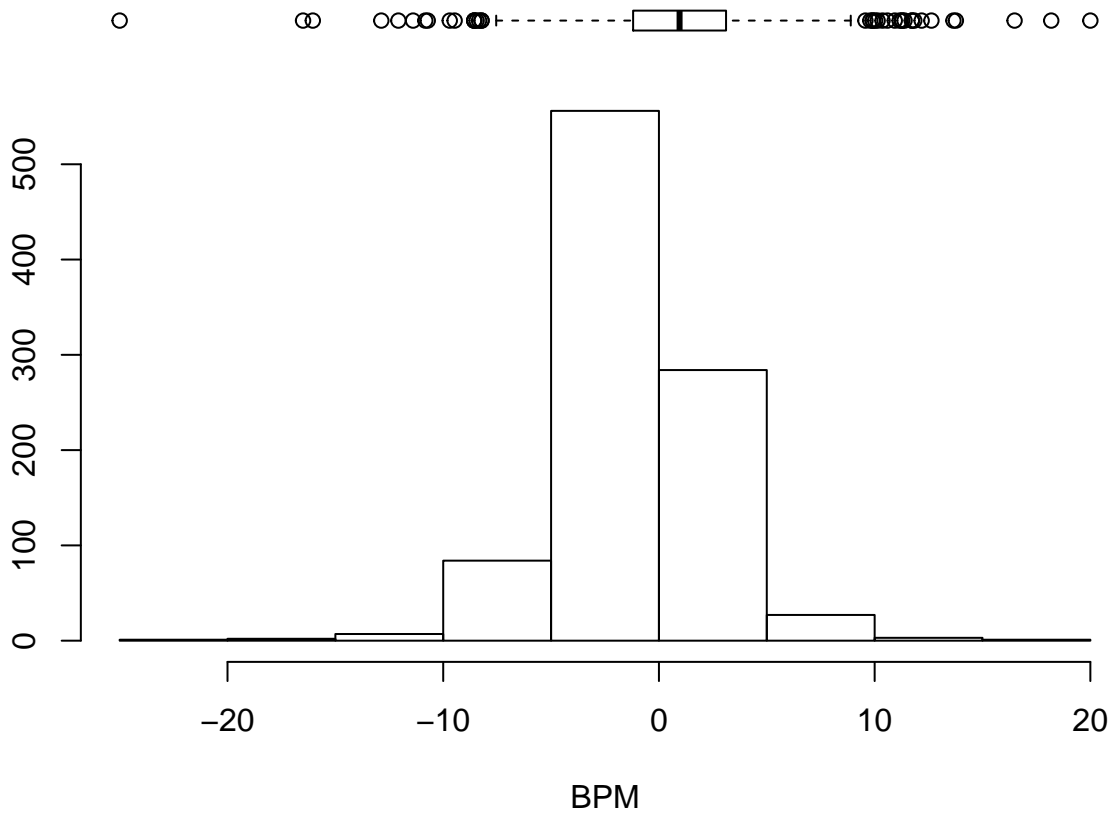
Top 10 Players by DBPM

A tibble: 10 x 51

| | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` |
|-------|-------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## 1 | 2017 | Brice~ | 1.33e6 | PF | 22 | LAC | 3 | 0 | 9 | 17.2 | 0.286 | 0 |
| ## 2 | 2016 | Cole ~ | 7.64e6 | C | 27 | LAC | 60 | 5 | 800 | 21.3 | 0.626 | 0 |
| ## 3 | 2016 | Sam D~ | 1.72e6 | SF | 21 | HOU | 3 | 0 | 6 | 10.8 | 0 | 0 |
| ## 4 | 2017 | Lucas~ | 2.95e6 | C | 24 | TOR | 57 | 6 | 1088 | 15.5 | 0.682 | 0.077 |
| ## 5 | 2017 | Andre~ | 2.33e6 | C | 32 | TOT | 27 | 21 | 583 | 9.3 | 0.46 | 0.012 |
| ## 6 | 2017 | Andre~ | 2.33e6 | C | 32 | DAL | 26 | 21 | 582 | 9.4 | 0.46 | 0.012 |
| ## 7 | 2016 | Andre~ | 1.10e7 | C | 31 | GSW | 70 | 66 | 1451 | 15.9 | 0.623 | 0.004 |
| ## 8 | 2017 | Draym~ | 1.64e7 | PF | 26 | GSW | 76 | 76 | 2471 | 16.5 | 0.522 | 0.405 |
| ## 9 | 2016 | Joel ~ | 6.64e5 | C | 33 | DET | 19 | 0 | 96 | 14.1 | 0.666 | 0 |
| ## 10 | 2016 | Tim D~ | 1.88e6 | C | 39 | SAS | 61 | 60 | 1536 | 16.9 | 0.523 | 0.005 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of BPM

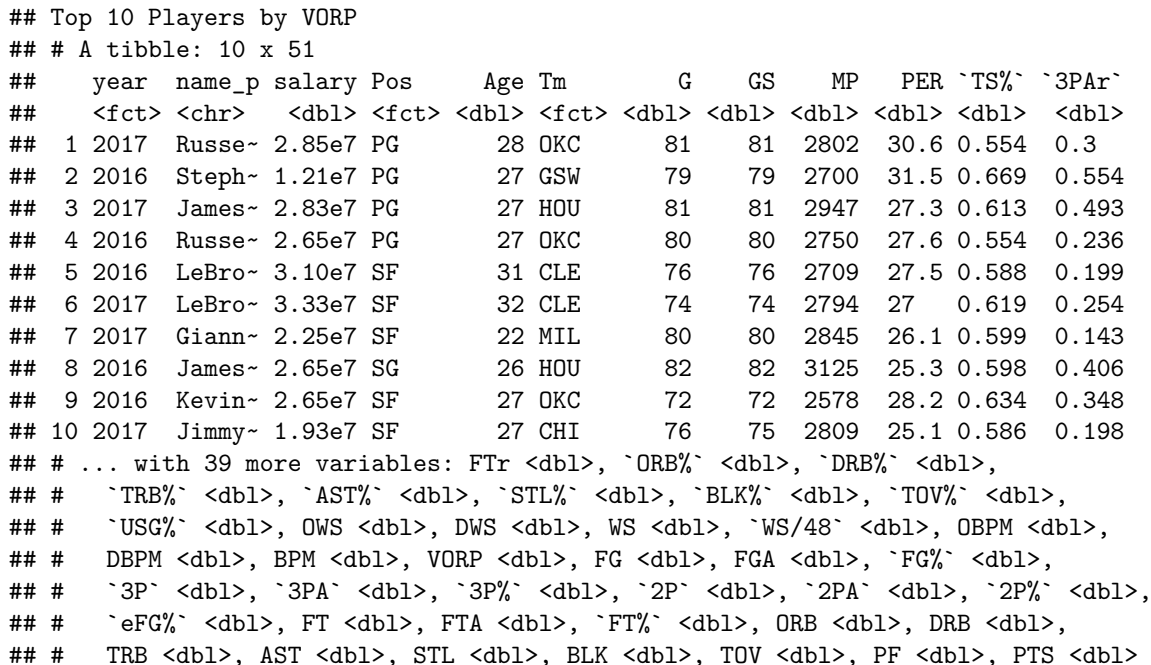


Top 10 Players by BPM

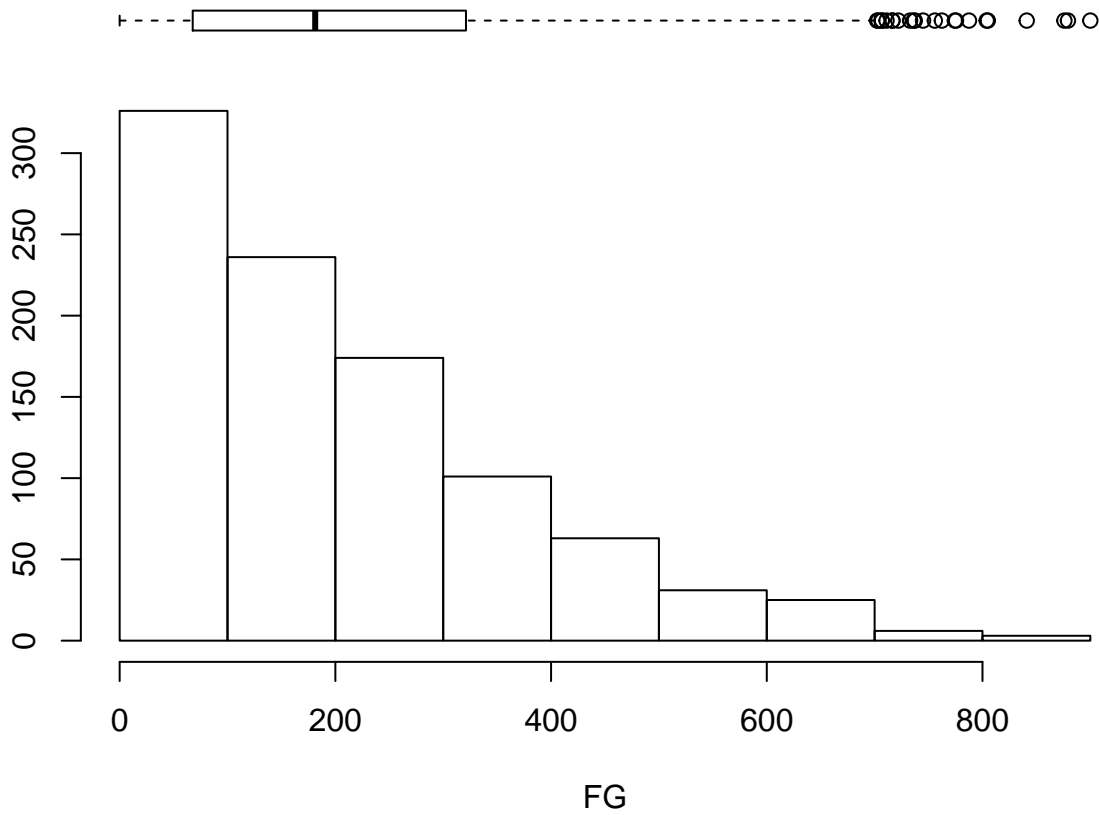
A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 2 | 2016 | Brian~ | 3.28e5 | PG | 23 | MIA | 1 | 0 | 3 | 39.3 | 1 | 0 |
| ## | 3 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 | 0.554 |
| ## | 4 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 | 0.493 |
| ## | 5 | 2016 | Russe~ | 2.65e7 | PG | 27 | OKC | 80 | 80 | 2750 | 27.6 | 0.554 | 0.236 |
| ## | 6 | 2016 | LeBro~ | 3.10e7 | SF | 31 | CLE | 76 | 76 | 2709 | 27.5 | 0.588 | 0.199 |
| ## | 7 | 2017 | Chris~ | 2.46e7 | PG | 31 | LAC | 61 | 61 | 1921 | 26.2 | 0.614 | 0.385 |
| ## | 8 | 2017 | LeBro~ | 3.33e7 | SF | 32 | CLE | 74 | 74 | 2794 | 27 | 0.619 | 0.254 |
| ## | 9 | 2016 | Kawhi~ | 1.76e7 | SF | 24 | SAS | 72 | 72 | 2380 | 26 | 0.616 | 0.267 |
| ## | 10 | 2017 | Nikol~ | 1.47e6 | C | 21 | DEN | 73 | 59 | 2038 | 26.4 | 0.64 | 0.163 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>



Histogram of FG



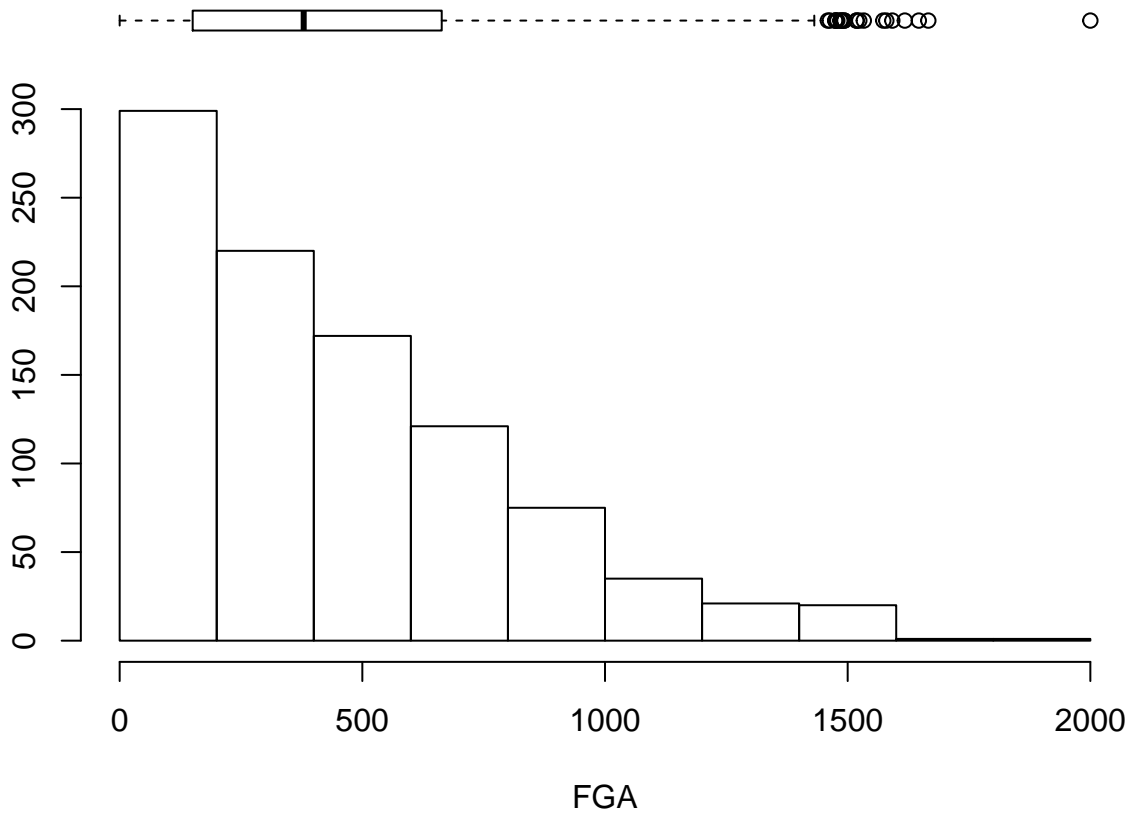
Top 10 Players by FG

A tibble: 10 x 51

| | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 2 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 | 0.554 |
| ## | 3 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 4 | 2017 | Antho~ | 2.38e7 | C | 23 | NOP | 75 | 75 | 2708 | 27.5 | 0.579 | 0.088 |
| ## | 5 | 2016 | LeBro~ | 3.10e7 | SF | 31 | CLE | 76 | 76 | 2709 | 27.5 | 0.588 | 0.199 |
| ## | 6 | 2017 | LeBro~ | 3.33e7 | SF | 32 | CLE | 74 | 74 | 2794 | 27 | 0.619 | 0.254 |
| ## | 7 | 2017 | DeMar~ | 2.77e7 | SG | 27 | TOR | 74 | 74 | 2620 | 24 | 0.552 | 0.08 |
| ## | 8 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 | 0.406 |
| ## | 9 | 2017 | Andre~ | 7.57e6 | SF | 21 | MIN | 82 | 82 | 3048 | 16.5 | 0.534 | 0.184 |
| ## | 10 | 2016 | Kevin~ | 2.65e7 | SF | 27 | OKC | 72 | 72 | 2578 | 28.2 | 0.634 | 0.348 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of FGA



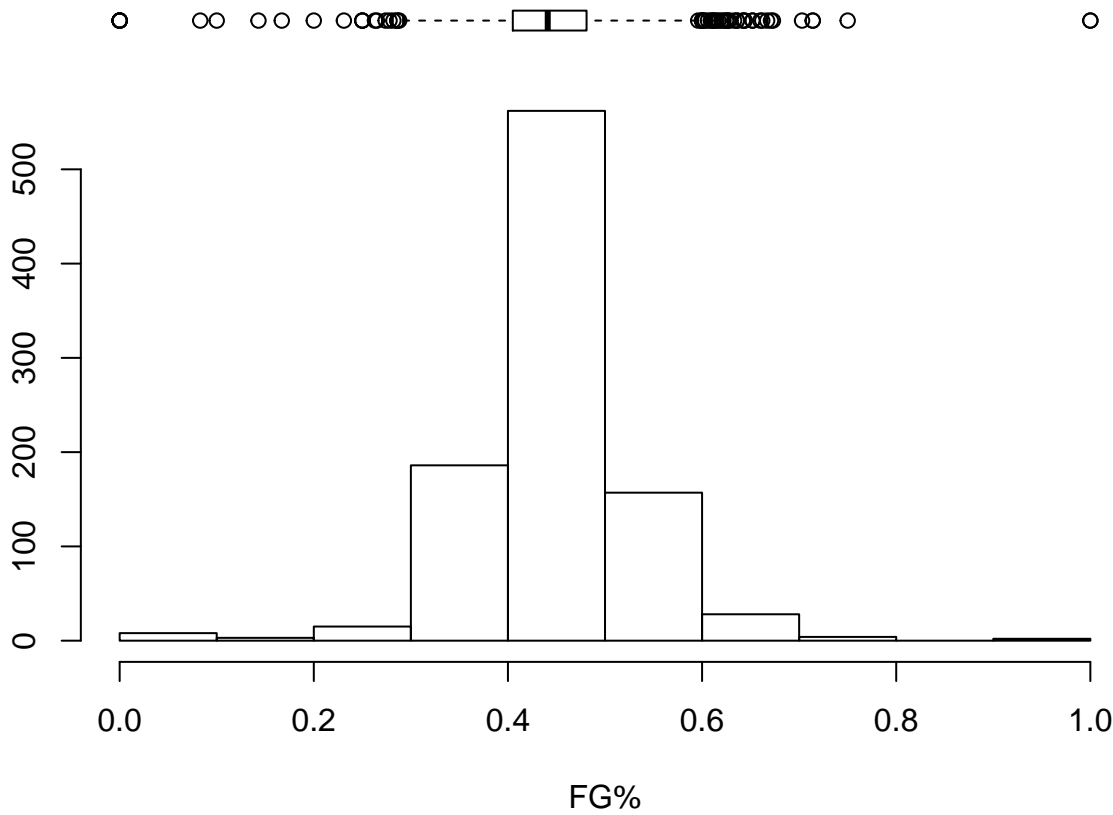
Top 10 Players by FGA

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 2 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 | 0.406 |
| ## | 3 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 | 0.554 |
| ## | 4 | 2017 | Andre~ | 7.57e6 | SF | 21 | MIN | 82 | 82 | 3048 | 16.5 | 0.534 | 0.184 |
| ## | 5 | 2017 | DeMar~ | 2.77e7 | SG | 27 | TOR | 74 | 74 | 2620 | 24 | 0.552 | 0.08 |
| ## | 6 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 | 0.493 |
| ## | 7 | 2017 | Antho~ | 2.38e7 | C | 23 | NOP | 75 | 75 | 2708 | 27.5 | 0.579 | 0.088 |
| ## | 8 | 2017 | Damia~ | 2.62e7 | PG | 26 | POR | 75 | 75 | 2694 | 24.1 | 0.586 | 0.388 |
| ## | 9 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 10 | 2016 | Damia~ | 2.43e7 | PG | 25 | POR | 75 | 75 | 2676 | 22.2 | 0.56 | 0.414 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of FG%

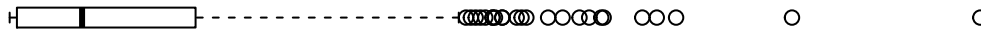


Top 10 Players by FG%

A tibble: 10 x 51

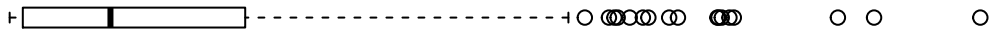
| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## | 1 | 2016 | Brian~ | 3.28e5 | PG | 23 | MIA | 1 | 0 | 3 | 39.3 | 1 |
| ## | 2 | 2016 | Rakee~ | 1.05e6 | PF | 24 | IND | 1 | 0 | 6 | 32 | 1 |
| ## | 3 | 2017 | Demet~ | 9.29e4 | PG | 22 | BOS | 5 | 0 | 17 | 30.8 | 0.753 |
| ## | 4 | 2017 | China~ | 1.31e6 | C | 20 | HOU | 5 | 1 | 52 | 12.3 | 0.799 |
| ## | 5 | 2017 | DeAnd~ | 2.26e7 | C | 28 | LAC | 81 | 81 | 2570 | 21.8 | 0.673 |
| ## | 6 | 2016 | DeAnd~ | 2.12e7 | C | 27 | LAC | 77 | 77 | 2598 | 20.6 | 0.628 |
| ## | 7 | 2016 | Brand~ | 5.70e6 | PF | 28 | MEM | 12 | 2 | 212 | 18.3 | 0.663 |
| ## | 8 | 2017 | Tyson~ | 1.30e7 | C | 34 | PHO | 47 | 46 | 1298 | 16.6 | 0.703 |
| ## | 9 | 2017 | Jarre~ | 2.33e6 | PG | 33 | NOP | 2 | 0 | 33 | 7.7 | 0.773 |
| ## | 10 | 2017 | Rudy ~ | 2.20e7 | C | 24 | UTA | 81 | 81 | 2744 | 23.3 | 0.682 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>



```
## # A tibble: 10 x 51
```

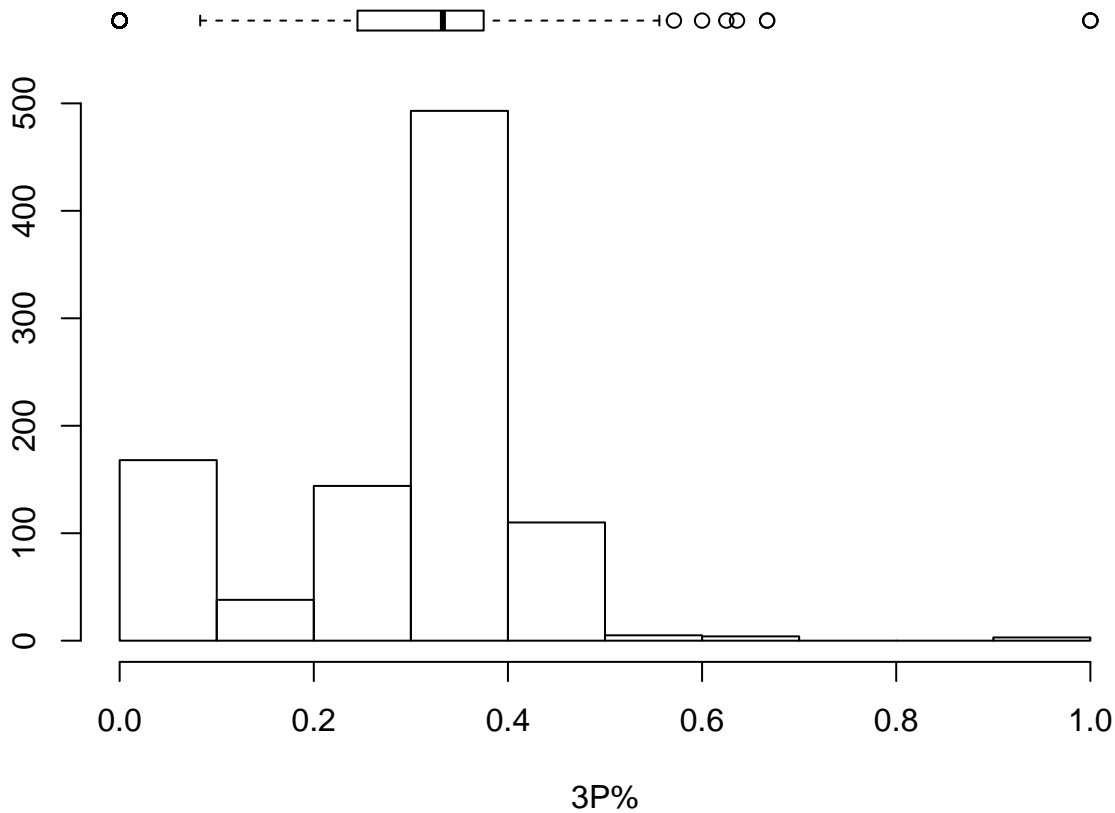
```
##      year name_p salary Pos      Age Tm      G      GS      MP      PER `TS`% `3PAr`
##      <fct> <chr>      <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Steph~ 1.21e7 PG      27 GSW      79      79 2700 31.5 0.669 0.554
## 2 2017 Steph~ 3.47e7 PG      28 GSW      79      79 2638 24.6 0.624 0.547
## 3 2016 Klay ~ 1.67e7 SG      25 GSW      80      80 2666 18.6 0.597 0.469
## 4 2017 Klay ~ 1.78e7 SG      26 GSW      78      78 2649 17.4 0.592 0.47
## 5 2017 James~ 2.83e7 PG      27 HOU      81      81 2947 27.3 0.613 0.493
## 6 2017 Eric ~ 1.29e7 SG      28 HOU      75      15 2323 13.1 0.557 0.651
## 7 2017 Isaia~ 6.26e6 PG      27 BOS      76      76 2569 26.5 0.625 0.439
## 8 2017 Kemba~ 1.20e7 PG      26 CHO      79      79 2739 21.3 0.569 0.415
## 9 2016 James~ 2.65e7 SG      26 HOU      82      82 3125 25.3 0.598 0.406
## 10 2016 Damia~ 2.43e7 PG      25 POR      75      75 2676 22.2 0.56 0.414
## # ... with 39 more variables: FTr <dbl>, `ORB`% <dbl>, `DRB`% <dbl>,
## # `TRB`% <dbl>, `AST`% <dbl>, `STL`% <dbl>, `BLK`% <dbl>, `TOV`% <dbl>,
## # `USG`% <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG`% <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P`% <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P`% <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>
```



```
## # A tibble: 10 x 51
```

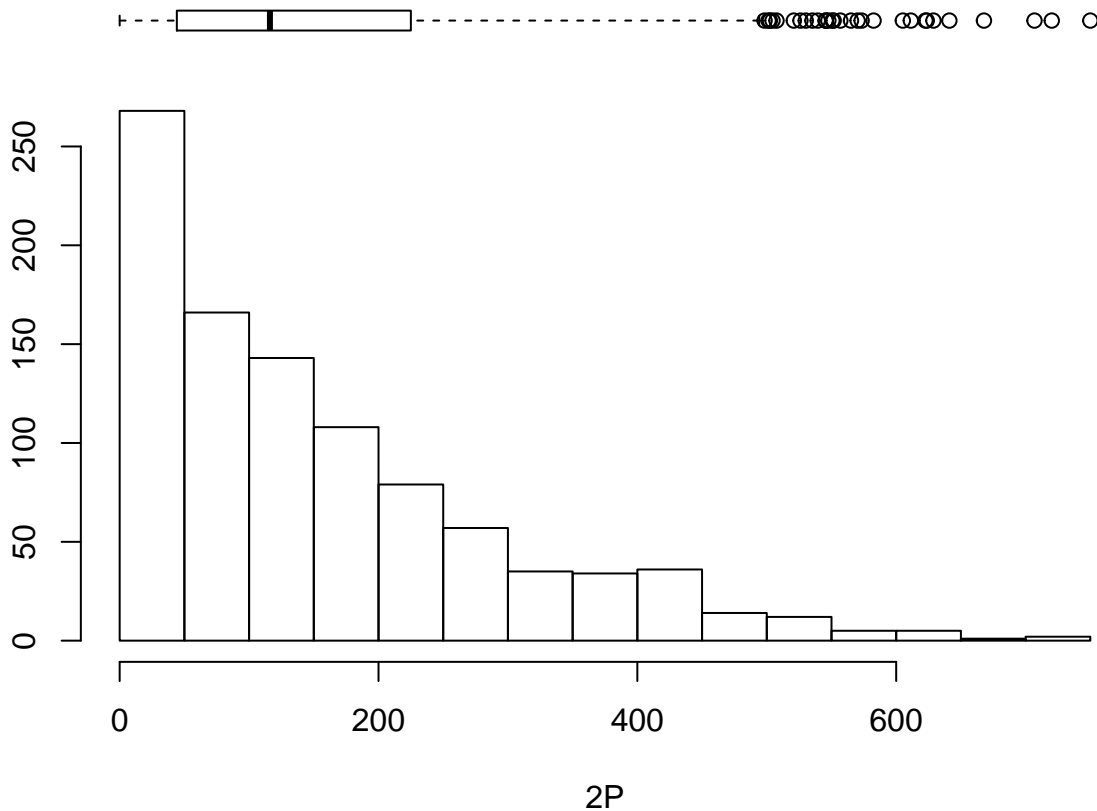
```
##      year name_p salary Pos      Age Tm      G      GS      MP      PER `TS%` `3PAR`
##      <fct> <chr>      <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 2016 Steph~ 1.21e7 PG      27 GSW      79      79 2700 31.5 0.669 0.554
##  2 2017 Steph~ 3.47e7 PG      28 GSW      79      79 2638 24.6 0.624 0.547
##  3 2017 James~ 2.83e7 PG      27 HOU      81      81 2947 27.3 0.613 0.493
##  4 2017 Eric ~ 1.29e7 SG      28 HOU      75     15 2323 13.1 0.557 0.651
##  5 2016 James~ 2.65e7 SG      26 HOU      82      82 3125 25.3 0.598 0.406
##  6 2016 Klay ~ 1.67e7 SG      25 GSW      80      80 2666 18.6 0.597 0.469
##  7 2017 Klay ~ 1.78e7 SG      26 GSW      78      78 2649 17.4 0.592 0.47
##  8 2017 Isaia~ 6.26e6 PG      27 BOS      76      76 2569 26.5 0.625 0.439
##  9 2016 Damia~ 2.43e7 PG      25 POR      75      75 2676 22.2 0.56 0.414
## 10 2017 Kemba~ 1.20e7 PG      26 CHO      79      79 2739 21.3 0.569 0.415
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

Histogram of 3P%



```
## Top 10 Players by 3P%
## # A tibble: 10 x 51
##   year name_p salary Pos Age Tm G GS MP PER `TS%` `3PAr`
##   <fct> <chr> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Andre~ 1.10e7 C 31 GSW 70 66 1451 15.9 0.623 0.004
## 2 2017 Demet~ 9.29e4 PG 22 BOS 5 0 17 30.8 0.753 0.25
## 3 2017 Taj G~ 1.40e7 PF 31 OKC 23 16 487 13.8 0.528 0.006
## 4 2016 Josh ~ 1.19e6 PF 24 OKC 5 0 55 6.7 0.509 0.5
## 5 2016 Marc ~ 2.12e7 C 31 MEM 52 52 1791 17.7 0.528 0.004
## 6 2016 Jorda~ 8.75e5 PG 24 CLE 15 1 113 14.2 0.537 0.212
## 7 2017 Lance~ 4.18e6 SG 26 IND 6 0 132 10.3 0.474 0.182
## 8 2017 Treve~ 1.31e6 SG 23 CHO 27 1 189 10.6 0.612 0.375
## 9 2017 Wayne~ 1.31e6 SG 22 NOP 3 3 47 10 0.82 0.875
## 10 2016 Steve~ 1.55e6 PF 32 OKC 7 0 24 20.8 0.708 0.75
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

Histogram of 2P



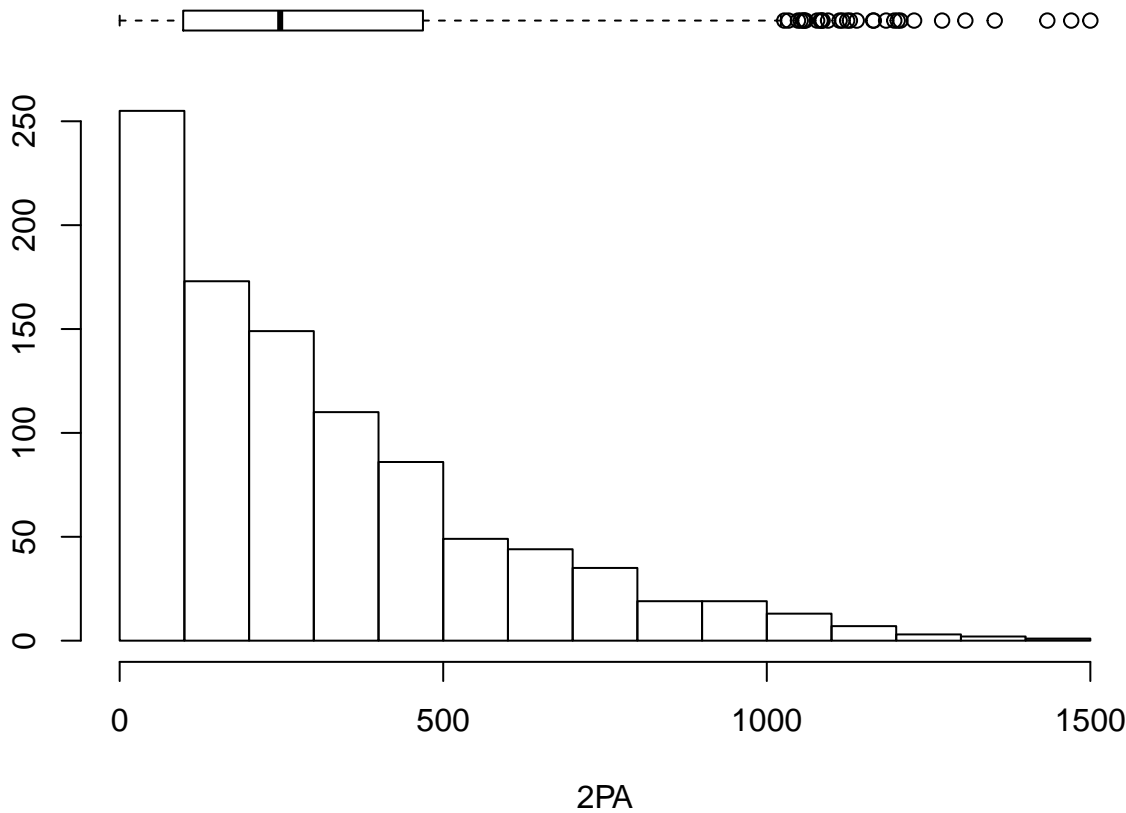
Top 10 Players by 2P

A tibble: 10 x 51

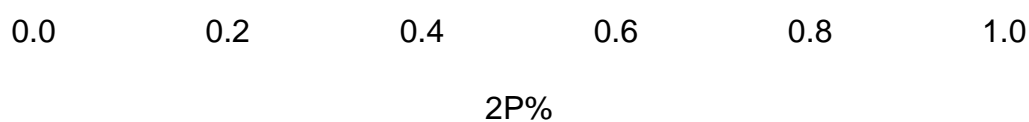
| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PA` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | Antho~ | 2.38e7 | C | 23 | NOP | 75 | 75 | 2708 | 27.5 | 0.579 | 0.088 |
| ## | 2 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 3 | 2017 | DeMar~ | 2.77e7 | SG | 27 | TOR | 74 | 74 | 2620 | 24 | 0.552 | 0.08 |
| ## | 4 | 2016 | LeBro~ | 3.10e7 | SF | 31 | CLE | 76 | 76 | 2709 | 27.5 | 0.588 | 0.199 |
| ## | 5 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 6 | 2017 | LeBro~ | 3.33e7 | SF | 32 | CLE | 74 | 74 | 2794 | 27 | 0.619 | 0.254 |
| ## | 7 | 2017 | Giann~ | 2.25e7 | SF | 22 | MIL | 80 | 80 | 2845 | 26.1 | 0.599 | 0.143 |
| ## | 8 | 2017 | Andre~ | 7.57e6 | SF | 21 | MIN | 82 | 82 | 3048 | 16.5 | 0.534 | 0.184 |
| ## | 9 | 2016 | Karl~ | 5.96e6 | C | 20 | MIN | 82 | 82 | 2627 | 22.5 | 0.59 | 0.076 |
| ## | 10 | 2016 | Brook~ | 2.12e7 | C | 27 | BRK | 73 | 73 | 2457 | 21.7 | 0.562 | 0.012 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of 2PA



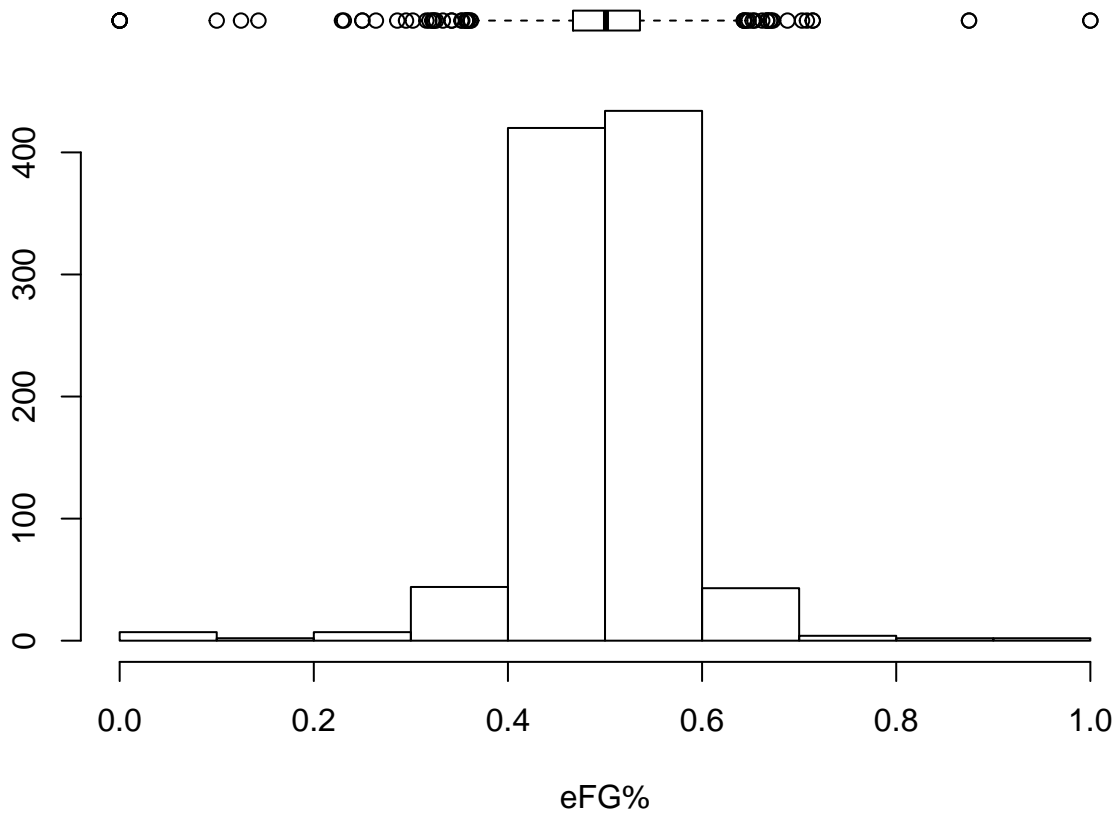
```
## Top 10 Players by 2PA
## # A tibble: 10 x 51
##   year name_p salary Pos Age Tm G GS MP PER `TS%` `3PAr`
##   <fct> <chr> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 DeMar~ 2.77e7 SG 27 TOR 74 74 2620 24 0.552 0.08
## 2 2017 Antho~ 2.38e7 C 23 NOP 75 75 2708 27.5 0.579 0.088
## 3 2017 Russe~ 2.85e7 PG 28 OKC 81 81 2802 30.6 0.554 0.3
## 4 2017 Andre~ 7.57e6 SF 21 MIN 82 82 3048 16.5 0.534 0.184
## 5 2016 DeMar~ 2.65e7 SG 26 TOR 78 78 2804 21.5 0.55 0.101
## 6 2017 Karl~ 6.22e6 C 21 MIN 82 82 3030 26 0.618 0.186
## 7 2017 John ~ 1.81e7 PG 26 WAS 78 78 2836 23.2 0.541 0.19
## 8 2016 Brook~ 2.12e7 C 27 BRK 73 73 2457 21.7 0.562 0.012
## 9 2016 Dwyan~ 2.32e7 SG 34 MIA 74 73 2258 20.3 0.517 0.037
## 10 2016 LeBro~ 3.10e7 SF 31 CLE 76 76 2709 27.5 0.588 0.199
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```



```
## # A tibble: 10 x 51
```

```
##      year name_p salary Pos      Age Tm      G      GS      MP      PER `TS`% `3PAr`
##      <fct> <chr>      <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 2016 Brian~ 3.28e5 PG      23 MIA      1      0      3 39.3 1      0
##  2 2017 Jarre~ 2.33e6 PG      33 NOP      2      0     33  7.7 0.773 0.333
##  3 2016 Rakee~ 1.05e6 PF      24 IND      1      0      6 32    1      0
##  4 2016 Sean ~ 9.80e5 SG      26 DEN      8      0     82  8    0.551 0.81
##  5 2017 Wayne~ 1.31e6 SG      22 NOP      3      3     47 10    0.82 0.875
##  6 2017 Axel ~ 2.50e4 SF      24 TOT      4      0     47  6.2 0.611 0.444
##  7 2017 Axel ~ 2.50e4 SF      24 NOP      2      0     41  8.6 0.688 0.375
##  8 2017 Ersan~ 6.00e6 PF      29 OKC      3      0     62  6.9 0.469 0.75
##  9 2017 DeAnd~ 2.26e7 C       28 LAC     81     81 2570 21.8 0.673 0.003
## 10 2017 China~ 1.31e6 C       20 HOU      5      1     52 12.3 0.799 0
## # ... with 39 more variables: FTr <dbl>, `ORB`% <dbl>, `DRB`% <dbl>,
## # `TRB`% <dbl>, `AST`% <dbl>, `STL`% <dbl>, `BLK`% <dbl>, `TOV`% <dbl>,
## # `USG`% <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG`% <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P`% <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P`% <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>
```

Histogram of eFG%



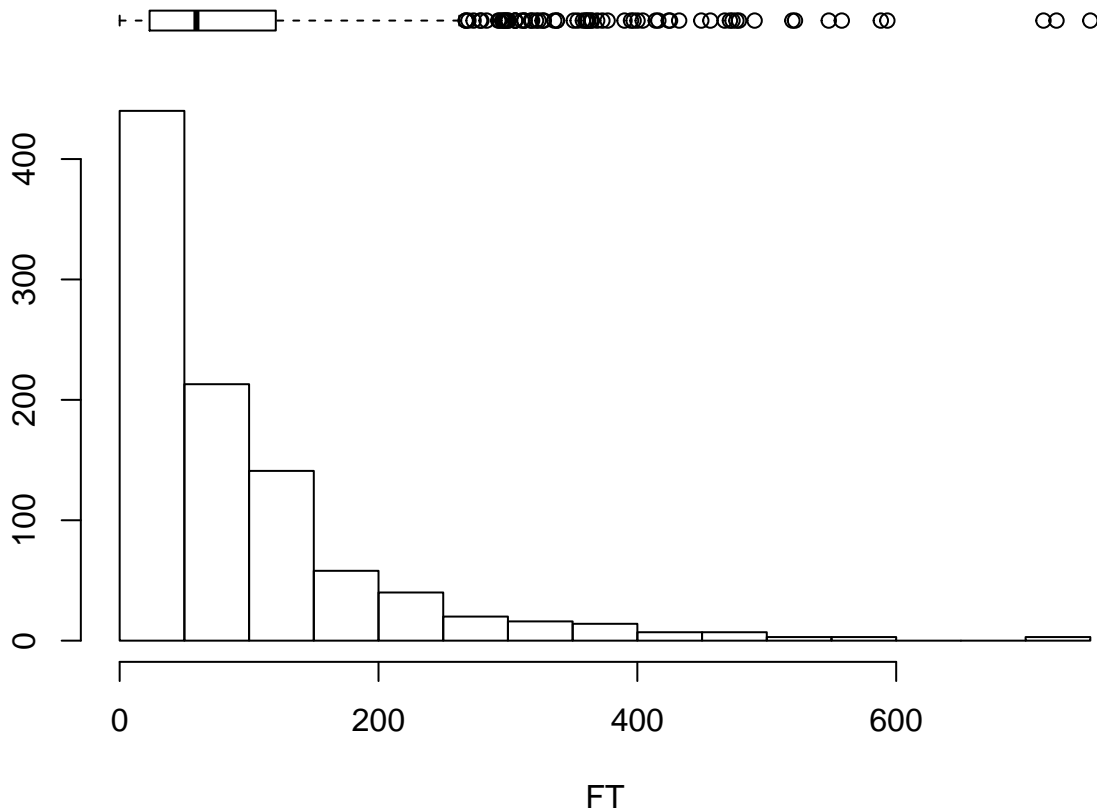
Top 10 Players by eFG%

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | TS% | 3Par |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## | 1 | 2016 | Brian~ | 3.28e5 | PG | 23 | MIA | 1 | 0 | 3 | 39.3 | 1 |
| ## | 2 | 2016 | Rakee~ | 1.05e6 | PF | 24 | IND | 1 | 0 | 6 | 32 | 1 |
| ## | 3 | 2017 | Demet~ | 9.29e4 | PG | 22 | BOS | 5 | 0 | 17 | 30.8 | 0.753 |
| ## | 4 | 2017 | Wayne~ | 1.31e6 | SG | 22 | NOP | 3 | 3 | 47 | 10 | 0.82 |
| ## | 5 | 2017 | China~ | 1.31e6 | C | 20 | HOU | 5 | 1 | 52 | 12.3 | 0.799 |
| ## | 6 | 2017 | DeAnd~ | 2.26e7 | C | 28 | LAC | 81 | 81 | 2570 | 21.8 | 0.673 |
| ## | 7 | 2016 | Steve~ | 1.55e6 | PF | 32 | OKC | 7 | 0 | 24 | 20.8 | 0.708 |
| ## | 8 | 2016 | DeAnd~ | 2.12e7 | C | 27 | LAC | 77 | 77 | 2598 | 20.6 | 0.628 |
| ## | 9 | 2017 | Axel ~ | 2.50e4 | SF | 24 | NOP | 2 | 0 | 41 | 8.6 | 0.688 |
| ## | 10 | 2016 | Brand~ | 5.70e6 | PF | 28 | MEM | 12 | 2 | 212 | 18.3 | 0.663 |

... with 39 more variables: FTr <dbl>, ORB% <dbl>, DRB% <dbl>, TRB% <dbl>, AST% <dbl>, STL% <dbl>, BLK% <dbl>, TOV% <dbl>, USG% <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, WS/48 <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, FG% <dbl>, 3P <dbl>, 3PA <dbl>, 3P% <dbl>, 2P <dbl>, 2PA <dbl>, 2P% <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>

Histogram of FT



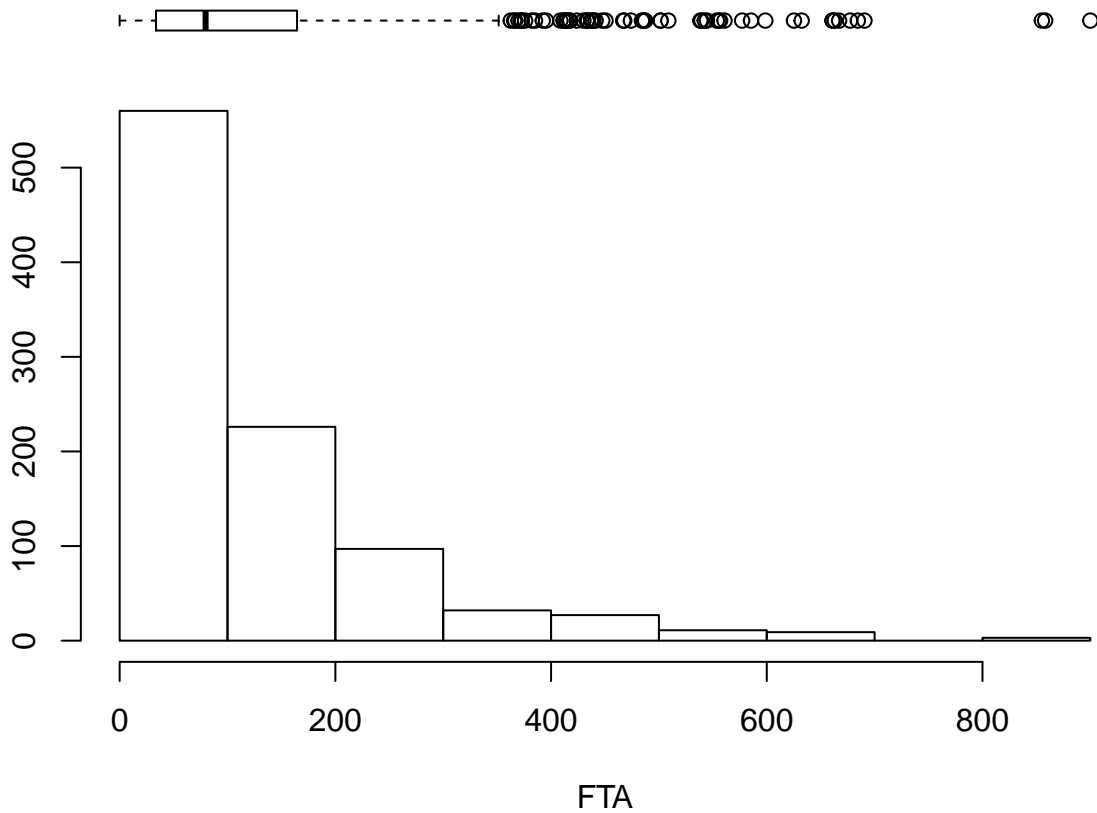
Top 10 Players by FT

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 | 0.493 |
| ## | 2 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 | 0.406 |
| ## | 3 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 4 | 2017 | Isaia~ | 6.26e6 | PG | 27 | BOS | 76 | 76 | 2569 | 26.5 | 0.625 | 0.439 |
| ## | 5 | 2017 | Jimmy~ | 1.93e7 | SF | 27 | CHI | 76 | 75 | 2809 | 25.1 | 0.586 | 0.198 |
| ## | 6 | 2016 | DeMar~ | 2.65e7 | SG | 26 | TOR | 78 | 78 | 2804 | 21.5 | 0.55 | 0.101 |
| ## | 7 | 2017 | DeMar~ | 2.77e7 | SG | 27 | TOR | 74 | 74 | 2620 | 24 | 0.552 | 0.08 |
| ## | 8 | 2017 | Antho~ | 2.38e7 | C | 23 | NOP | 75 | 75 | 2708 | 27.5 | 0.579 | 0.088 |
| ## | 9 | 2017 | DeMar~ | 1.81e7 | C | 26 | TOT | 72 | 72 | 2465 | 25.7 | 0.562 | 0.254 |
| ## | 10 | 2017 | Damia~ | 2.62e7 | PG | 26 | POR | 75 | 75 | 2694 | 24.1 | 0.586 | 0.388 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of FTA



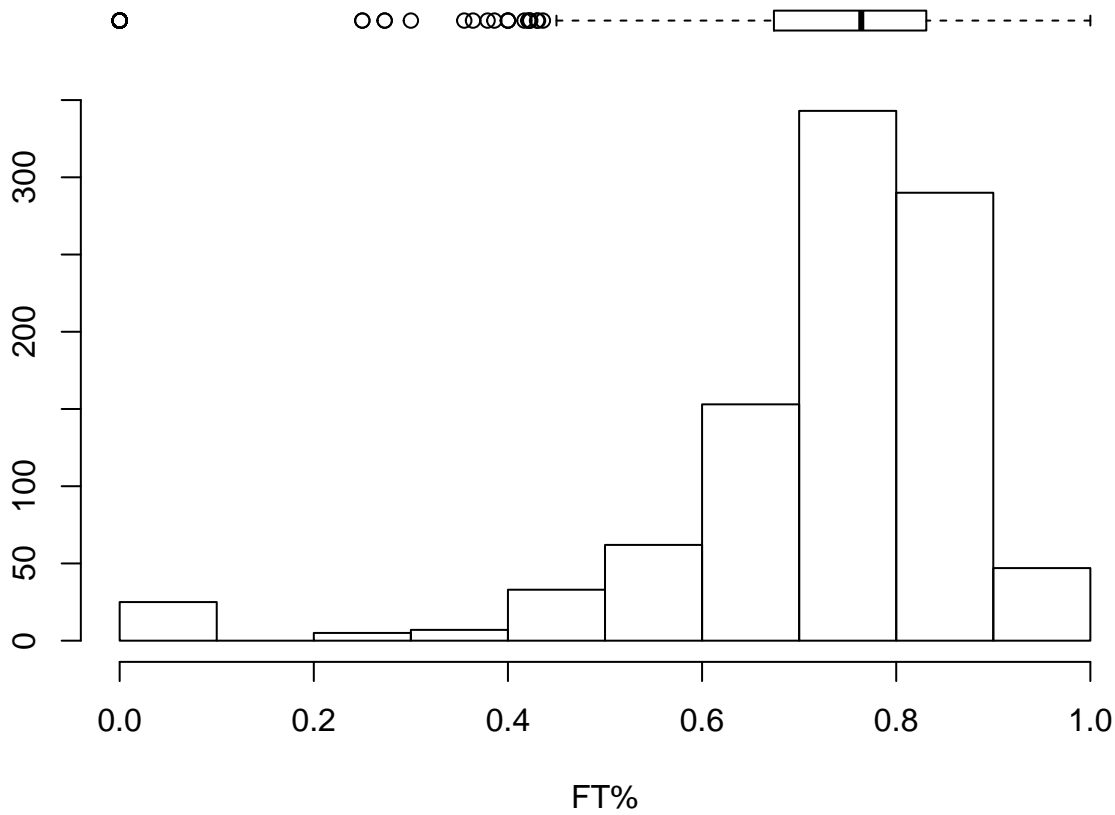
Top 10 Players by FTA

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 | 0.493 |
| ## | 2 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 3 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 | 0.406 |
| ## | 4 | 2017 | Jimmy~ | 1.93e7 | SF | 27 | CHI | 76 | 75 | 2809 | 25.1 | 0.586 | 0.198 |
| ## | 5 | 2017 | DeMar~ | 1.81e7 | C | 26 | TOT | 72 | 72 | 2465 | 25.7 | 0.562 | 0.254 |
| ## | 6 | 2016 | DeMar~ | 1.70e7 | C | 25 | SAC | 65 | 65 | 2246 | 23.6 | 0.538 | 0.158 |
| ## | 7 | 2016 | DeMar~ | 2.65e7 | SG | 26 | TOR | 78 | 78 | 2804 | 21.5 | 0.55 | 0.101 |
| ## | 8 | 2017 | Isaia~ | 6.26e6 | PG | 27 | BOS | 76 | 76 | 2569 | 26.5 | 0.625 | 0.439 |
| ## | 9 | 2017 | Antho~ | 2.38e7 | C | 23 | NOP | 75 | 75 | 2708 | 27.5 | 0.579 | 0.088 |
| ## | 10 | 2017 | DeMar~ | 2.77e7 | SG | 27 | TOR | 74 | 74 | 2620 | 24 | 0.552 | 0.08 |

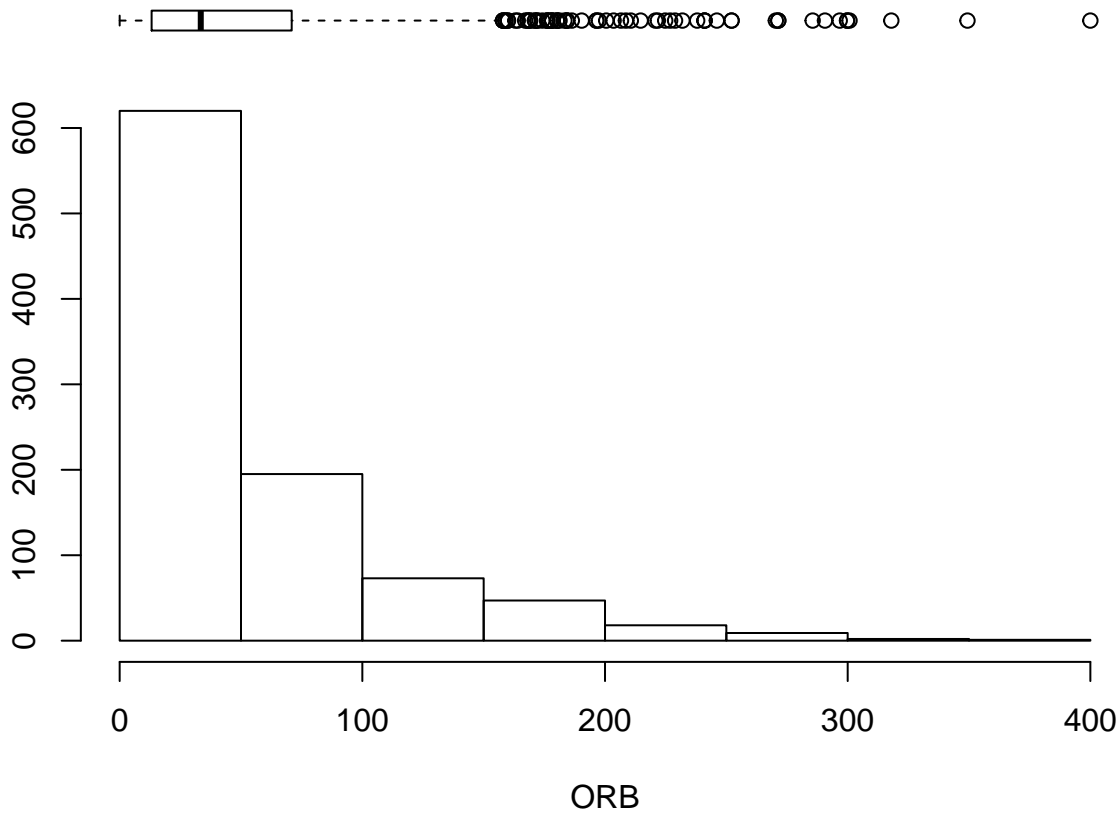
... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of FT%



```
## Top 10 Players by FT%
## # A tibble: 10 x 51
##   year name_p salary Pos Age Tm G GS MP PER `TS%` `3PAr`
##   <dbl> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Andre~ 2.84e6 PF 27 BRK 10 0 111 5 0.43 0.324
## 2 2016 Beno ~ 1.55e6 PG 33 MEM 8 0 120 12.6 0.497 0.239
## 3 2017 Bobby~ 1.52e6 PG 32 HOU 25 0 123 10.8 0.509 0.583
## 4 2017 Camer~ 2.20e6 PG 22 OKC 20 0 320 6.2 0.402 0.4
## 5 2017 Chass~ 1.31e6 PG 23 PHI 8 0 74 17.7 0.671 0.577
## 6 2017 China~ 1.31e6 C 20 HOU 5 1 52 12.3 0.799 0
## 7 2016 Damja~ 9.80e5 SF 29 MIN 33 0 277 5.5 0.572 0.806
## 8 2017 Diamo~ 1.31e6 C 19 LAC 7 0 24 -1.2 0.339 0
## 9 2017 Georg~ 1.00e5 PF 23 IND 23 0 93 0.1 0.285 0.333
## 10 2016 Jarel~ 1.75e5 SF 24 WAS 26 0 147 11 0.46 0.723
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

Histogram of ORB



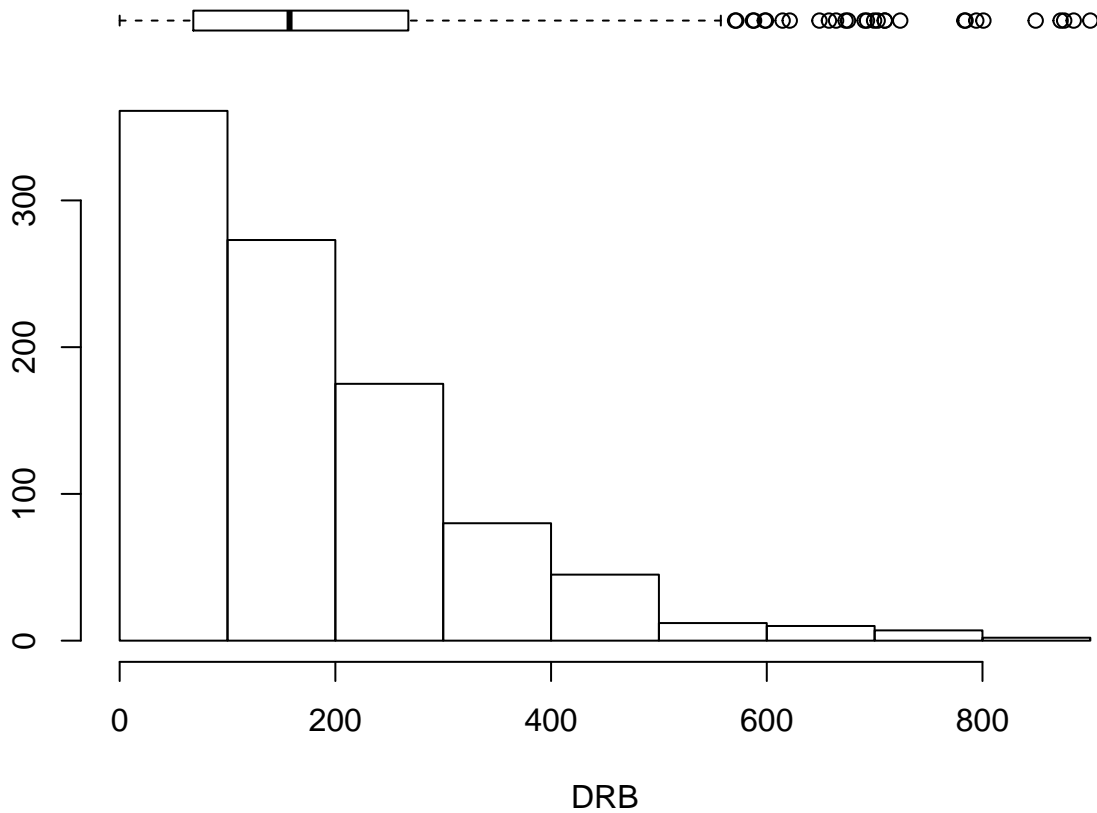
Top 10 Players by ORB

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | TS% | 3PAr | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Andre~ | 2.21e7 | C | 22 | DET | 81 | 81 | 2666 | 21.2 | 0.499 | 0.006 |
| ## | 2 | 2017 | Andre~ | 2.38e7 | C | 23 | DET | 81 | 81 | 2409 | 20.9 | 0.518 | 0.008 |
| ## | 3 | 2017 | Rudy ~ | 2.20e7 | C | 24 | UTA | 81 | 81 | 2744 | 23.3 | 0.682 | 0.002 |
| ## | 4 | 2017 | DeAnd~ | 2.26e7 | C | 28 | LAC | 81 | 81 | 2570 | 21.8 | 0.673 | 0.003 |
| ## | 5 | 2017 | Dwigh~ | 2.35e7 | C | 31 | ATL | 74 | 74 | 2199 | 20.8 | 0.627 | 0.003 |
| ## | 6 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 7 | 2017 | Hassa~ | 2.38e7 | C | 27 | MIA | 77 | 77 | 2513 | 22.6 | 0.579 | 0 |
| ## | 8 | 2017 | Trist~ | 1.64e7 | C | 25 | CLE | 78 | 78 | 2336 | 15.3 | 0.594 | 0.007 |
| ## | 9 | 2017 | Steve~ | 2.25e7 | C | 23 | OKC | 80 | 80 | 2389 | 16.5 | 0.589 | 0.002 |
| ## | 10 | 2016 | Robin~ | 1.32e7 | C | 27 | NYK | 82 | 82 | 2219 | 17.6 | 0.574 | 0.002 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of DRB



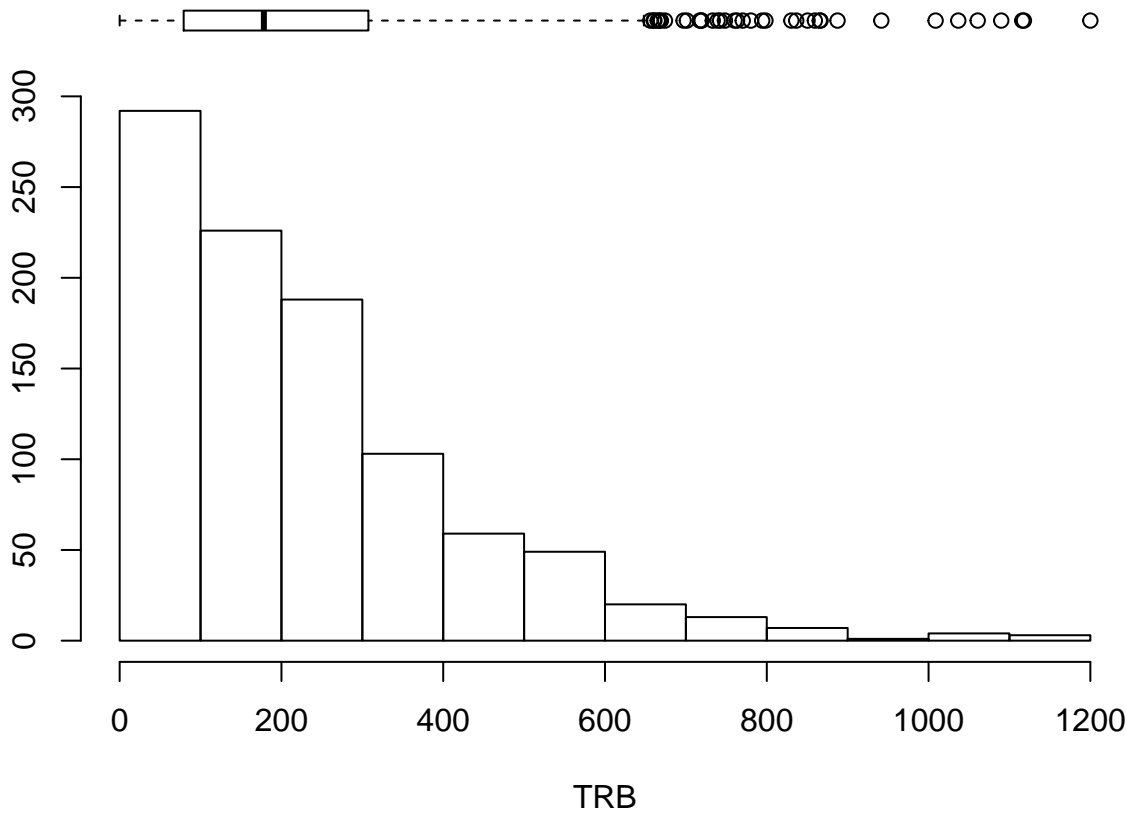
Top 10 Players by DRB

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3Par` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | DeAnd~ | 2.26e7 | C | 28 | LAC | 81 | 81 | 2570 | 21.8 | 0.673 | 0.003 |
| ## | 2 | 2016 | Andre~ | 2.21e7 | C | 22 | DET | 81 | 81 | 2666 | 21.2 | 0.499 | 0.006 |
| ## | 3 | 2017 | Hassa~ | 2.38e7 | C | 27 | MIA | 77 | 77 | 2513 | 22.6 | 0.579 | 0 |
| ## | 4 | 2016 | DeAnd~ | 2.12e7 | C | 27 | LAC | 77 | 77 | 2598 | 20.6 | 0.628 | 0.002 |
| ## | 5 | 2017 | Andre~ | 2.38e7 | C | 23 | DET | 81 | 81 | 2409 | 20.9 | 0.518 | 0.008 |
| ## | 6 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 | 0.3 |
| ## | 7 | 2017 | Rudy ~ | 2.20e7 | C | 24 | UTA | 81 | 81 | 2744 | 23.3 | 0.682 | 0.002 |
| ## | 8 | 2017 | Antho~ | 2.38e7 | C | 23 | NOP | 75 | 75 | 2708 | 27.5 | 0.579 | 0.088 |
| ## | 9 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 10 | 2016 | Juliu~ | 3.27e6 | PF | 21 | LAL | 81 | 60 | 2286 | 13.9 | 0.482 | 0.043 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of TRB

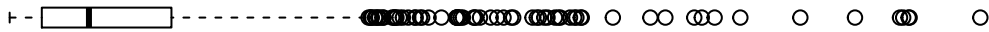


Top 10 Players by TRB

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Andre~ | 2.21e7 | C | 22 | DET | 81 | 81 | 2666 | 21.2 | 0.499 | 0.006 |
| ## | 2 | 2017 | Andre~ | 2.38e7 | C | 23 | DET | 81 | 81 | 2409 | 20.9 | 0.518 | 0.008 |
| ## | 3 | 2017 | DeAnd~ | 2.26e7 | C | 28 | LAC | 81 | 81 | 2570 | 21.8 | 0.673 | 0.003 |
| ## | 4 | 2017 | Hassa~ | 2.38e7 | C | 27 | MIA | 77 | 77 | 2513 | 22.6 | 0.579 | 0 |
| ## | 5 | 2016 | DeAnd~ | 2.12e7 | C | 27 | LAC | 77 | 77 | 2598 | 20.6 | 0.628 | 0.002 |
| ## | 6 | 2017 | Rudy ~ | 2.20e7 | C | 24 | UTA | 81 | 81 | 2744 | 23.3 | 0.682 | 0.002 |
| ## | 7 | 2017 | Karl~ | 6.22e6 | C | 21 | MIN | 82 | 82 | 3030 | 26 | 0.618 | 0.186 |
| ## | 8 | 2017 | Dwigh~ | 2.35e7 | C | 31 | ATL | 74 | 74 | 2199 | 20.8 | 0.627 | 0.003 |
| ## | 9 | 2017 | Antho~ | 2.38e7 | C | 23 | NOP | 75 | 75 | 2708 | 27.5 | 0.579 | 0.088 |
| ## | 10 | 2016 | Hassa~ | 2.21e7 | C | 26 | MIA | 73 | 43 | 2125 | 25.7 | 0.629 | 0 |

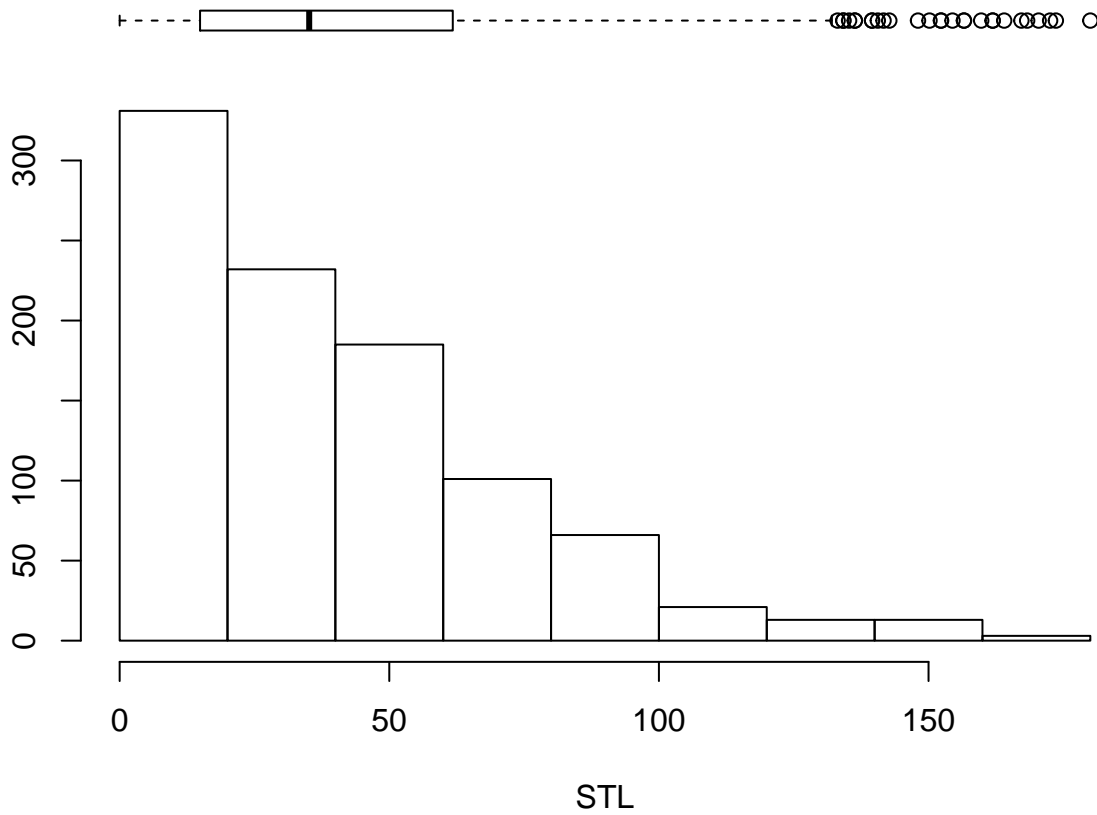
... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
 ## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
 ## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
 ## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
 ## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>



```
## # A tibble: 10 x 51
```

[illegible]

Histogram of STL



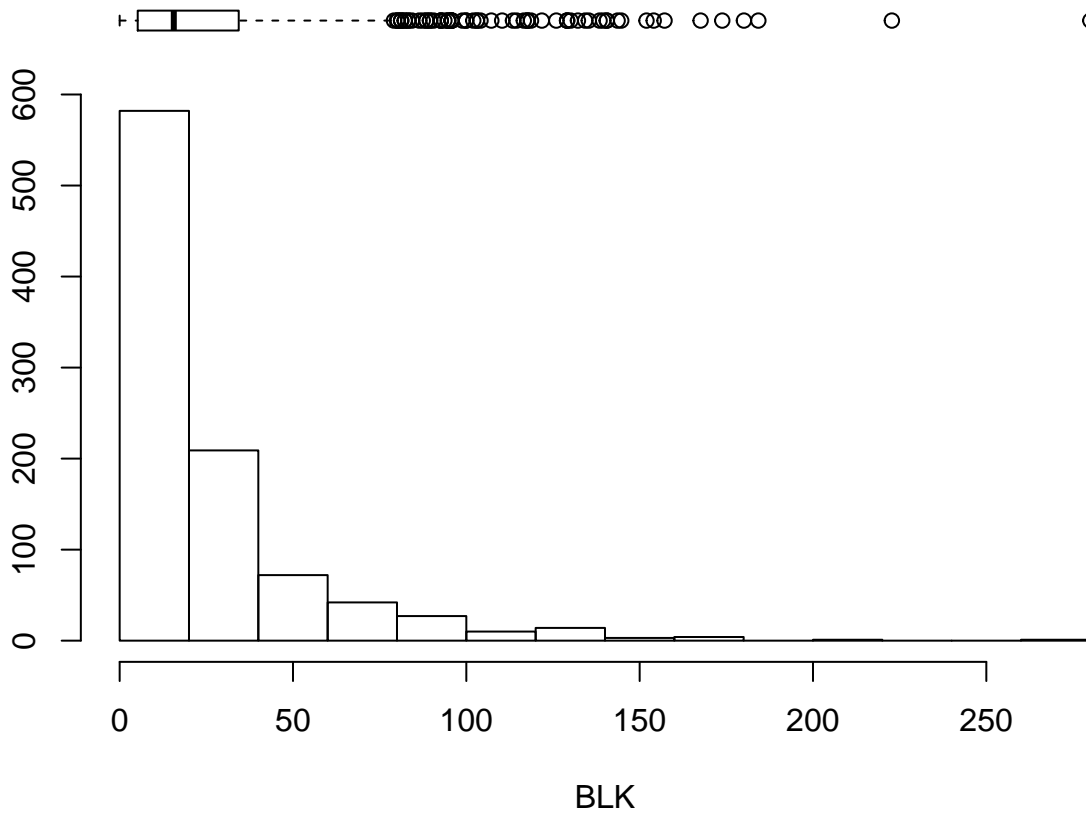
Top 10 Players by STL

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2016 | Steph~ | 1.21e7 | PG | 27 | GSW | 79 | 79 | 2700 | 31.5 | 0.669 | 0.554 |
| ## | 2 | 2016 | Russe~ | 2.65e7 | PG | 27 | OKC | 80 | 80 | 2750 | 27.6 | 0.554 | 0.236 |
| ## | 3 | 2016 | Ricky~ | 1.36e7 | PG | 25 | MIN | 76 | 76 | 2323 | 17.6 | 0.529 | 0.324 |
| ## | 4 | 2016 | Trevo~ | 7.81e6 | SF | 30 | HOU | 81 | 81 | 2859 | 12.9 | 0.551 | 0.581 |
| ## | 5 | 2016 | Kyle ~ | 1.20e7 | PG | 29 | TOR | 77 | 77 | 2851 | 22.2 | 0.578 | 0.457 |
| ## | 6 | 2017 | John ~ | 1.81e7 | PG | 26 | WAS | 78 | 78 | 2836 | 23.2 | 0.541 | 0.19 |
| ## | 7 | 2017 | Draym~ | 1.64e7 | PF | 26 | GSW | 76 | 76 | 2471 | 16.5 | 0.522 | 0.405 |
| ## | 8 | 2016 | Chris~ | 2.29e7 | PG | 30 | LAC | 74 | 74 | 2420 | 26.2 | 0.575 | 0.295 |
| ## | 9 | 2016 | Paul ~ | 1.83e7 | SF | 25 | IND | 81 | 81 | 2819 | 20.9 | 0.557 | 0.391 |
| ## | 10 | 2016 | Monta~ | 1.08e7 | SG | 30 | IND | 81 | 81 | 2734 | 13.7 | 0.504 | 0.276 |

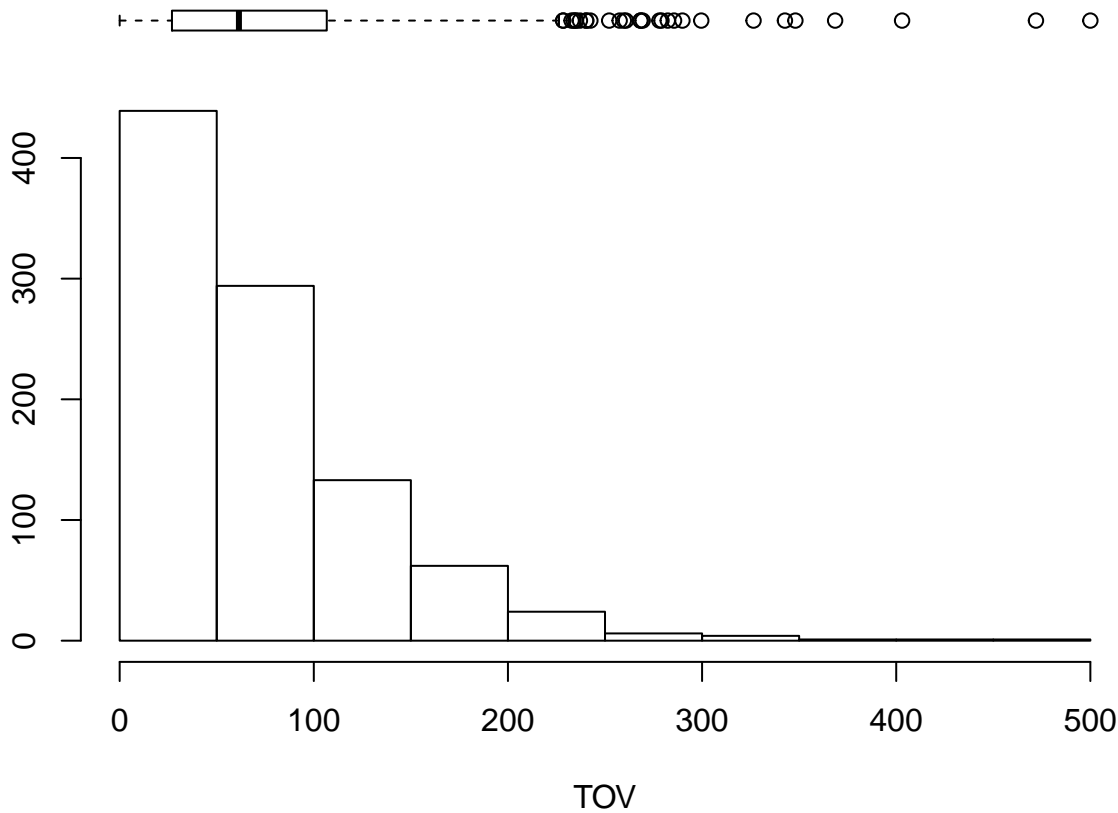
... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
`TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
`USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
`3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of BLK



```
## Top 10 Players by BLK
## # A tibble: 10 x 51
##   year name_p salary Pos Age Tm G GS MP PER `TS%` `3PAR`
##   <fct> <chr> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Hassa~ 2.21e7 C 26 MIA 73 43 2125 25.7 0.629 0
## 2 2017 Rudy ~ 2.20e7 C 24 UTA 81 81 2744 23.3 0.682 0.002
## 3 2016 DeAnd~ 2.12e7 C 27 LAC 77 77 2598 20.6 0.628 0.002
## 4 2017 Myles~ 2.57e6 C 20 IND 81 81 2541 18.5 0.585 0.132
## 5 2017 Antho~ 2.38e7 C 23 NOP 75 75 2708 27.5 0.579 0.088
## 6 2017 Hassa~ 2.38e7 C 27 MIA 77 77 2513 22.6 0.579 0
## 7 2017 Giann~ 2.25e7 SF 22 MIL 80 80 2845 26.1 0.599 0.143
## 8 2016 Serge~ 1.23e7 PF 26 OKC 78 78 2500 13.9 0.533 0.212
## 9 2016 Pau G~ 1.55e7 C 35 CHI 72 72 2291 21.7 0.529 0.069
## 10 2016 Paul ~ 2.01e7 PF 30 ATL 81 81 2647 21.3 0.556 0.218
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```


Histogram of TOV



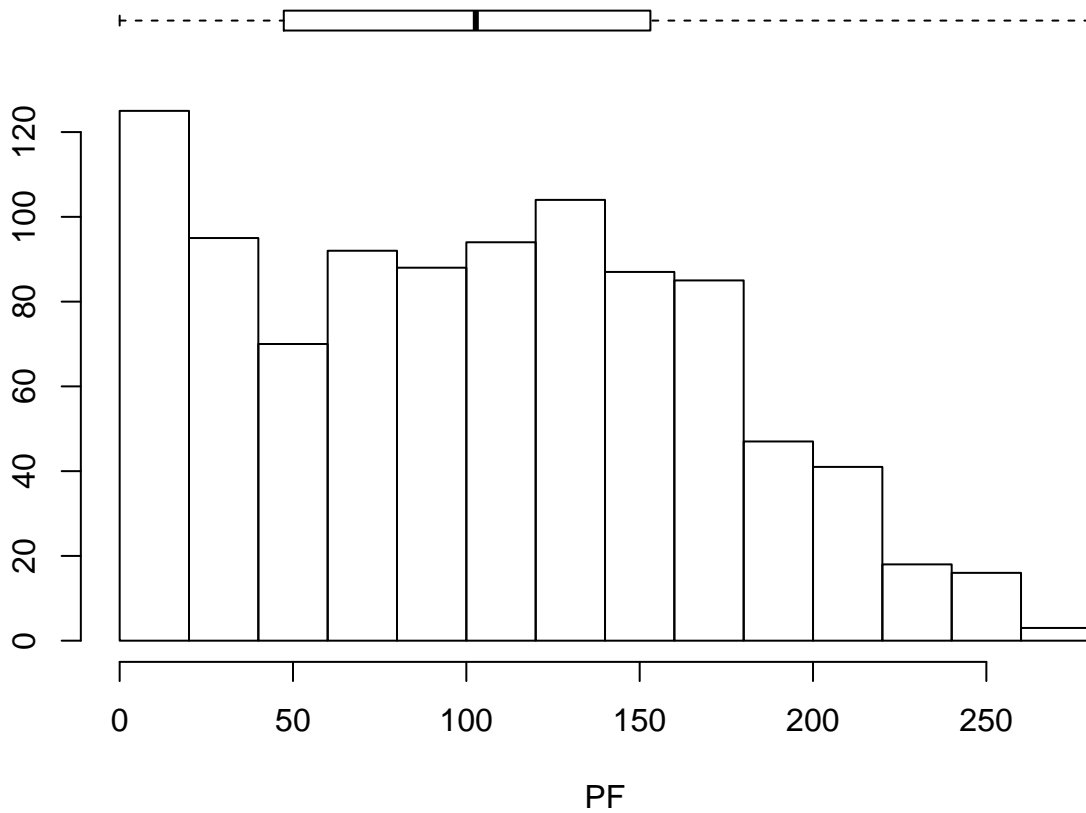
Top 10 Players by TOV

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| ## | 1 | 2017 | James~ | 2.83e7 | PG | 27 | HOU | 81 | 81 | 2947 | 27.3 | 0.613 |
| ## | 2 | 2017 | Russe~ | 2.85e7 | PG | 28 | OKC | 81 | 81 | 2802 | 30.6 | 0.554 |
| ## | 3 | 2016 | James~ | 2.65e7 | SG | 26 | HOU | 82 | 82 | 3125 | 25.3 | 0.598 |
| ## | 4 | 2016 | Russe~ | 2.65e7 | PG | 27 | OKC | 80 | 80 | 2750 | 27.6 | 0.554 |
| ## | 5 | 2017 | John ~ | 1.81e7 | PG | 26 | WAS | 78 | 78 | 2836 | 23.2 | 0.541 |
| ## | 6 | 2016 | John ~ | 1.70e7 | PG | 25 | WAS | 77 | 77 | 2784 | 19.8 | 0.51 |
| ## | 7 | 2017 | LeBro~ | 3.33e7 | SF | 32 | CLE | 74 | 74 | 2794 | 27 | 0.619 |
| ## | 8 | 2016 | Rajon~ | 1.40e7 | PG | 29 | SAC | 72 | 72 | 2537 | 16.9 | 0.506 |
| ## | 9 | 2017 | DeMar~ | 1.81e7 | C | 26 | TOT | 72 | 72 | 2465 | 25.7 | 0.562 |
| ## | 10 | 2016 | Paul ~ | 1.83e7 | SF | 25 | IND | 81 | 81 | 2819 | 20.9 | 0.557 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of PF



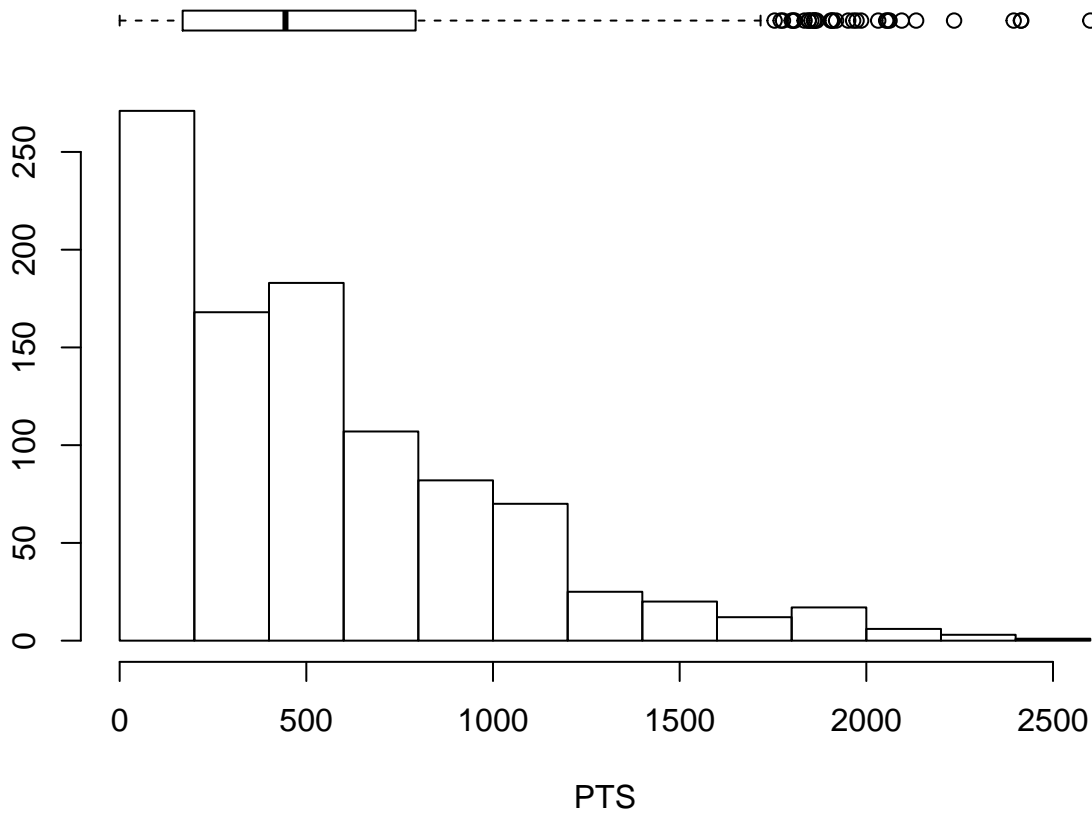
Top 10 Players by PF

A tibble: 10 x 51

| ## | year | name_p | salary | Pos | Age | Tm | G | GS | MP | PER | `TS%` | `3PAr` | |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
| ## | <fct> | <chr> | <dbl> | <fct> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | |
| ## | 1 | 2017 | DeMar~ | 1.81e7 | C | 26 | TOT | 72 | 72 | 2465 | 25.7 | 0.562 | 0.254 |
| ## | 2 | 2017 | Marqu~ | 3.07e6 | PF | 19 | PHO | 82 | 75 | 1743 | 12.3 | 0.529 | 0.354 |
| ## | 3 | 2017 | Myles~ | 2.57e6 | C | 20 | IND | 81 | 81 | 2541 | 18.5 | 0.585 | 0.132 |
| ## | 4 | 2016 | Giann~ | 3.00e6 | PG | 21 | MIL | 80 | 79 | 2823 | 18.8 | 0.566 | 0.108 |
| ## | 5 | 2017 | Gorgu~ | 1.41e7 | PF | 27 | MIN | 82 | 82 | 2653 | 14.2 | 0.555 | 0.065 |
| ## | 6 | 2017 | Marki~ | 8.00e6 | PF | 27 | WAS | 76 | 76 | 2374 | 13.7 | 0.54 | 0.22 |
| ## | 7 | 2016 | Mason~ | 2.33e6 | C | 25 | POR | 82 | 82 | 2084 | 17.2 | 0.564 | 0.008 |
| ## | 8 | 2016 | Roy H~ | 5.00e6 | C | 29 | LAL | 81 | 81 | 1878 | 11.2 | 0.507 | 0.005 |
| ## | 9 | 2017 | JaMyc~ | 8.53e6 | PF | 26 | MEM | 77 | 75 | 2101 | 13.5 | 0.601 | 0.290 |
| ## | 10 | 2017 | Juliu~ | 4.15e6 | PF | 22 | LAL | 74 | 73 | 2132 | 16.3 | 0.543 | 0.082 |

... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>, `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>, `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>, DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>, `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>

Histogram of PTS

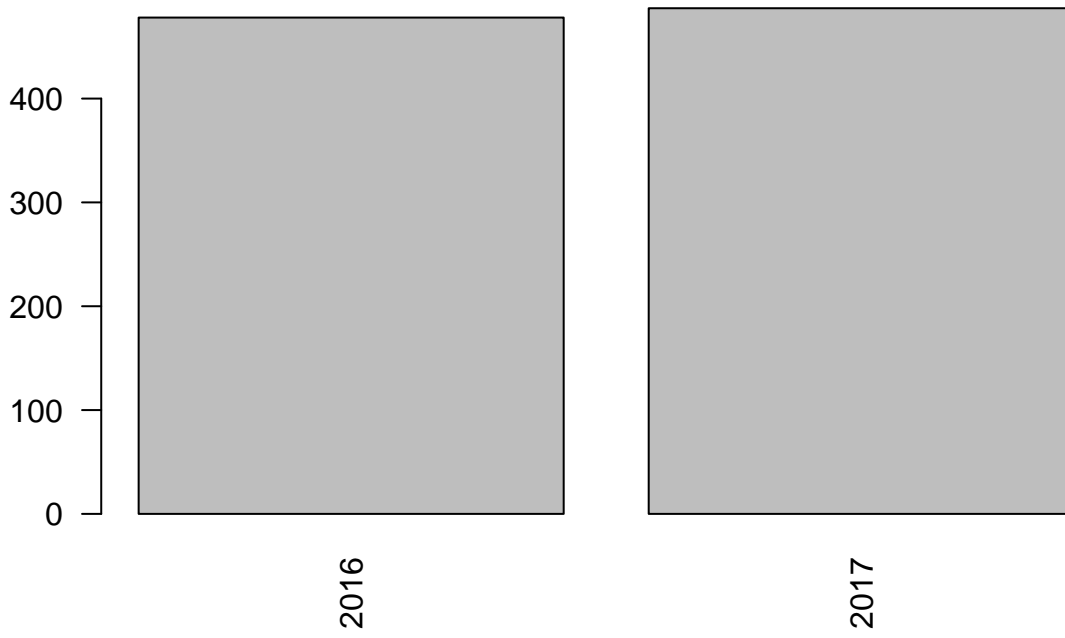


```
## Top 10 Players by PTS
## # A tibble: 10 x 51
##   year name_p salary Pos Age Tm G GS MP PER `TS%` `3PAr`
##   <fct> <chr> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Russe~ 2.85e7 PG 28 OKC 81 81 2802 30.6 0.554 0.3
## 2 2016 James~ 2.65e7 SG 26 HOU 82 82 3125 25.3 0.598 0.406
## 3 2016 Steph~ 1.21e7 PG 27 GSW 79 79 2700 31.5 0.669 0.554
## 4 2017 James~ 2.83e7 PG 27 HOU 81 81 2947 27.3 0.613 0.493
## 5 2017 Isaia~ 6.26e6 PG 27 BOS 76 76 2569 26.5 0.625 0.439
## 6 2017 Antho~ 2.38e7 C 23 NOP 75 75 2708 27.5 0.579 0.088
## 7 2017 Karl~ 6.22e6 C 21 MIN 82 82 3030 26 0.618 0.186
## 8 2016 Kevin~ 2.65e7 SF 27 OKC 72 72 2578 28.2 0.634 0.348
## 9 2017 Damia~ 2.62e7 PG 26 POR 75 75 2694 24.1 0.586 0.388
## 10 2017 DeMar~ 2.77e7 SG 27 TOR 74 74 2620 24 0.552 0.08
## # ... with 39 more variables: FTr <dbl>, `ORB%` <dbl>, `DRB%` <dbl>,
## # `TRB%` <dbl>, `AST%` <dbl>, `STL%` <dbl>, `BLK%` <dbl>, `TOV%` <dbl>,
## # `USG%` <dbl>, OWS <dbl>, DWS <dbl>, WS <dbl>, `WS/48` <dbl>, OBPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VORP <dbl>, FG <dbl>, FGA <dbl>, `FG%` <dbl>,
## # `3P` <dbl>, `3PA` <dbl>, `3P%` <dbl>, `2P` <dbl>, `2PA` <dbl>, `2P%` <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>
```

Histograms for Categorical Variables

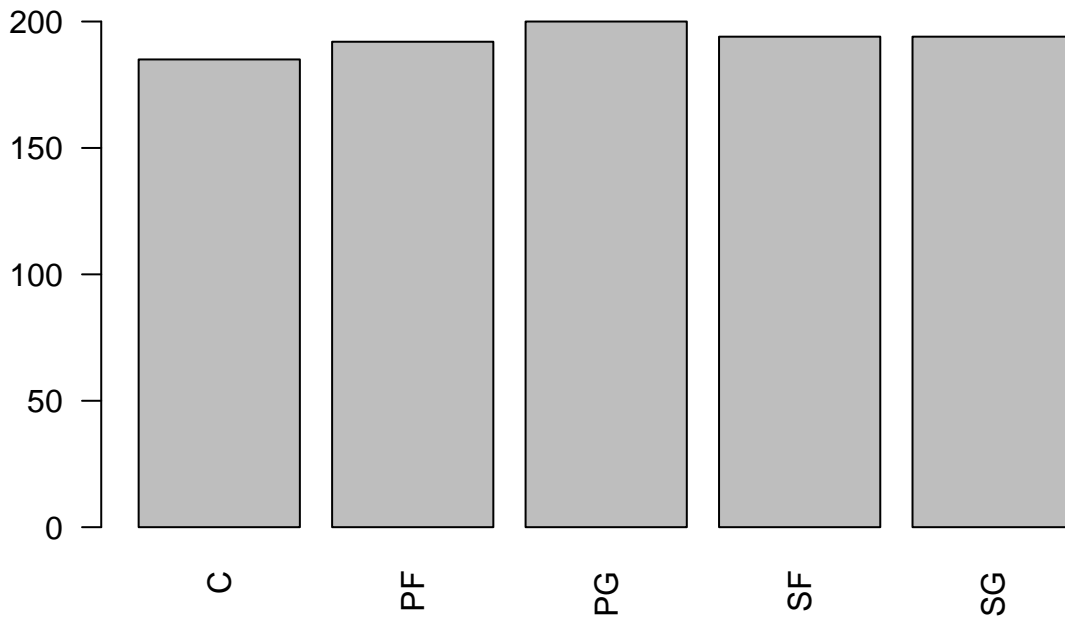
```
categorical_vars <- c('year', 'Pos', 'Tm')
for (col in categorical_vars){
  data <- df_primary[[col]]
  barplot(table(data), main=sprintf('Histogram of %s', col), las=2)
  print('\n')}
```

Histogram of year



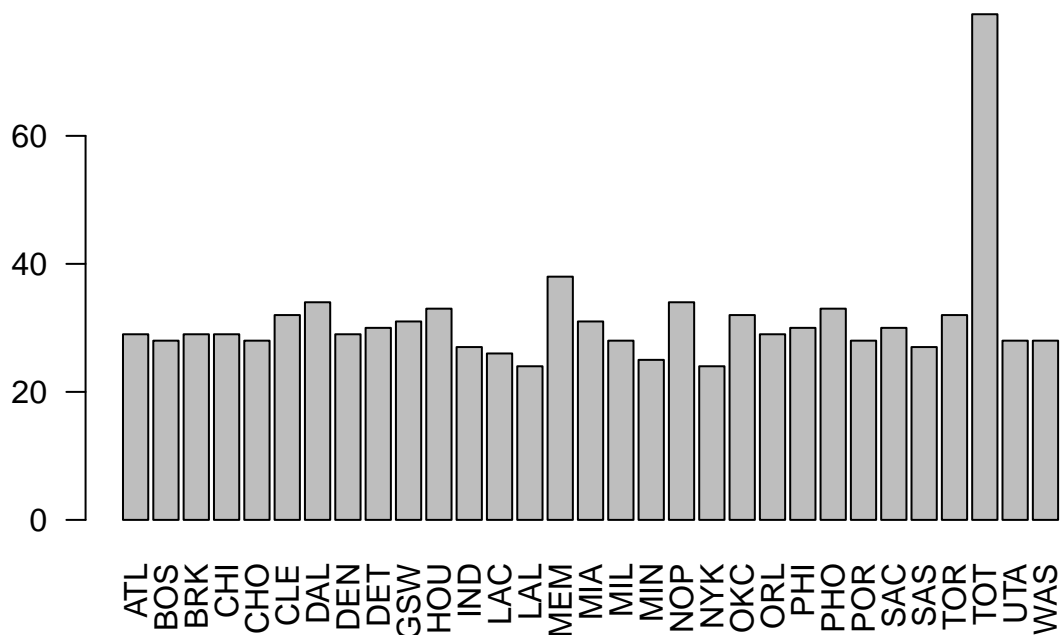
[1] "\n"

Histogram of Pos



[1] "\n"

Histogram of Tm



```
## [1] "\n"
```

Pool Together and Clean NBA 2K Data (Secondary Dataset)

```
secondary_attriutes <- c('name_s','position_s','ovr','out','ins','pla','ath','def','reb')
df_secondary <- vector('list',9)
names(df_secondary) <- secondary_attriutes
path_f = '../data/raw/nba2k/nba2k_%d.csv'
for (year in c(16:20)){
  df_year <- read.csv(sprintf(path_f,year))
  headers <- names(df_year)
  names(df_year) <- c('drop1',headers[1:length(headers)-1])
  df_year <- df_year[,c('name','position','ovr','out','ins','pla','ath','def','reb')]
  names(df_year) <- secondary_attriutes
  df_year[, 'year'] <- 2000+year
  df_secondary <- rbind(df_secondary,df_year)}
df_secondary[is.na(df_secondary)] <- 0
df_secondary <- df_secondary[df_secondary$year%in%c(2016,2017),] # take 2016-2017 2K ratings data
head(df_secondary)
```

```
##           name_s position_s ovr out ins pla ath def reb year
## 1      '96 Michael Jordan      SG  99  95  88  91  93  92  75 2016
## 2      '15 Kobe Bryant      SG  99  97  79  95  84  88  65 2016
## 3      Stephen Curry      PG  99  98  66  98  89  78  54 2016
## 4      LeBron James      SF  99  94  89  91  92  91  91 2016
## 5 '71 Kareem Abdul-Jabbar      C  99  75  93  56  89  86  98 2016
## 6      Kyrie Irving      PG  98  98  70  95  91  74  49 2016
```

```
summary(df_secondary)
```

```
##           name_s      position_s      ovr      out
## Jimmy Butler      : 10      PG      :812      Min.      :40.00      Min.      :25.0
## Kyrie Irving      : 10      SF      :782      1st Qu.:71.00      1st Qu.:62.0
## Russell Westbrook: 10      SG      :749      Median :78.00      Median :73.0
## Damian Lillard    : 9      PF      :710      Mean     :78.89      Mean     :71.3
## Demar Derozan     : 9      C      :708      3rd Qu.:86.00      3rd Qu.:82.0
## James Harden      : 9      C/PF     : 0      Max.     :99.00      Max.     :99.0
## (Other)           :3704      (Other): 0
##           ins      pla      ath      def
## Min.      :25.00      Min.      :25.00      Min.      :25.00      Min.      :25.00
```

```
## 1st Qu.:58.00 1st Qu.:48.00 1st Qu.:68.00 1st Qu.:58.00
## Median :64.00 Median :61.00 Median :74.00 Median :65.00
## Mean :65.43 Mean :62.04 Mean :73.68 Mean :66.28
## 3rd Qu.:72.00 3rd Qu.:76.00 3rd Qu.:80.00 3rd Qu.:73.00
## Max. :98.00 Max. :99.00 Max. :98.00 Max. :98.00
##
## reb year
## Min. :25.00 Min. :2016
## 1st Qu.:43.00 1st Qu.:2016
## Median :57.00 Median :2016
## Mean :59.62 Mean :2016
## 3rd Qu.:75.00 3rd Qu.:2017
## Max. :99.00 Max. :2017
##
```

Numeric / Factor Variables

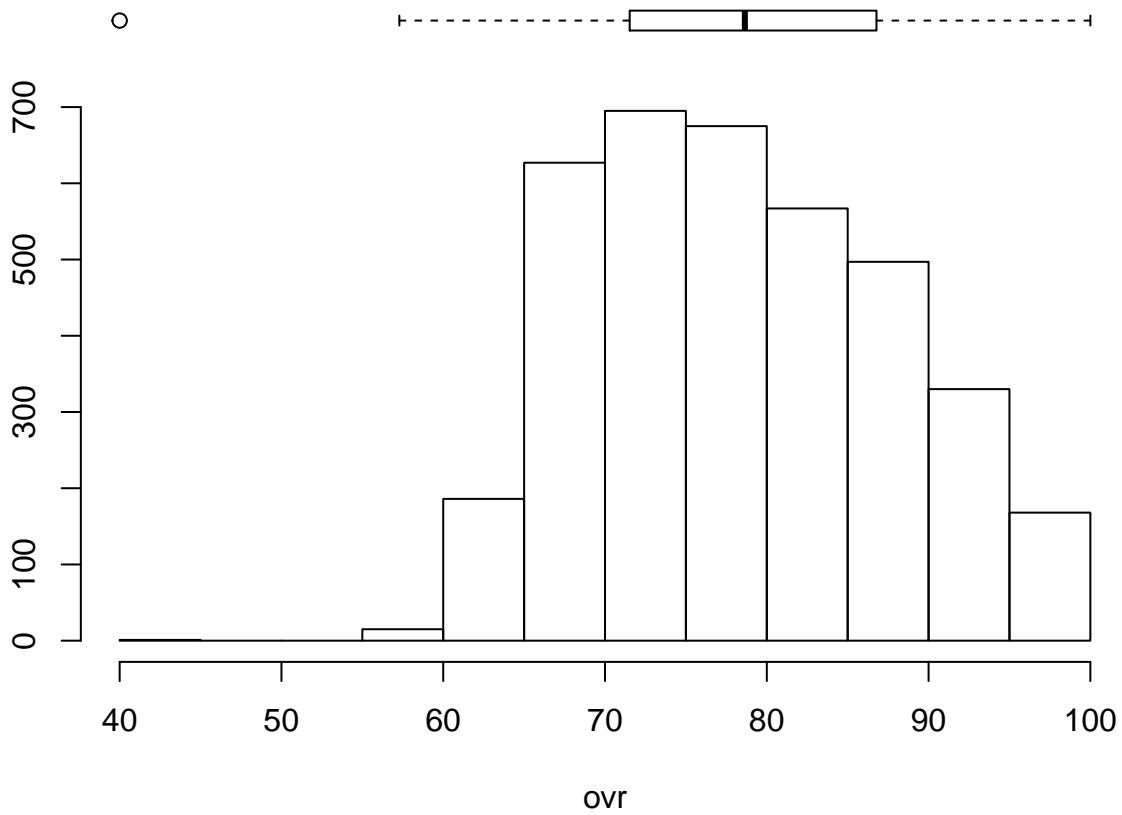
```
df_secondary$name_s <- as.character(df_secondary$name_s)
df_secondary$year <- as.factor(df_secondary$year) # make year a factor variable
df_secondary$position_s <- factor(df_secondary$position_s) # make position a factor variable
str(df_secondary)
```

```
## 'data.frame': 3761 obs. of 10 variables:
## $ name_s : chr "'96 Michael Jordan" "'15 Kobe Bryant" "Stephen Curry" "LeBron James" ...
## $ position_s: Factor w/ 5 levels "C","PF","PG",...: 5 5 3 4 1 3 3 5 2 5 ...
## $ ovr : int 99 99 99 99 99 98 98 98 98 98 ...
## $ out : int 95 97 98 94 75 98 92 90 84 96 ...
## $ ins : int 88 79 66 89 93 70 78 82 89 81 ...
## $ pla : int 91 95 98 91 56 95 98 93 76 81 ...
## $ ath : int 93 84 89 92 89 91 90 92 81 88 ...
## $ def : int 92 88 78 91 86 74 84 83 87 83 ...
## $ reb : int 75 65 54 91 98 49 88 76 98 60 ...
## $ year : Factor w/ 2 levels "2016","2017": 1 1 1 1 1 1 1 1 1 1 ...
```

Histogram Barcharts for Numeric Variables

```
df_s_numeric <- Filter(is.numeric,df_secondary) # numeric variables
for (col in names(df_s_numeric)){
  data <- df_s_numeric[[col]]
  layout(mat = matrix(c(1,2),2,1, byrow=TRUE), height = c(1,8))
  par(mar=c(0, 3.1, 1.1, 2.1))
  boxplot(data , horizontal=TRUE , xaxt="n", frame=F, main=sprintf('Histogram of %s',col))
  par(mar=c(4, 3.1, 1.1, 2.1))
  hist(data,xlab=col,main='')
  # print top players in this category
  cat(sprintf('Top 10 Players by %s\n',col))
  df_top <- df_secondary[order(df_secondary[[col]],decreasing=T),]
  print(df_top[1:10,])}
```

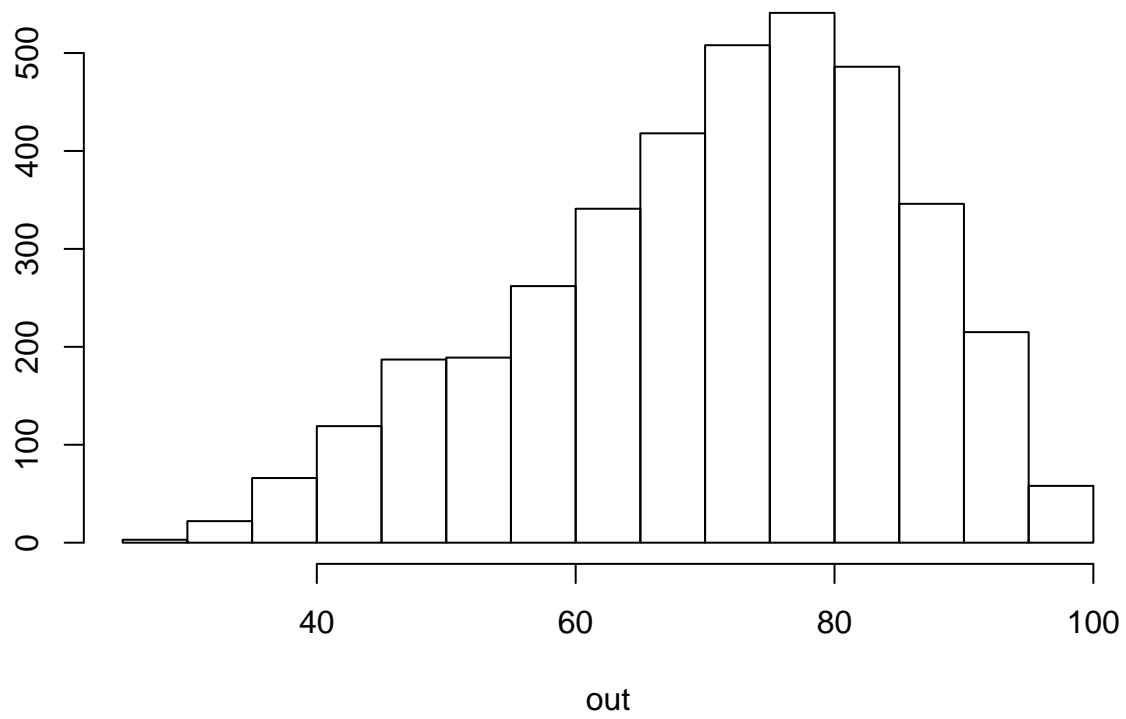
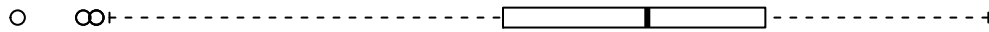
Histogram of ovr



Top 10 Players by ovr

| ## | name_s | position_s | ovr | out | ins | pla | ath | def | reb | year |
|---------|-------------------------|------------|-----|-----|-----|-----|-----|-----|-----|------|
| ## 1 | '96 Michael Jordan | SG | 99 | 95 | 88 | 91 | 93 | 92 | 75 | 2016 |
| ## 2 | '15 Kobe Bryant | SG | 99 | 97 | 79 | 95 | 84 | 88 | 65 | 2016 |
| ## 3 | Stephen Curry | PG | 99 | 98 | 66 | 98 | 89 | 78 | 54 | 2016 |
| ## 4 | LeBron James | SF | 99 | 94 | 89 | 91 | 92 | 91 | 91 | 2016 |
| ## 5 | '71 Kareem Abdul-Jabbar | C | 99 | 75 | 93 | 56 | 89 | 86 | 98 | 2016 |
| ## 2082 | Kobe Bryant | SG | 99 | 98 | 93 | 91 | 94 | 91 | 74 | 2017 |
| ## 2083 | Wilt Chamberlain | C | 99 | 65 | 95 | 68 | 89 | 88 | 98 | 2017 |
| ## 2084 | Jerry West | PG | 99 | 97 | 68 | 94 | 90 | 85 | 65 | 2017 |
| ## 2085 | Kobe Bryant | SG | 99 | 97 | 82 | 82 | 89 | 83 | 60 | 2017 |
| ## 2086 | Michael Jordan | SG | 99 | 94 | 85 | 86 | 91 | 91 | 66 | 2017 |

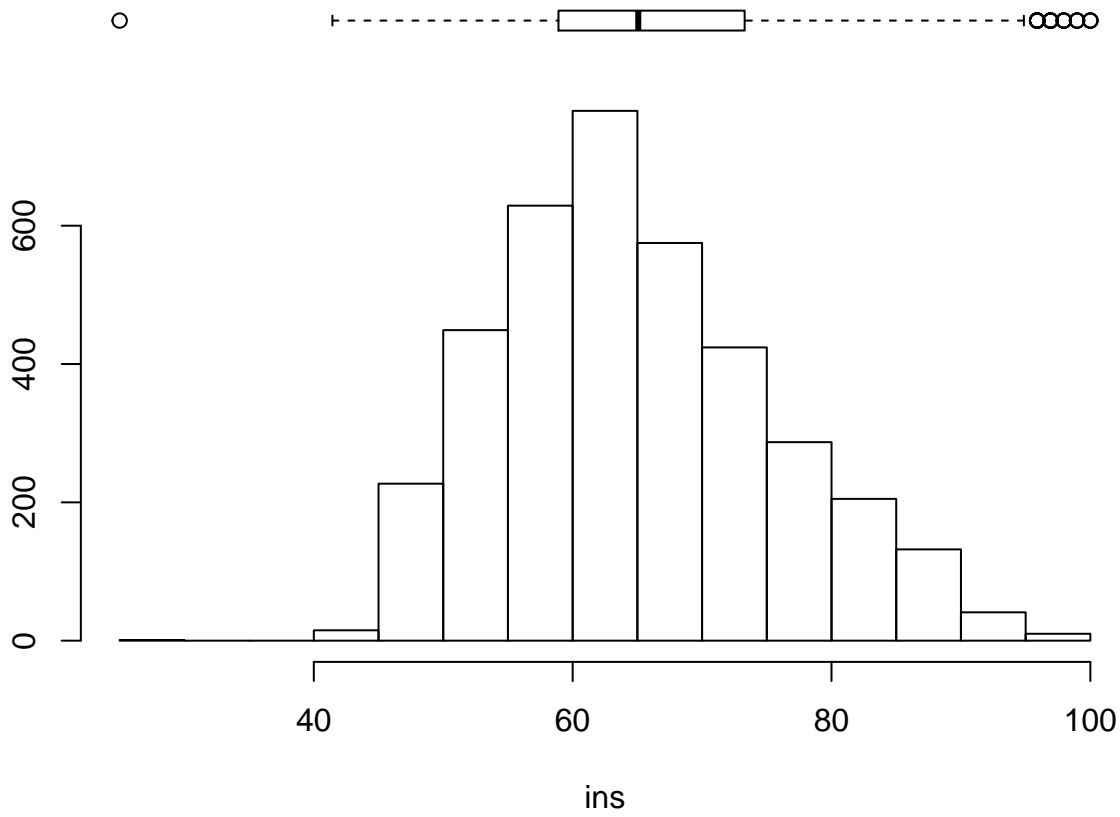
Histogram of out



Top 10 Players by out

| ## | name_s | position_s | ovr | out | ins | pla | ath | def | reb | year |
|---------|------------------|------------|-----|-----|-----|-----|-----|-----|-----|------|
| ## 2098 | Stephen Curry | PG | 98 | 99 | 70 | 98 | 92 | 86 | 78 | 2017 |
| ## 3 | Stephen Curry | PG | 99 | 98 | 66 | 98 | 89 | 78 | 54 | 2016 |
| ## 6 | Kyrie Irving | PG | 98 | 98 | 70 | 95 | 91 | 74 | 49 | 2016 |
| ## 17 | Kyrie Irving TBT | PG | 97 | 98 | 67 | 94 | 86 | 74 | 49 | 2016 |
| ## 35 | Klay Thompson | SG | 97 | 98 | 77 | 79 | 89 | 86 | 49 | 2016 |
| ## 2082 | Kobe Bryant | SG | 99 | 98 | 93 | 91 | 94 | 91 | 74 | 2017 |
| ## 2100 | Kevin Durant | SF | 98 | 98 | 88 | 85 | 84 | 91 | 82 | 2017 |
| ## 2105 | James Harden | SG | 98 | 98 | 85 | 98 | 91 | 80 | 88 | 2017 |
| ## 2143 | Isaiah Thomas | PG | 97 | 98 | 62 | 97 | 92 | 71 | 54 | 2017 |
| ## 2146 | Klay Thompson | SG | 97 | 98 | 81 | 84 | 88 | 93 | 55 | 2017 |

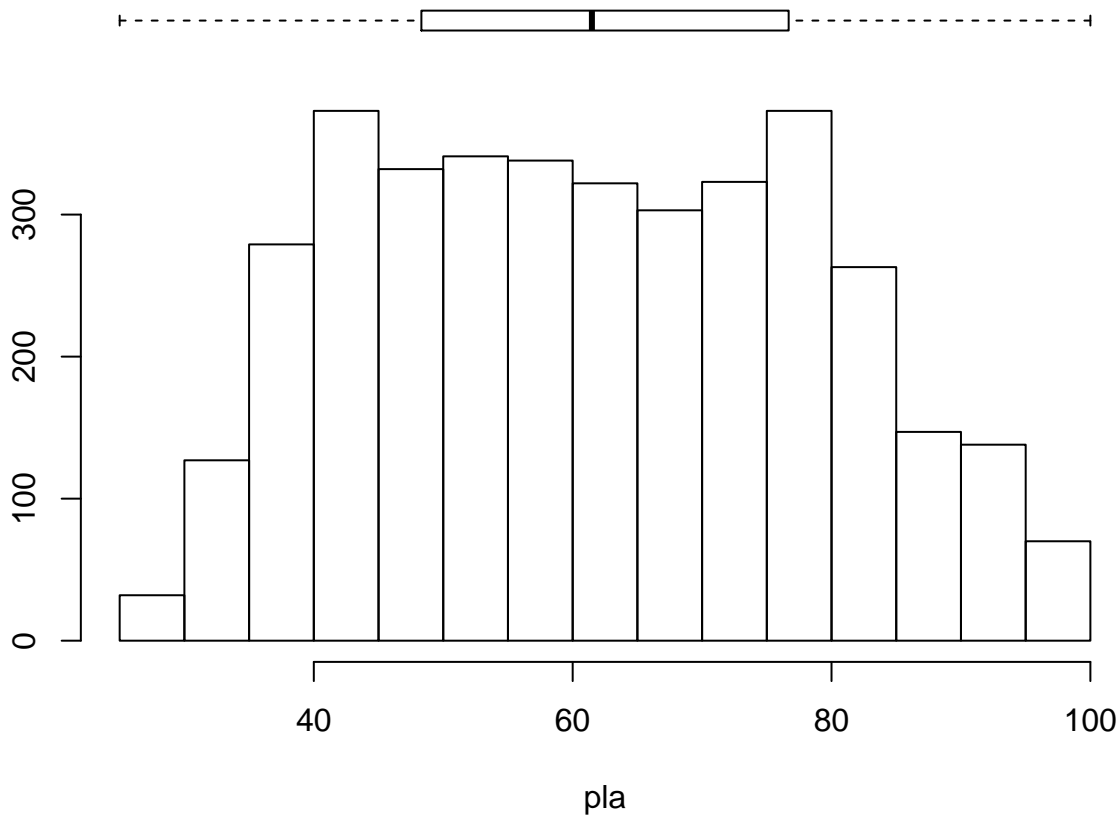
Histogram of ins



Top 10 Players by ins

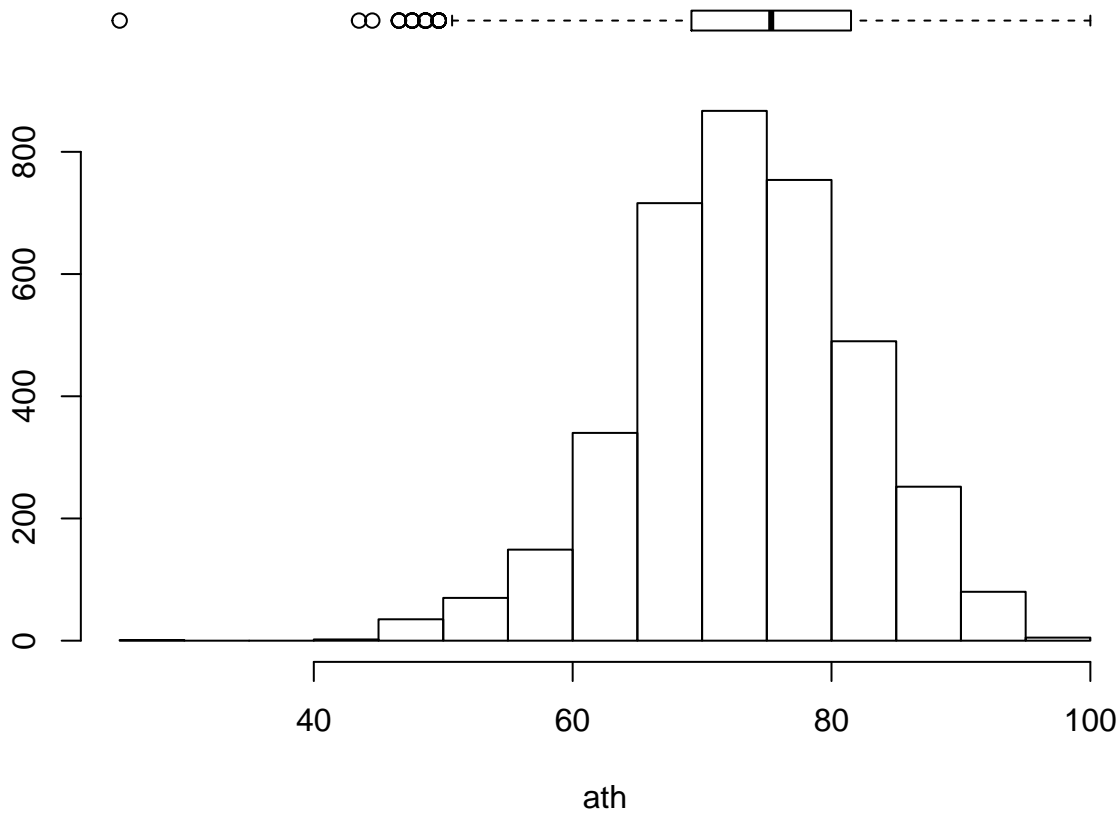
| ## | name_s | position_s | ovr | out | ins | pla | ath | def | reb | year |
|---------|---------------------|------------|-----|-----|-----|-----|-----|-----|-----|------|
| ## 2091 | Charles Barkley | PF | 99 | 95 | 98 | 89 | 95 | 97 | 98 | 2017 |
| ## 2106 | Kareem Abdul-Jabbar | C | 98 | 83 | 98 | 85 | 90 | 95 | 98 | 2017 |
| ## 2107 | Karl Malone | PF | 98 | 88 | 98 | 80 | 96 | 94 | 98 | 2017 |
| ## 2110 | Anthony Davis | PF | 98 | 87 | 97 | 65 | 91 | 94 | 97 | 2017 |
| ## 2135 | Kevin Garnett | PF | 97 | 86 | 97 | 80 | 93 | 94 | 98 | 2017 |
| ## 2227 | Amar'e Stoudemire | PF | 95 | 82 | 97 | 63 | 88 | 82 | 94 | 2017 |
| ## 2096 | Michael Jordan | SG | 99 | 97 | 96 | 95 | 96 | 95 | 80 | 2017 |
| ## 2103 | Bill Russell | C | 98 | 57 | 96 | 79 | 92 | 97 | 99 | 2017 |
| ## 2205 | Wes Unseld | C | 95 | 78 | 96 | 89 | 90 | 94 | 98 | 2017 |
| ## 2213 | Shawn Kemp | PF | 95 | 83 | 96 | 63 | 92 | 85 | 95 | 2017 |

Histogram of pla



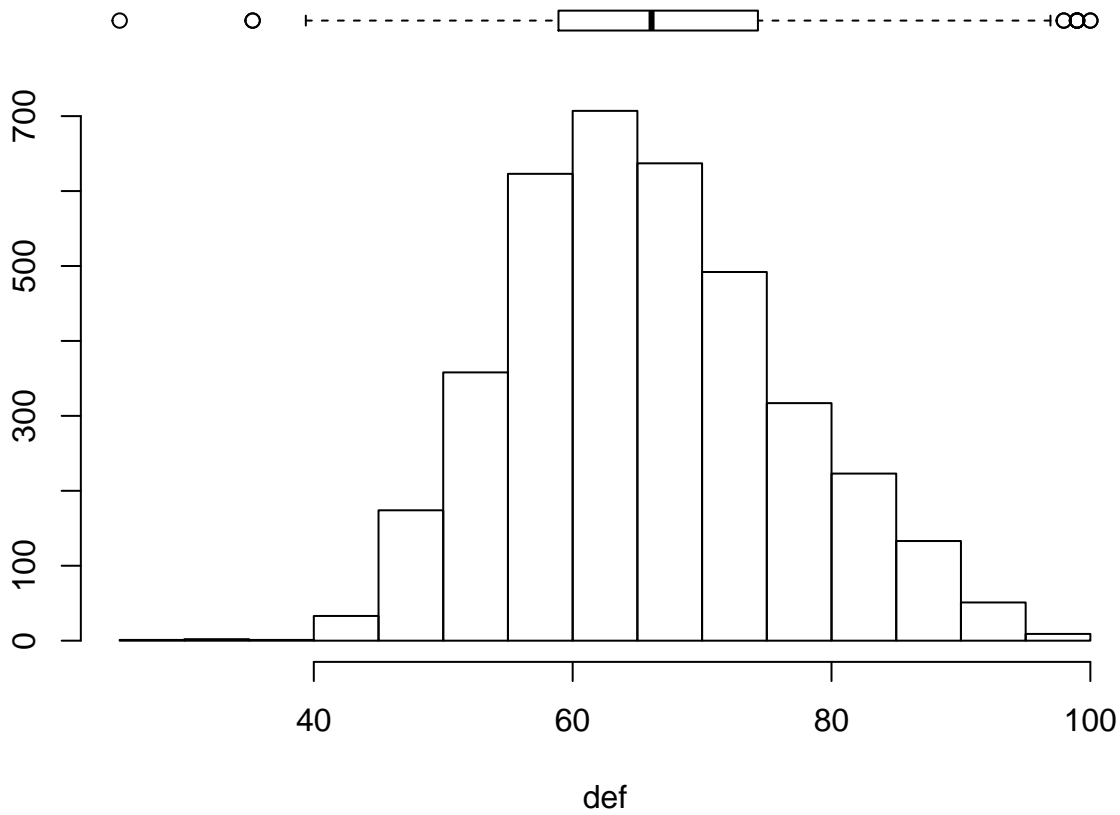
| | | | | | | | | | | |
|--------------------------|---------------------|------------|-----|-----|-----|-----|-----|-----|-----|------|
| ## Top 10 Players by pla | | | | | | | | | | |
| ## | name_s | position_s | ovr | out | ins | pla | ath | def | reb | year |
| ## 2101 | Magic Johnson | PG | 98 | 95 | 90 | 99 | 96 | 95 | 89 | 2017 |
| ## 2115 | John Stockton | PG | 98 | 97 | 70 | 99 | 92 | 90 | 51 | 2017 |
| ## 3 | Stephen Curry | PG | 99 | 98 | 66 | 98 | 89 | 78 | 54 | 2016 |
| ## 7 | '62 Oscar Robertson | PG | 98 | 92 | 78 | 98 | 90 | 84 | 88 | 2016 |
| ## 14 | '90 John Stockton | PG | 97 | 93 | 64 | 98 | 86 | 86 | 38 | 2016 |
| ## 61 | '57 Bob Cousy | PG | 96 | 92 | 65 | 98 | 83 | 82 | 59 | 2016 |
| ## 72 | '07 Steve Nash | PG | 96 | 95 | 61 | 98 | 85 | 75 | 42 | 2016 |
| ## 79 | '85 Isiah Thomas | PG | 95 | 87 | 65 | 98 | 90 | 82 | 52 | 2016 |
| ## 94 | '02 Jason Kidd | PG | 95 | 86 | 64 | 98 | 83 | 85 | 75 | 2016 |
| ## 2090 | Isiah Thomas | PG | 99 | 94 | 69 | 98 | 90 | 83 | 40 | 2017 |

Histogram of ath



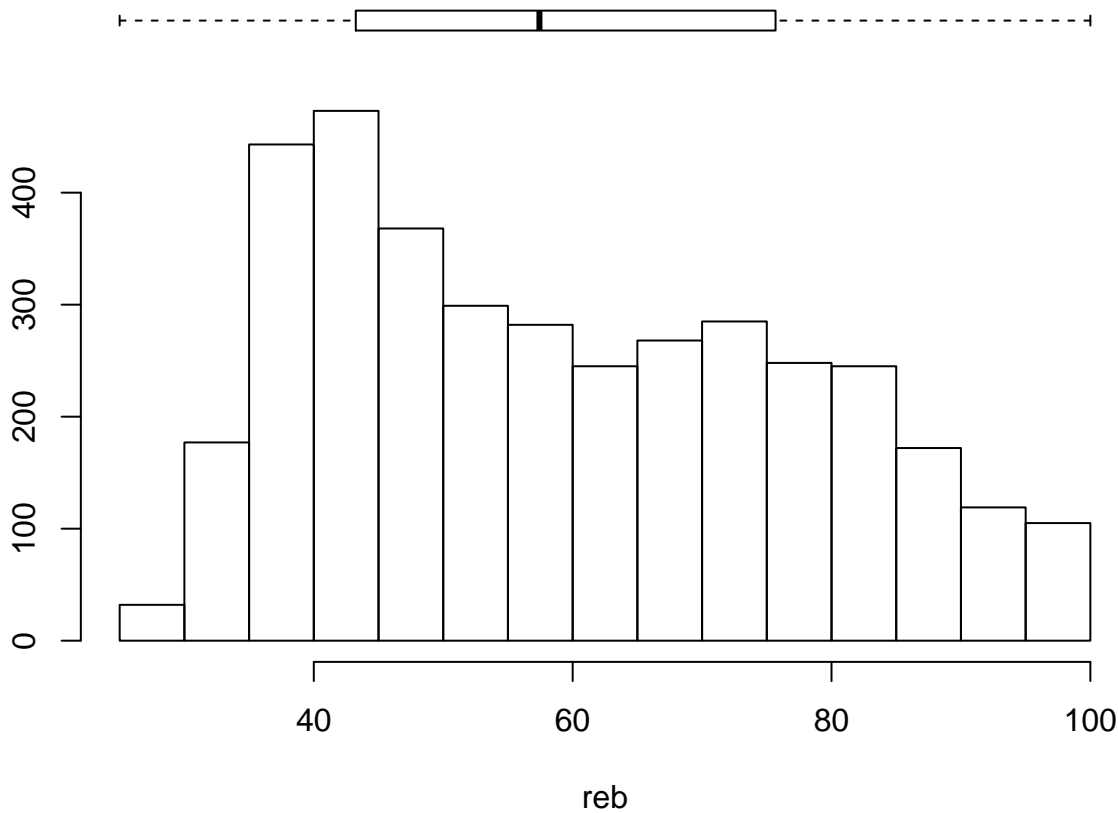
| ## Top 10 Players by ath | | | | | | | | | | |
|--------------------------|-------------------|------------|-----|-----|-----|-----|-----|-----|-----|------|
| ## | name_s | position_s | ovr | out | ins | pla | ath | def | reb | year |
| ## 2097 | Russell Westbrook | PG | 99 | 97 | 83 | 98 | 98 | 91 | 97 | 2017 |
| ## 2099 | Lebron James | SF | 98 | 94 | 94 | 95 | 97 | 92 | 74 | 2017 |
| ## 2096 | Michael Jordan | SG | 99 | 97 | 96 | 95 | 96 | 95 | 80 | 2017 |
| ## 2101 | Magic Johnson | PG | 98 | 95 | 90 | 99 | 96 | 95 | 89 | 2017 |
| ## 2107 | Karl Malone | PF | 98 | 88 | 98 | 80 | 96 | 94 | 98 | 2017 |
| ## 2091 | Charles Barkley | PF | 99 | 95 | 98 | 89 | 95 | 97 | 98 | 2017 |
| ## 2108 | Allen Iverson | SG | 98 | 96 | 71 | 97 | 95 | 84 | 53 | 2017 |
| ## 2112 | Russell Westbrook | PG | 98 | 96 | 76 | 97 | 95 | 86 | 93 | 2017 |
| ## 2148 | Chauncey Billups | PG | 96 | 97 | 64 | 97 | 95 | 90 | 43 | 2017 |
| ## 2200 | Bob Cousy | PG | 95 | 97 | 70 | 98 | 95 | 83 | 75 | 2017 |

Histogram of def



| | | | | | | | | | | |
|--------------------------|------------------|------------|-----|-----|-----|-----|-----|-----|-----|------|
| ## Top 10 Players by def | | | | | | | | | | |
| ## | name_s | position_s | ovr | out | ins | pla | ath | def | reb | year |
| ## 2109 | Dennis Rodman | PF | 98 | 77 | 88 | 61 | 92 | 98 | 99 | 2017 |
| ## 2233 | Draymond Green | PF | 95 | 90 | 86 | 94 | 91 | 98 | 92 | 2017 |
| ## 2091 | Charles Barkley | PF | 99 | 95 | 98 | 89 | 95 | 97 | 98 | 2017 |
| ## 2103 | Bill Russell | C | 98 | 57 | 96 | 79 | 92 | 97 | 99 | 2017 |
| ## 2111 | Hakeem Olajuwon | C | 98 | 82 | 94 | 74 | 84 | 97 | 98 | 2017 |
| ## 2195 | Ben Wallace | C | 95 | 46 | 85 | 52 | 89 | 97 | 97 | 2017 |
| ## 2321 | Dave Debusschere | PF | 92 | 90 | 90 | 82 | 85 | 97 | 96 | 2017 |
| ## 2102 | Shaquille O'Neal | C | 98 | 55 | 95 | 75 | 90 | 96 | 99 | 2017 |
| ## 2104 | Larry Bird | SF | 98 | 96 | 88 | 92 | 89 | 96 | 94 | 2017 |
| ## 2087 | Tim Duncan | PF | 99 | 74 | 95 | 73 | 87 | 95 | 98 | 2017 |

Histogram of reb



Top 10 Players by reb

| ## | name_s | position_s | ovr | out | ins | pla | ath | def | reb | year |
|---------|-------------------------|------------|-----|-----|-----|-----|-----|-----|-----|------|
| ## 2102 | Shaquille O'Neal | C | 98 | 55 | 95 | 75 | 90 | 96 | 99 | 2017 |
| ## 2103 | Bill Russell | C | 98 | 57 | 96 | 79 | 92 | 97 | 99 | 2017 |
| ## 2109 | Dennis Rodman | PF | 98 | 77 | 88 | 61 | 92 | 98 | 99 | 2017 |
| ## 2345 | Dennis Rodman | PF | 92 | 59 | 74 | 49 | 85 | 87 | 99 | 2017 |
| ## 2598 | Dennis Rodman | PF | 88 | 54 | 68 | 49 | 80 | 85 | 99 | 2017 |
| ## 5 | '71 Kareem Abdul-Jabbar | C | 99 | 75 | 93 | 56 | 89 | 86 | 98 | 2016 |
| ## 9 | '03 Tim Duncan | PF | 98 | 84 | 89 | 76 | 81 | 87 | 98 | 2016 |
| ## 18 | '60 Bill Russell | C | 97 | 58 | 88 | 73 | 90 | 93 | 98 | 2016 |
| ## 21 | '62 Bill Russell | C | 97 | 57 | 89 | 73 | 88 | 92 | 98 | 2016 |
| ## 24 | Anthony Davis | PF | 97 | 91 | 89 | 64 | 87 | 87 | 98 | 2016 |

Histograms for Categorical Variables

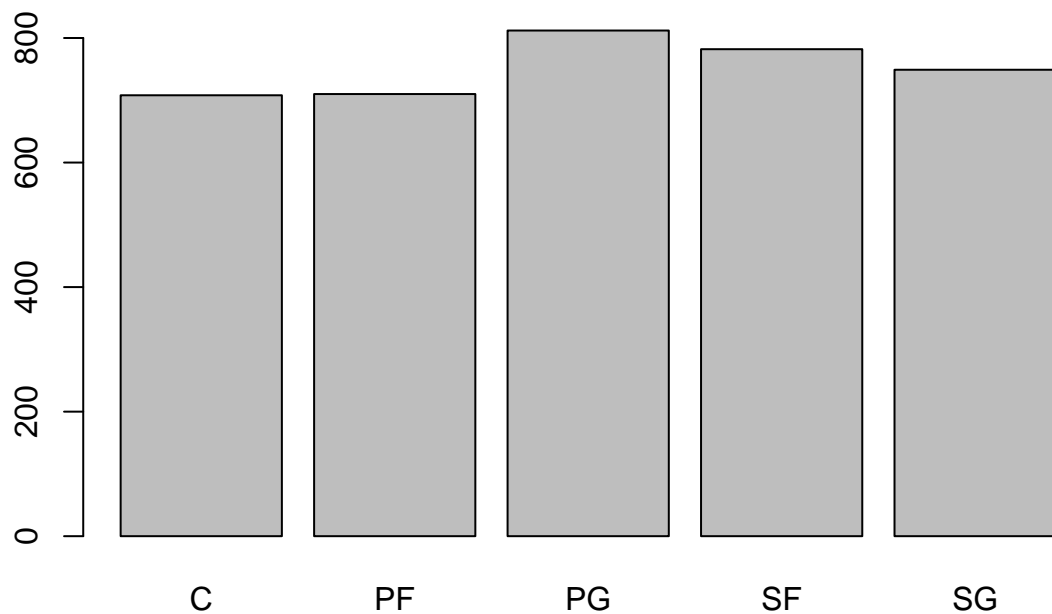
```
categorical_vars <- c('year','position_s')
for (col in categorical_vars){
  data <- df_secondary[[col]]
  barplot(table(data),main=sprintf('Histogram of %s',col))
  print('\n')}
```

Histogram of year



```
## [1] "\n"
```

Histogram of position_s



```
## [1] "\n"
```

Merge Primary and Secondary Datasets

Name Cleaning

```
library(stringdist)
library(reshape)
library(stringr)
sub_n_diff_score <- function(ppl,n){
  str_dist <- melt(stringdistmatrix(a=ppl, b=ppl, method = 'lv', useNames = 'strings'))
  str_dist_1_to_n <- str_dist[str_dist$value<=n & str_dist$value>0,]
  return (str_dist_1_to_n[order(str_dist_1_to_n$value),])}
clean_names <- function(names){
  names <- tolower(names)
  names <- str_squish(names)
  names <- gsub('\\\\.', '', names)
```

```

names <- gsub('-', ' ', names)
return (names)}

df_primary$name <- clean_names(df_primary[['name_p']]) # clean primary dataset names
df_secondary$name <- clean_names(df_secondary[['name_s']]) # clean secondary dataset
df_primary$name <- iconv(df_primary$name, to='ASCII//TRANSLIT') # convert to ascii
df_secondary$name <- iconv(df_secondary$name, to='ASCII//TRANSLIT') # convert to asii
df_secondary <- df_secondary[!grepl("\\d", df_secondary$name),]
# remove players with numbers in name as this signifies a legendary player
df_secondary <- df_secondary[!grepl("dynamic", df_secondary$name),]
# remove dynamic versions of players
replace_names <- list(
  `isiaah thomas` = 'isaiah thomas',
  `jonathan simmons` = 'jonathon simmons',
  `lance stepheson` = 'lance stephenson',
  `luke babbitt` = 'luke babbitt',
  `luke babbitt` = 'luke babbitt',
  `patrick beverly` = 'patrick beverley',
  `willis reed` = 'willie reed',
  `kiki vanderweghe` = 'kiki vandeweghe',
  `mychael thompson` = 'mychal thompson',
  `drayamond green` = 'draymond green',
  `louis amundson` = 'lou amundson',
  `louis williams` = 'lou williams')
for (n in names(replace_names)){
  df_primary$name <- gsub(n, replace_names[[n]], df_primary$name)
  df_secondary$name <- gsub(n, replace_names[[n]], df_secondary$name)
}
all_names <- unique(c(df_primary$name, df_secondary$name))
sub_n_diff_score(unique(all_names), 2)

```

| ## | X1 | X2 | value |
|-----------|---------------------|---------------------|-------|
| ## 162251 | zoran dragic | goran dragic | 1 |
| ## 608586 | goran dragic | zoran dragic | 1 |
| ## 7431 | ryan anderson | alan anderson | 2 |
| ## 8946 | alvin williams | alan williams | 2 |
| ## 90346 | damon jones | damian jones | 2 |
| ## 104062 | david wear | david west | 2 |
| ## 133344 | dryamond green | draymond green | 2 |
| ## 161525 | flynn robinson | glenn robinson | 2 |
| ## 178959 | josh smith | ish smith | 2 |
| ## 216793 | brian grant | jerian grant | 2 |
| ## 247163 | ish smith | josh smith | 2 |
| ## 295513 | mo williams | lou williams | 2 |
| ## 310901 | darius morris | marcus morris | 2 |
| ## 319182 | alvin williams | marvin williams | 2 |
| ## 337639 | lou williams | mo williams | 2 |
| ## 375368 | paul pressey | phil pressey | 2 |
| ## 403616 | alan anderson | ryan anderson | 2 |
| ## 478426 | willie green | willie reed | 2 |
| ## 486897 | joe bryant | kobe bryant | 2 |
| ## 523561 | willie reed | willie green | 2 |
| ## 586233 | darius miles | darius miller | 2 |
| ## 608388 | drew gooden | drew gordon | 2 |
| ## 651700 | david west | david wear | 2 |
| ## 666966 | marcus morris | darius morris | 2 |
| ## 685574 | charles oakley | charles barkley | 2 |
| ## 748935 | ervin johnson | kevin johnson | 2 |
| ## 805078 | shareef abdur rahim | shareef abdur rahim | 2 |
| ## 818393 | draymond green | dryamond green | 2 |
| ## 849063 | charles barkley | charles oakley | 2 |
| ## 877294 | shareef abdur rahim | shareef abdur rahim | 2 |
| ## 878874 | phil pressey | paul pressey | 2 |
| ## 887697 | glenn robinson | flynn robinson | 2 |

```
## 904184      darius miller      darius miles      2
## 916661      alan williams     alvin williams     2
## 916970      marvin williams    alvin williams     2
## 935944      jerian grant      brian grant       2
## 958562      kevin johnson     ervin johnson      2
## 964325      kobe bryant       joe bryant         2
## 971474      drew gordon      drew gooden        2
## 993046      damian jones      damon jones        2
```

Joining Datasets

```
# if multiple versions of a player, take the one with the max overall
df_secondary_max <- aggregate(df_secondary['ovr'],df_secondary[c('name','year')],max)
df_secondary_max <- merge(df_secondary_max,df_secondary,by=c('name','year','ovr'),all=F)
df_secondary_max_2 <- aggregate(df_secondary_max['out'],df_secondary_max[c('name','year')],max)
df_full_s <- merge(df_secondary_max,df_secondary_max_2,by=c('name','year','out'),all=F)
# only take totals from players who changed teams mid-year
df_p_tot <- df_primary[df_primary$Tm=='TOT',]
traded_player_years <- interaction(df_primary[,c('year','name')]) %in%
  interaction(df_p_tot[,c('year','name')])
df_p_wo_tot <- df_primary[!traded_player_years,]
df_full_p <- rbind(df_p_wo_tot,df_p_tot)
# join datasets
df_full <- merge(df_full_p,df_full_s,by=c('name','year'),all=F)
df_full <- df_full[order(df_full$name,df_full$year),]
df_full <- unique(df_full)
head(df_full[,1:5])
```

```
##           name year      name_p salary Pos
## 1  aaron brooks 2016  Aaron Brooks 2700000 PG
## 2  aaron brooks 2017  Aaron Brooks 2116955 PG
## 3  aaron gordon 2016  Aaron Gordon 4351320 PF
## 4  aaron gordon 2017  Aaron Gordon 5504420 SF
## 5 adreian payne 2016 Adreian Payne 2022240 PF
## 6    aj hammons 2017   A.J. Hammons 1312611  C
```

```
# joined datasets checks
max(table(df_full$name)) # should be 2 (2016,2017)
```

```
## [1] 2
nrow(df_full)
```

```
## [1] 734
```

Clean Up Joined Data

```
drop_cols <- c('name_p','name_s','position_s')
df_final <- df_full[,!(names(df_full)%in%drop_cols)]
names(df_final)[names(df_final)=='position_p'] <- 'position'
s_columns <- c('ovr','out','ins','pla','ath','def','reb')
df_p_final <- df_final[,!(names(df_final)%in%s_columns)] # final primary dataset
df_s_final <- df_final[,c('name',s_columns)] # final secondary dataset
summary(df_final)
```

```
##      name      year      salary      Pos      Age
## Length:734    2016:369  Min.   : 11534  C :157  Min.   :19.00
## Class :character 2017:365  1st Qu.: 2113599  PF:147  1st Qu.:23.00
## Mode  :character      Median : 5196000  PG:138  Median :26.00
##                        Mean    : 7829827  SF:143  Mean   :26.55
##                        3rd Qu.:12012640  SG:149  3rd Qu.:29.00
##                        Max.    :34682550      Max.   :40.00
##
```



```

##          Tm          G          GS          MP          PER
## TOT      : 70   Min.    : 1.00   Min.    : 0.00   Min.    : 6   Min.    : -7.70
## GSW      : 27   1st Qu.:52.00   1st Qu.: 3.00   1st Qu.: 848   1st Qu.:10.90
## TOR      : 27   Median :68.00   Median :20.50   Median :1506   Median :13.70
## PHO      : 25   Mean    :61.18   Mean    :31.72   Mean    :1474   Mean    :14.17
## UTA      : 25   3rd Qu.:77.00   3rd Qu.:62.75   3rd Qu.:2118   3rd Qu.:16.90
## DET      : 24   Max.    :82.00   Max.    :82.00   Max.    :3125   Max.    :32.00
## (Other):536
##          TS%          3PAr          FTr          ORB%
## Min.      :0.0000   Min.      :0.0000   Min.      :0.0000   Min.      : 0.000
## 1st Qu.:0.5090   1st Qu.:0.1060   1st Qu.:0.1760   1st Qu.: 2.000
## Median :0.5415   Median :0.3050   Median :0.2480   Median : 3.600
## Mean      :0.5382   Mean      :0.2933   Mean      :0.2717   Mean      : 5.068
## 3rd Qu.:0.5720   3rd Qu.:0.4427   3rd Qu.:0.3397   3rd Qu.: 7.500
## Max.      :1.0000   Max.      :0.9000   Max.      :1.2190   Max.      :21.800
##
##          DRB%          TRB%          AST%          STL%
## Min.      : 0.00   Min.      : 0.0   Min.      : 0.000   Min.      : 0.000
## 1st Qu.:10.53   1st Qu.: 6.3   1st Qu.: 7.125   1st Qu.: 1.100
## Median :14.60   Median : 9.3   Median :10.300   Median : 1.500
## Mean      :15.53   Mean      :10.3   Mean      :13.386   Mean      : 1.586
## 3rd Qu.:19.57   3rd Qu.:13.3   3rd Qu.:17.650   3rd Qu.: 1.900
## Max.      :36.30   Max.      :25.6   Max.      :57.300   Max.      :11.100
##
##          BLK%          TOV%          USG%          OWS
## Min.      :0.000   Min.      : 0.00   Min.      : 0.00   Min.      : -3.30
## 1st Qu.:0.600   1st Qu.:10.00   1st Qu.:15.40   1st Qu.: 0.20
## Median :1.200   Median :12.50   Median :18.50   Median : 1.10
## Mean      :1.736   Mean      :12.74   Mean      :19.18   Mean      : 1.72
## 3rd Qu.:2.500   3rd Qu.:15.10   3rd Qu.:22.18   3rd Qu.: 2.50
## Max.      :9.700   Max.      :43.60   Max.      :41.70   Max.      :13.80
##
##          DWS          WS          WS/48          OBPM
## Min.      :0.000   Min.      : -2.100   Min.      : -0.28300   Min.      : -17.3000
## 1st Qu.:0.700   1st Qu.: 1.100   1st Qu.: 0.05600   1st Qu.: -2.1000
## Median :1.300   Median : 2.500   Median : 0.09100   Median : -0.7000
## Mean      :1.521   Mean      : 3.241   Mean      : 0.09281   Mean      : -0.6693
## 3rd Qu.:2.200   3rd Qu.: 4.475   3rd Qu.: 0.12700   3rd Qu.: 0.5000
## Max.      :6.000   Max.      :17.900   Max.      : 0.34300   Max.      :12.4000
##
##          DBPM          BPM          VORP          FG
## Min.      : -8.20000   Min.      : -24.1000   Min.      : -1.4000   Min.      : 0.0
## 1st Qu.: -1.30000   1st Qu.: -2.6750   1st Qu.: -0.1000   1st Qu.:114.5
## Median : -0.10000   Median : -0.7000   Median : 0.4000   Median :208.0
## Mean      : -0.08965   Mean      : -0.7583   Mean      : 0.8349   Mean      :240.2
## 3rd Qu.: 1.10000   3rd Qu.: 1.0000   3rd Qu.: 1.3000   3rd Qu.:338.0
## Max.      :12.00000   Max.      :15.6000   Max.      :12.4000   Max.      :824.0
##
##          FGA          FG%          3P          3PA
## Min.      : 0.0   Min.      :0.0000   Min.      : 0.00   Min.      : 0.00
## 1st Qu.: 256.8   1st Qu.:0.4110   1st Qu.: 4.00   1st Qu.:16.25
## Median : 459.0   Median :0.4450   Median :42.50   Median :120.00
## Mean      : 526.0   Mean      :0.4527   Mean      :56.36   Mean      :157.22
## 3rd Qu.: 730.8   3rd Qu.:0.4880   3rd Qu.:90.75   3rd Qu.:256.00
## Max.      :1941.0   Max.      :1.0000   Max.      :402.00   Max.      :886.00
##
##          3P%          2P          2PA          2P%
## Min.      :0.0000   Min.      : 0.00   Min.      : 0.0   Min.      :0.0000
## 1st Qu.:0.2500   1st Qu.:74.25   1st Qu.:157.0   1st Qu.:0.4522
## Median :0.3330   Median :153.00   Median :308.0   Median :0.4850
## Mean      :0.2847   Mean      :183.82   Mean      :368.8   Mean      :0.4881
## 3rd Qu.:0.3738   3rd Qu.:258.00   3rd Qu.:513.0   3rd Qu.:0.5308
## Max.      :1.0000   Max.      :730.00   Max.      :1421.0   Max.      :1.0000
##

```

```
##          eFG%          FT          FTA          FT%
## Min.      :0.0000   Min.      : 0.0   Min.      : 0.0   Min.      :0.0000
## 1st Qu.:0.4730   1st Qu.: 38.0   1st Qu.: 51.0   1st Qu.:0.6943
## Median :0.5060   Median : 79.5   Median :109.5   Median :0.7685
## Mean      :0.5038   Mean      :111.4   Mean      :145.1   Mean      :0.7438
## 3rd Qu.:0.5370   3rd Qu.:145.0   3rd Qu.:194.0   3rd Qu.:0.8310
## Max.      :1.0000   Max.      :746.0   Max.      :881.0   Max.      :1.0000
##
##          ORB          DRB          TRB          AST
## Min.      : 0.00   Min.      : 0.0   Min.      : 0.0   Min.      : 0.0
## 1st Qu.: 21.00   1st Qu.:103.0   1st Qu.: 128.0   1st Qu.: 47.0
## Median : 44.00   Median :180.0   Median : 229.0   Median : 97.0
## Mean      : 63.27   Mean      :206.7   Mean      : 269.9   Mean      :137.1
## 3rd Qu.: 86.00   3rd Qu.:278.2   3rd Qu.: 364.2   3rd Qu.:176.0
## Max.      :395.00   Max.      :817.0   Max.      :1198.0   Max.      :906.0
##
##          STL          BLK          TOV          PF
## Min.      : 0.00   Min.      : 0.00   Min.      : 0.0   Min.      : 0.0
## 1st Qu.: 22.00   1st Qu.: 9.00   1st Qu.: 39.0   1st Qu.: 79.0
## Median : 42.00   Median : 20.00   Median : 69.0   Median :125.0
## Mean      : 47.36   Mean      : 30.19   Mean      : 83.2   Mean      :121.6
## 3rd Qu.: 65.75   3rd Qu.: 39.00   3rd Qu.:113.8   3rd Qu.:165.0
## Max.      :169.00   Max.      :269.00   Max.      :464.0   Max.      :278.0
##
##          PTS          out          ovr          ins
## Min.      : 0.0   Min.      :30.00   Min.      :61.00   Min.      :44.00
## 1st Qu.: 307.0   1st Qu.:61.00   1st Qu.:71.00   1st Qu.:58.00
## Median : 543.5   Median :72.00   Median :76.00   Median :64.00
## Mean      : 648.1   Mean      :71.18   Mean      :78.48   Mean      :65.33
## 3rd Qu.: 897.0   3rd Qu.:82.75   3rd Qu.:85.00   3rd Qu.:71.00
## Max.      :2558.0   Max.      :99.00   Max.      :99.00   Max.      :97.00
##
##          pla          ath          def          reb
## Min.      :28.00   Min.      :49.00   Min.      :43.00   Min.      :27.00
## 1st Qu.:47.00   1st Qu.:68.00   1st Qu.:58.00   1st Qu.:44.00
## Median :59.00   Median :73.00   Median :64.00   Median :59.00
## Mean      :61.25   Mean      :73.47   Mean      :65.47   Mean      :60.85
## 3rd Qu.:75.75   3rd Qu.:79.00   3rd Qu.:72.00   3rd Qu.:74.00
## Max.      :98.00   Max.      :98.00   Max.      :98.00   Max.      :98.00
##
```

```
# Output final complete, primary, and secondary datasets
write.csv(df_final,'../data/pooled/complete.csv',row.names=F)
write.csv(df_p_final,'../data/pooled/primary.csv',row.names=F)
write.csv(df_s_final,'../data/pooled/secondary.csv',row.names=F)
# preview datasets
head(df_p_final)
```

```
##          name year salary Pos Age Tm G GS MP PER TS% 3Par FTr ORB%
## 1 aaron brooks 2016 2700000 PG 31 CHI 69 0 1108 11.8 0.494 0.394 0.136 2.0
## 2 aaron brooks 2017 2116955 PG 32 IND 65 0 894 9.5 0.507 0.427 0.133 2.3
## 3 aaron gordon 2016 4351320 PF 20 ORL 78 37 1863 17.0 0.541 0.245 0.333 9.0
## 4 aaron gordon 2017 5504420 SF 21 ORL 80 72 2298 14.4 0.530 0.309 0.251 5.3
## 5 adreian payne 2016 2022240 PF 24 MIN 52 2 486 5.6 0.422 0.221 0.179 4.8
## 6 aj hammons 2017 1312611 C 24 DAL 22 0 163 8.4 0.472 0.238 0.476 5.4
## DRB% TRB% AST% STL% BLK% TOV% USG% OWS DWS WS WS/48 OBPM DBPM BPM VORP
## 1 7.5 4.8 26.0 1.4 0.7 14.2 22.9 0.2 0.7 0.9 0.040 -0.5 -2.8 -3.3 -0.4
## 2 6.3 4.3 20.7 1.4 0.9 17.2 19.2 -0.2 0.5 0.3 0.016 -2.1 -2.6 -4.6 -0.6
## 3 21.3 15.1 10.3 1.6 2.4 9.0 17.3 3.2 2.2 5.4 0.139 0.6 1.2 1.8 1.8
## 4 14.1 9.6 10.5 1.4 1.4 8.5 20.1 2.0 1.7 3.7 0.076 -0.2 -0.4 -0.7 0.8
## 5 21.5 13.3 8.9 1.7 1.8 18.7 17.7 -0.9 0.4 -0.5 -0.047 -5.9 -0.2 -6.1 -0.5
## 6 20.9 12.8 3.8 0.3 7.2 16.4 17.6 -0.2 0.2 0.0 -0.001 -7.5 1.9 -5.6 -0.1
## FG FGA FG% 3P 3PA 3P% 2P 2PA 2P% eFG% FT FTA FT% ORB DRB TRB AST
## 1 188 469 0.401 66 185 0.357 122 284 0.430 0.471 49 64 0.766 21 80 101 180
```

```
## 2 121 300 0.403 48 128 0.375 73 172 0.424 0.483 32 40 0.800 18 51 69 125
## 3 274 579 0.473 42 142 0.296 232 437 0.531 0.509 129 193 0.668 154 353 507 128
## 4 393 865 0.454 77 267 0.288 316 598 0.528 0.499 156 217 0.719 116 289 405 150
## 5 53 145 0.366 9 32 0.281 44 113 0.389 0.397 17 26 0.654 20 91 111 29
## 6 17 42 0.405 5 10 0.500 12 32 0.375 0.464 9 20 0.450 8 28 36 4
## STL BLK TOV PF PTS
## 1 30 10 82 132 491
## 2 25 9 66 93 322
## 3 59 55 66 153 719
## 4 64 40 89 172 1019
## 5 16 11 36 77 132
## 6 1 13 10 21 48
```

```
head(df_s_final)
```

```
##           name ovr out ins pla ath def reb
## 1 aaron brooks 75 79 52 74 77 52 36
## 2 aaron brooks 85 87 51 81 82 57 37
## 3 aaron gordon 90 87 91 69 86 69 87
## 4 aaron gordon 92 86 91 49 86 75 94
## 5 adreian payne 69 56 65 43 66 64 68
## 6 aj hammons 66 47 64 40 58 57 71
```

```
head(df_final)
```

```
##           name year salary Pos Age Tm G GS MP PER TS% 3PAr FTr ORB%
## 1 aaron brooks 2016 2700000 PG 31 CHI 69 0 1108 11.8 0.494 0.394 0.136 2.0
## 2 aaron brooks 2017 2116955 PG 32 IND 65 0 894 9.5 0.507 0.427 0.133 2.3
## 3 aaron gordon 2016 4351320 PF 20 ORL 78 37 1863 17.0 0.541 0.245 0.333 9.0
## 4 aaron gordon 2017 5504420 SF 21 ORL 80 72 2298 14.4 0.530 0.309 0.251 5.3
## 5 adreian payne 2016 2022240 PF 24 MIN 52 2 486 5.6 0.422 0.221 0.179 4.8
## 6 aj hammons 2017 1312611 C 24 DAL 22 0 163 8.4 0.472 0.238 0.476 5.4
## DRB% TRB% AST% STL% BLK% TOV% USG% OWS DWS WS WS/48 OBPM DBPM BPM VORP
## 1 7.5 4.8 26.0 1.4 0.7 14.2 22.9 0.2 0.7 0.9 0.040 -0.5 -2.8 -3.3 -0.4
## 2 6.3 4.3 20.7 1.4 0.9 17.2 19.2 -0.2 0.5 0.3 0.016 -2.1 -2.6 -4.6 -0.6
## 3 21.3 15.1 10.3 1.6 2.4 9.0 17.3 3.2 2.2 5.4 0.139 0.6 1.2 1.8 1.8
## 4 14.1 9.6 10.5 1.4 1.4 8.5 20.1 2.0 1.7 3.7 0.076 -0.2 -0.4 -0.7 0.8
## 5 21.5 13.3 8.9 1.7 1.8 18.7 17.7 -0.9 0.4 -0.5 -0.047 -5.9 -0.2 -6.1 -0.5
## 6 20.9 12.8 3.8 0.3 7.2 16.4 17.6 -0.2 0.2 0.0 -0.001 -7.5 1.9 -5.6 -0.1
## FG FGA FG% 3P 3PA 3P% 2P 2PA 2P% eFG% FT FTA FT% ORB DRB TRB AST
## 1 188 469 0.401 66 185 0.357 122 284 0.430 0.471 49 64 0.766 21 80 101 180
## 2 121 300 0.403 48 128 0.375 73 172 0.424 0.483 32 40 0.800 18 51 69 125
## 3 274 579 0.473 42 142 0.296 232 437 0.531 0.509 129 193 0.668 154 353 507 128
## 4 393 865 0.454 77 267 0.288 316 598 0.528 0.499 156 217 0.719 116 289 405 150
## 5 53 145 0.366 9 32 0.281 44 113 0.389 0.397 17 26 0.654 20 91 111 29
## 6 17 42 0.405 5 10 0.500 12 32 0.375 0.464 9 20 0.450 8 28 36 4
## STL BLK TOV PF PTS out ovr ins pla ath def reb
## 1 30 10 82 132 491 79 75 52 74 77 52 36
## 2 25 9 66 93 322 87 85 51 81 82 57 37
## 3 59 55 66 153 719 87 90 91 69 86 69 87
## 4 64 40 89 172 1019 86 92 91 49 86 75 94
## 5 16 11 36 77 132 56 69 65 43 66 64 68
## 6 1 13 10 21 48 47 66 64 40 58 57 71
```

Explore Data

Summarize Datasets

```
# primary dataset
str(df_p_final)
```

```
## 'data.frame': 734 obs. of 51 variables:
## $ name : chr "aaron brooks" "aaron brooks" "aaron gordon" "aaron gordon" ...
```

```
## $ year : Factor w/ 2 levels "2016","2017": 1 2 1 2 1 2 1 2 1 2 ...
## $ salary: num 2700000 2116955 4351320 5504420 2022240 ...
## $ Pos : Factor w/ 5 levels "C","PF","PG",...: 3 3 2 4 2 1 4 4 1 1 ...
## $ Age : num 31 32 20 21 24 24 25 26 29 30 ...
## $ Tm : Factor w/ 31 levels "ATL","BOS","BRK",...: 4 12 22 22 18 7 25 25 1 2 ...
## $ G : num 69 65 78 80 52 22 82 61 82 68 ...
## $ GS : num 0 0 37 72 2 0 82 25 82 68 ...
## $ MP : num 1108 894 1863 2298 486 ...
## $ PER : num 11.8 9.5 17 14.4 5.6 8.4 12.7 11.3 19.4 17.7 ...
## $ TS% : num 0.494 0.507 0.541 0.53 0.422 0.472 0.533 0.506 0.565 0.553 ...
## $ 3PAr : num 0.394 0.427 0.245 0.309 0.221 0.238 0.485 0.455 0.244 0.302 ...
## $ FTr : num 0.136 0.133 0.333 0.251 0.179 0.476 0.217 0.292 0.123 0.169 ...
## $ ORB% : num 2 2.3 9 5.3 4.8 5.4 4.5 4.8 6.3 4.9 ...
## $ DRB% : num 7.5 6.3 21.3 14.1 21.5 20.9 18.6 23.5 18.2 18.6 ...
## $ TRB% : num 4.8 4.3 15.1 9.6 13.3 12.8 11.5 14.1 12.4 11.8 ...
## $ AST% : num 26 20.7 10.3 10.5 8.9 3.8 8.8 7.9 16.7 24.4 ...
## $ STL% : num 1.4 1.4 1.6 1.4 1.7 0.3 1.5 1.7 1.3 1.2 ...
## $ BLK% : num 0.7 0.9 2.4 1.4 1.8 7.2 1.8 2 3.6 3.3 ...
## $ TOV% : num 14.2 17.2 9 8.5 18.7 16.4 13.2 15.2 8.8 11.9 ...
## $ USG% : num 22.9 19.2 17.3 20.1 17.7 17.6 16.9 15.4 20.6 19.8 ...
## $ OWS : num 0.2 -0.2 3.2 2 -0.9 -0.2 1.7 -0.1 4.9 3.6 ...
## $ DWS : num 0.7 0.5 2.2 1.7 0.4 0.2 2.3 2 4.5 2.7 ...
## $ WS : num 0.9 0.3 5.4 3.7 -0.5 0 4 1.9 9.4 6.3 ...
## $ WS/48 : num 0.04 0.016 0.139 0.076 -0.047 -0.001 0.082 0.051 0.172 0.137 ...
## $ OBPM : num -0.5 -2.1 0.6 -0.2 -5.9 -7.5 -0.4 -2.3 1.5 1 ...
## $ DBPM : num -2.8 -2.6 1.2 -0.4 -0.2 1.9 0.7 1.2 2.6 2.1 ...
## $ BPM : num -3.3 -4.6 1.8 -0.7 -6.1 -5.6 0.2 -1.1 4.1 3.1 ...
## $ VORP : num -0.4 -0.6 1.8 0.8 -0.5 -0.1 1.3 0.4 4.1 2.8 ...
## $ FG : num 188 121 274 393 53 17 299 183 529 379 ...
## $ FGA : num 469 300 579 865 145 ...
## $ FG% : num 0.401 0.403 0.473 0.454 0.366 0.405 0.416 0.393 0.505 0.473 ...
## $ 3P : num 66 48 42 77 9 5 126 70 88 86 ...
## $ 3PA : num 185 128 142 267 32 10 349 212 256 242 ...
## $ 3P% : num 0.357 0.375 0.296 0.288 0.281 0.5 0.361 0.33 0.344 0.355 ...
## $ 2P : num 122 73 232 316 44 12 173 113 441 293 ...
## $ 2PA : num 284 172 437 598 113 32 370 254 792 559 ...
## $ 2P% : num 0.43 0.424 0.531 0.528 0.389 0.375 0.468 0.445 0.557 0.524 ...
## $ eFG% : num 0.471 0.483 0.509 0.499 0.397 0.464 0.503 0.468 0.547 0.527 ...
## $ FT : num 49 32 129 156 17 9 115 96 103 108 ...
## $ FTA : num 64 40 193 217 26 20 156 136 129 135 ...
## $ FT% : num 0.766 0.8 0.668 0.719 0.654 0.45 0.737 0.706 0.798 0.8 ...
## $ ORB : num 21 18 154 116 20 8 98 77 148 95 ...
## $ DRB : num 80 51 353 289 91 28 401 374 448 369 ...
## $ TRB : num 101 69 507 405 111 36 499 451 596 464 ...
## $ AST : num 180 125 128 150 29 4 138 99 263 337 ...
## $ STL : num 30 25 59 64 16 1 72 60 68 52 ...
## $ BLK : num 10 9 55 40 11 13 53 44 121 87 ...
## $ TOV : num 82 66 66 89 36 10 120 94 107 116 ...
## $ PF : num 132 93 153 172 77 21 171 102 163 138 ...
## $ PTS : num 491 322 719 1019 132 ...
```

```
# secondary dataset
```

```
str(df_s_final)
```

```
## 'data.frame': 734 obs. of 8 variables:
## $ name: chr "aaron brooks" "aaron brooks" "aaron gordon" "aaron gordon" ...
## $ ovr : int 75 85 90 92 69 66 91 83 83 91 ...
## $ out : int 79 87 87 86 56 47 90 75 81 80 ...
## $ ins : int 52 51 91 91 65 64 77 72 76 82 ...
## $ pla : int 74 81 69 49 43 40 60 59 58 82 ...
## $ ath : int 77 82 86 86 66 58 81 75 75 77 ...
## $ def : int 52 57 69 75 64 57 76 66 70 80 ...
## $ reb : int 36 37 87 94 68 71 94 65 73 87 ...
```

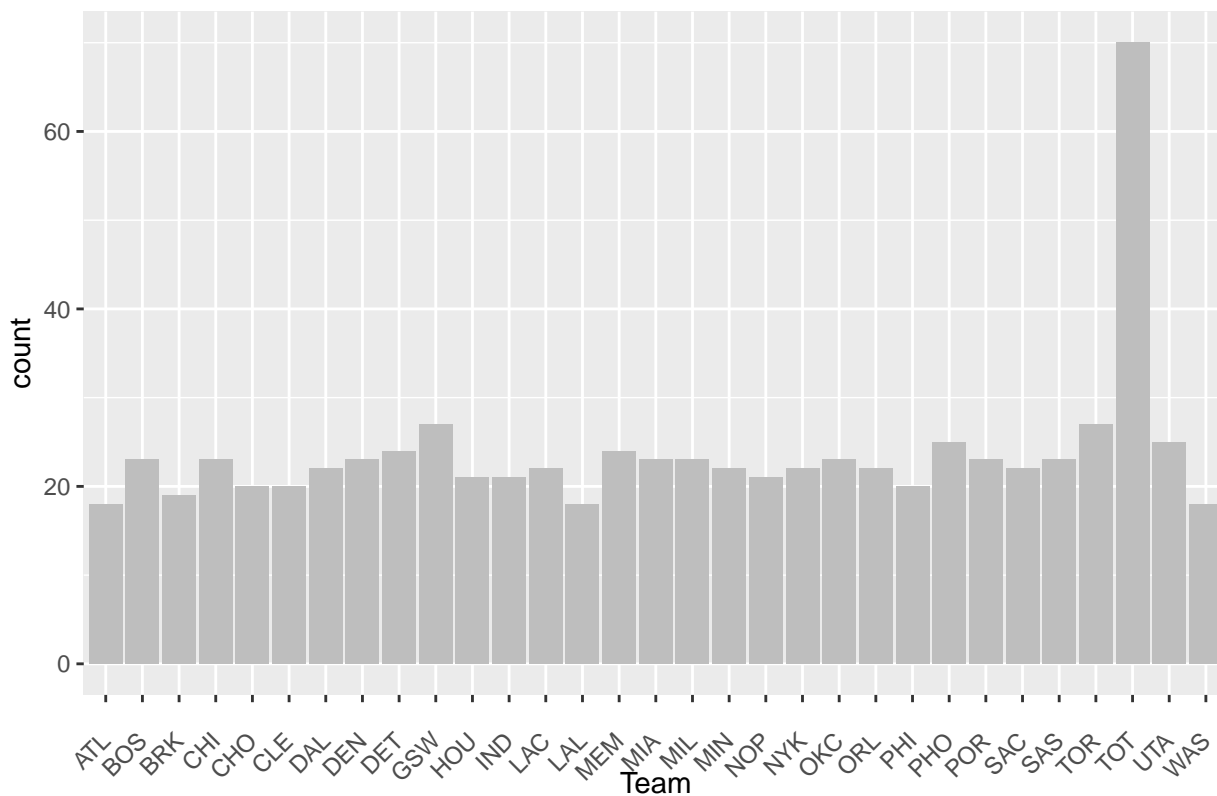
Complete Dataset Histograms

```
library(purrr)
library(tidyr)
library(ggplot2)
df_final %>%
  keep(is.numeric) %>%
  gather() %>%
  ggplot(aes(value)) +
    facet_wrap(~ key, scales = "free") +
    geom_histogram(aes(y=..density..), fill = "grey") +
    geom_density()
ggsave("../figures/hist_complete_vars.png", width=15, height=13)
```

Bar Chart of Player by Team from Complete Dataset

```
library(ggplot2)
ggplot(df_final, aes(x = Tm)) +
  geom_bar(fill = "grey") +
  labs(x = "Team", title = "Players per team") +
  theme(axis.text.x=element_text(angle=45,hjust=1,vjust=0.5))
```

Players per team

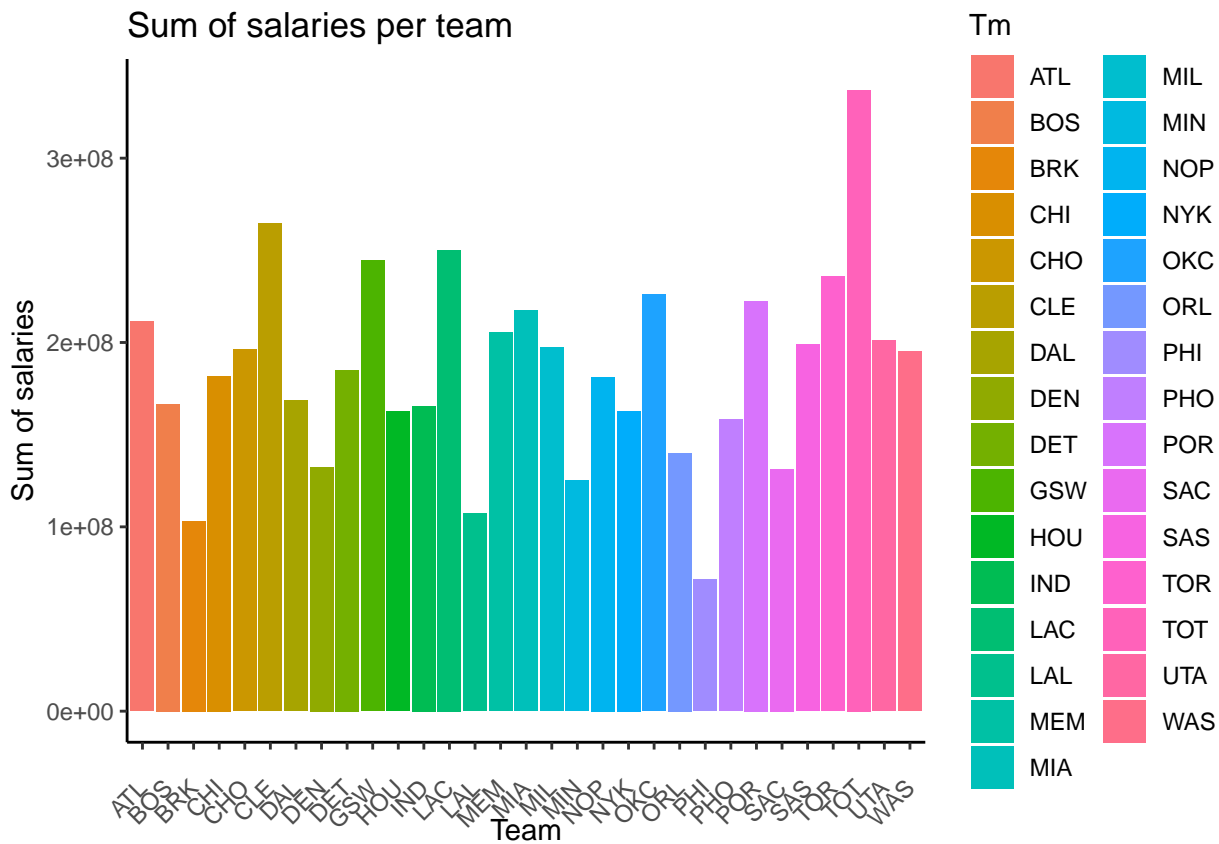


```
ggsave("../figures/bar_complete_player_per_team.png", width=10, height=7)
```

Sum of Salaries per Team for Complete Dataset

```
library(ggplot2)
library(tidyr)
library(dplyr)
df_final %>%
  group_by(Tm) %>%
  summarise(sum_salary = sum(salary)) %>%
  ggplot(aes(x = Tm, y = sum_salary, fill = Tm)) +
```

```
geom_bar(stat = "identity") +
theme_classic() +
labs(
  x = "Team",
  y = "Sum of salaries",
  title = paste("Sum of salaries per team")) +
theme(axis.text.x=element_text(angle=45,hjust=1,vjust=0.5))
```

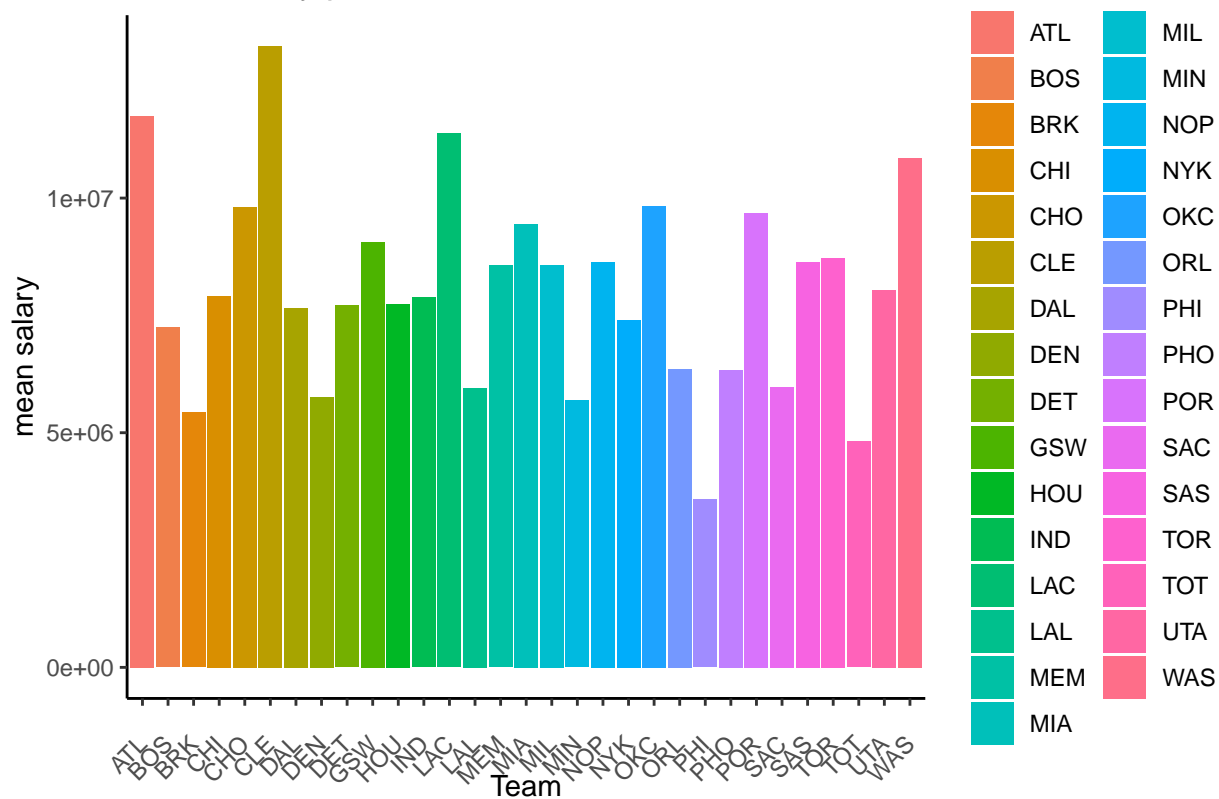


```
ggsave("../figures/bar_complete_sum_salaries_per_team.png", width=10, height=7)
```

Mean Salaries per Team for Complete Dataset

```
library(ggplot2)
library(tidyr)
library(dplyr)
df_final %>%
  group_by(Tm) %>%
  summarise(mean_salary = mean(salary)) %>%
  ggplot(aes(x = Tm, y = mean_salary, fill = Tm)) +
  geom_bar(stat = "identity") +
  theme_classic() +
  labs(
    x = "Team",
    y = "mean salary",
    title = paste(
      "mean salary per team")) +
  theme(axis.text.x=element_text(angle=45,hjust=1,vjust=0.5))
```

mean salary per team

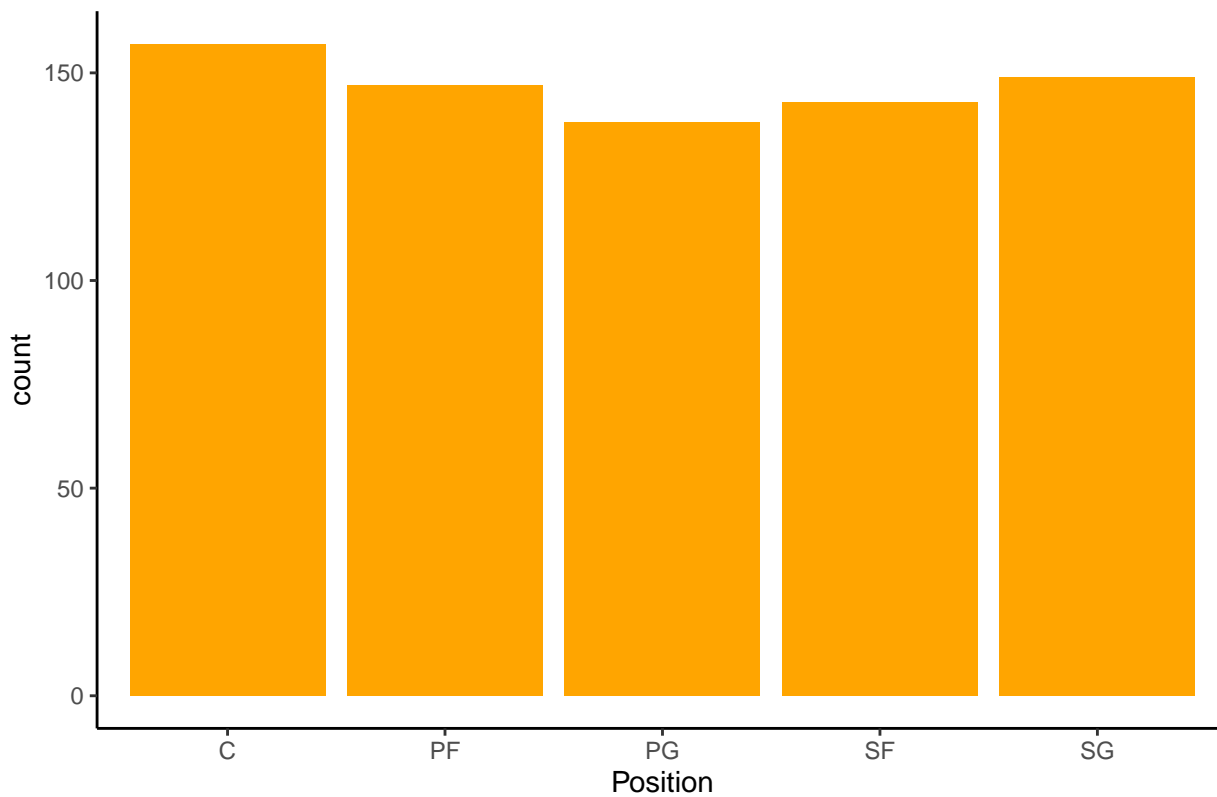


```
ggsave("../figures/bar_complete_mean_salaries_per_team.png", width=10, height=7)
```

Players in each position

```
library(ggplot2)
ggplot(df_final, aes(x = Pos)) +
  geom_bar(fill = "orange") +
  labs(x = "Position", title = "No of Player for each position") +
  theme_classic()
```

No of Player for each position

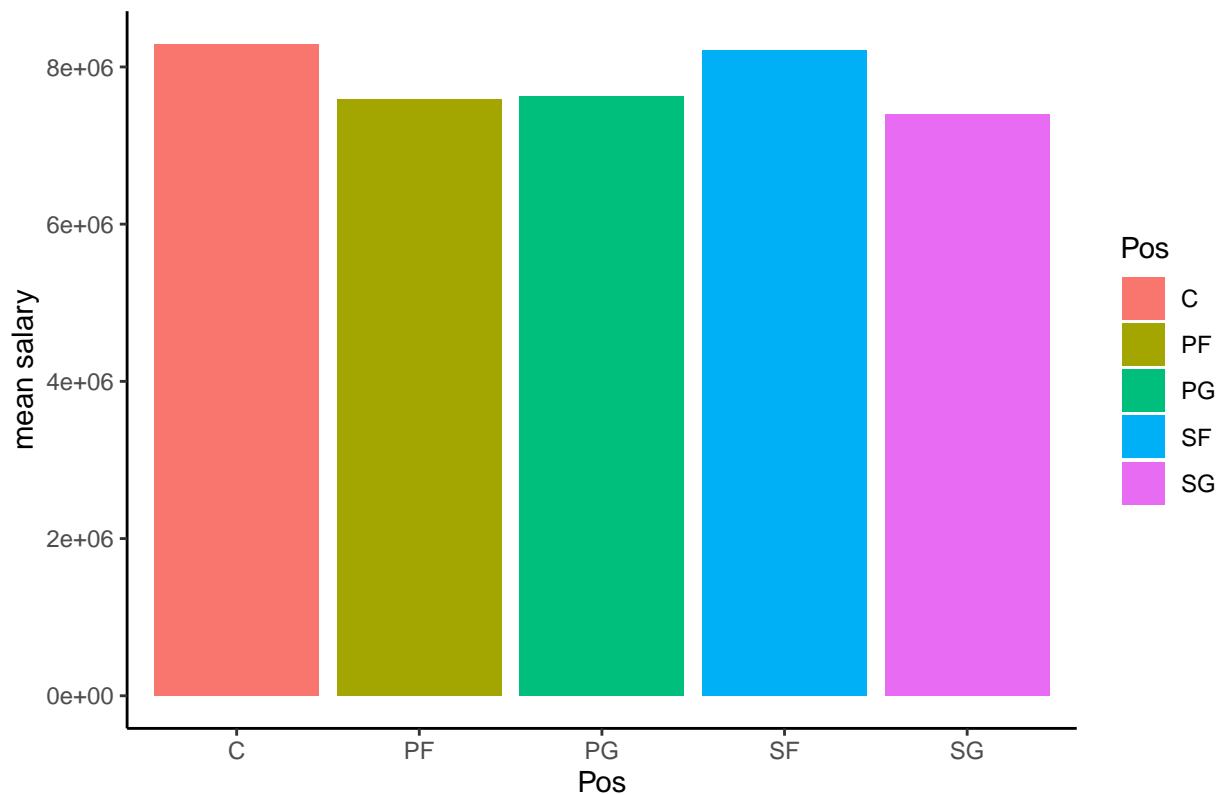


```
ggsave("../figures/bar_complete_player_Position.png", width=10, height=7)
```

Mean salaries for each position

```
library(ggplot2)
library(tidyr)
library(dplyr)
df_final %>%
  group_by(Pos) %>%
  summarise(mean_salary = mean(salary)) %>%
  ggplot(aes(x = Pos, y = mean_salary, fill = Pos)) +
  geom_bar(stat = "identity") +
  theme_classic() +
  labs(
    x = "Pos",
    y = "mean salary",
    title = paste(
      "mean salary for Position")
  )
```


mean salary for Position



```
ggsave("../figures/bar_complete_mean_salaries_for_Position.png", width=10, height=7)
```

Correlation Matrix for complete dataset

```
corr_matrix_c <- cor(Filter(is.numeric,df_final[2:ncol(df_final)]),method = "pearson")
correlation_salary_c <- sort(corr_matrix_c[, 'salary'],decreasing = TRUE)
correlation_salary_c
```

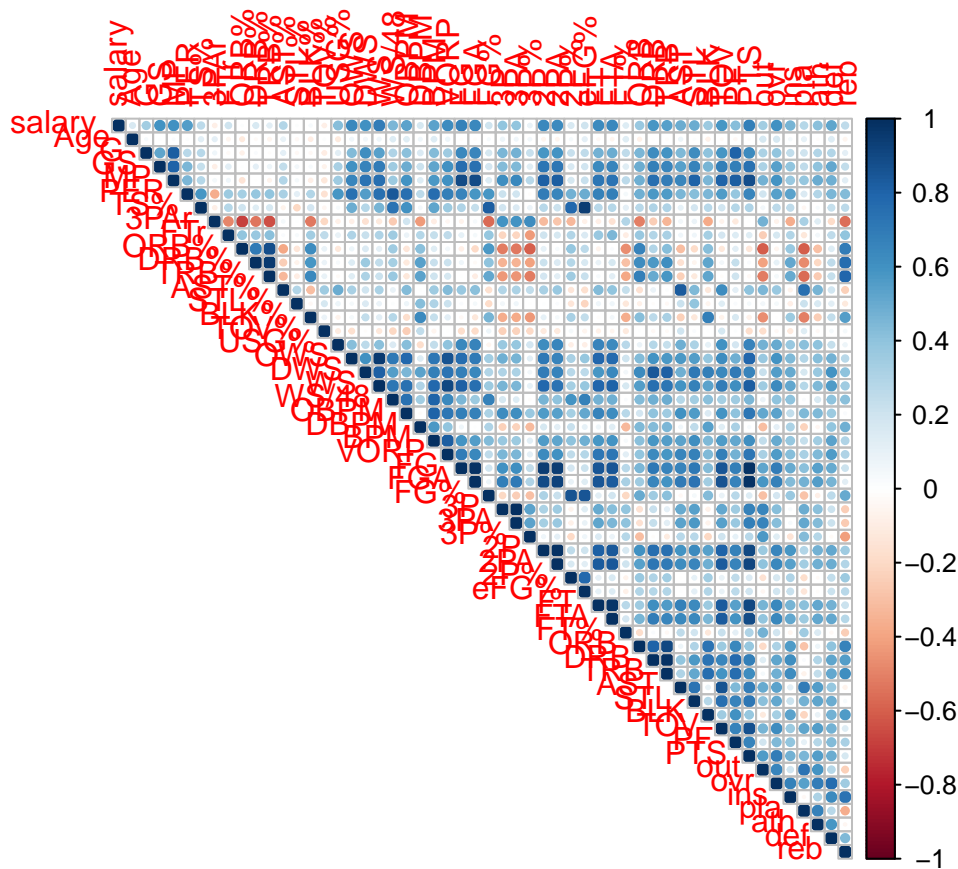
| ## | salary | WS | PTS | FG | FGA | 2P |
|----|------------|------------|------------|------------|-------------|-------------|
| ## | 1.00000000 | 0.69665645 | 0.68170454 | 0.67897723 | 0.64990139 | 0.64984231 |
| ## | FTA | OWS | FT | 2PA | VORP | ovr |
| ## | 0.64858045 | 0.64772438 | 0.63666066 | 0.63574577 | 0.62415826 | 0.60432440 |
| ## | MP | DWS | GS | TOV | DRB | PER |
| ## | 0.60407111 | 0.60108603 | 0.59325726 | 0.58481090 | 0.58151748 | 0.55121387 |
| ## | TRB | BPM | OBPM | def | AST | STL |
| ## | 0.54226307 | 0.53975078 | 0.53647127 | 0.52260324 | 0.49614270 | 0.49000118 |
| ## | ins | WS/48 | USG% | PF | ath | 3P |
| ## | 0.46685385 | 0.45292509 | 0.42859771 | 0.42263528 | 0.41158151 | 0.39131386 |
| ## | 3PA | ORB | BLK | out | G | AST% |
| ## | 0.39000035 | 0.37269808 | 0.36909524 | 0.34625196 | 0.34620759 | 0.29725170 |
| ## | pla | TS% | reb | FTr | eFG% | FG% |
| ## | 0.28611363 | 0.26743221 | 0.25553148 | 0.20526496 | 0.20501001 | 0.20170219 |
| ## | DBPM | 2P% | Age | DRB% | FT% | TRB% |
| ## | 0.17092923 | 0.16971591 | 0.16958607 | 0.16788419 | 0.14395542 | 0.12124045 |
| ## | 3P% | BLK% | STL% | ORB% | 3PAr | TOV% |
| ## | 0.09261836 | 0.03617047 | 0.01753918 | 0.01613656 | -0.08847379 | -0.08924334 |

Correlation Plot for complete dataset

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
corrplot(corr_matrix_c,type = "upper")
```

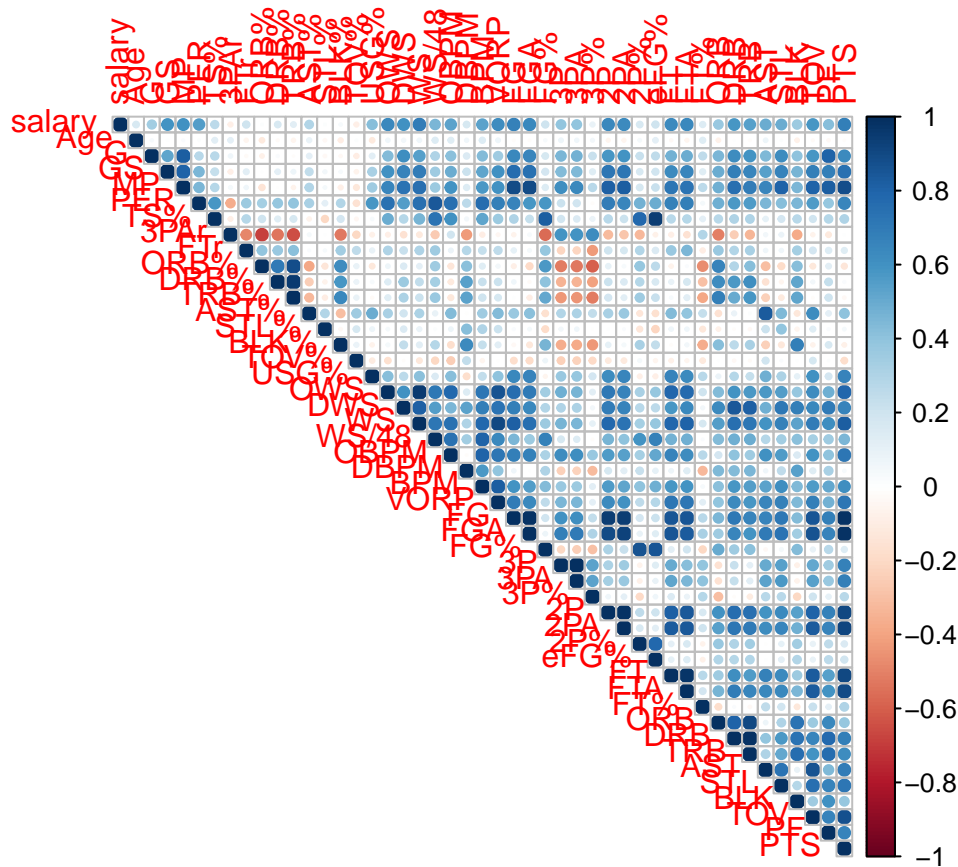


Correlation plot for Primary dataset

```
corr_matrix_p <- cor(Filter(is.numeric,df_p_final[2:ncol(df_p_final)]),method = "pearson")
correlation_salary_p <- sort(corr_matrix_p[, 'salary'],decreasing = TRUE)
correlation_salary_p
```

| | | | | | | |
|----|------------|------------|------------|-------------|-------------|------------|
| ## | salary | WS | PTS | FG | FGA | 2P |
| ## | 1.00000000 | 0.69665645 | 0.68170454 | 0.67897723 | 0.64990139 | 0.64984231 |
| ## | FTA | OWS | FT | 2PA | VORP | MP |
| ## | 0.64858045 | 0.64772438 | 0.63666066 | 0.63574577 | 0.62415826 | 0.60407111 |
| ## | DWS | GS | TOV | DRB | PER | TRB |
| ## | 0.60108603 | 0.59325726 | 0.58481090 | 0.58151748 | 0.55121387 | 0.54226307 |
| ## | BPM | OBPM | AST | STL | WS/48 | USG% |
| ## | 0.53975078 | 0.53647127 | 0.49614270 | 0.49000118 | 0.45292509 | 0.42859771 |
| ## | PF | 3P | 3PA | ORB | BLK | G |
| ## | 0.42263528 | 0.39131386 | 0.39000035 | 0.37269808 | 0.36909524 | 0.34620759 |
| ## | AST% | TS% | FT% | eFG% | FG% | DBPM |
| ## | 0.29725170 | 0.26743221 | 0.20526496 | 0.20501001 | 0.20170219 | 0.17092923 |
| ## | 2P% | Age | DRB% | FT% | TRB% | 3P% |
| ## | 0.16971591 | 0.16958607 | 0.16788419 | 0.14395542 | 0.12124045 | 0.09261836 |
| ## | BLK% | STL% | ORB% | 3PAr | TOV% | |
| ## | 0.03617047 | 0.01753918 | 0.01613656 | -0.08847379 | -0.08924334 | |

```
library(corrplot)
corrplot(corr_matrix_p,type = "upper")
```



Save correlation plots.

```
# complete dataset
png(file = "../figures/Correlation_plot_c.png")
corrplot(corr_matrix_c,type = "upper")
# primary dataset
png(file = "../figures/Correlation_plot_p.png")
corrplot(corr_matrix_p,type = "upper")
dev.off()
```

```
## pdf
## 2
```

Detecting Outliers

```
plot = function(variable)
{
  print(variable)
  ggplot(df_final,aes(x = df_final[,variable], y = salary)) + geom_point() + theme_classic() + labs(x=variable)
}
```

```
library(gridExtra)
```

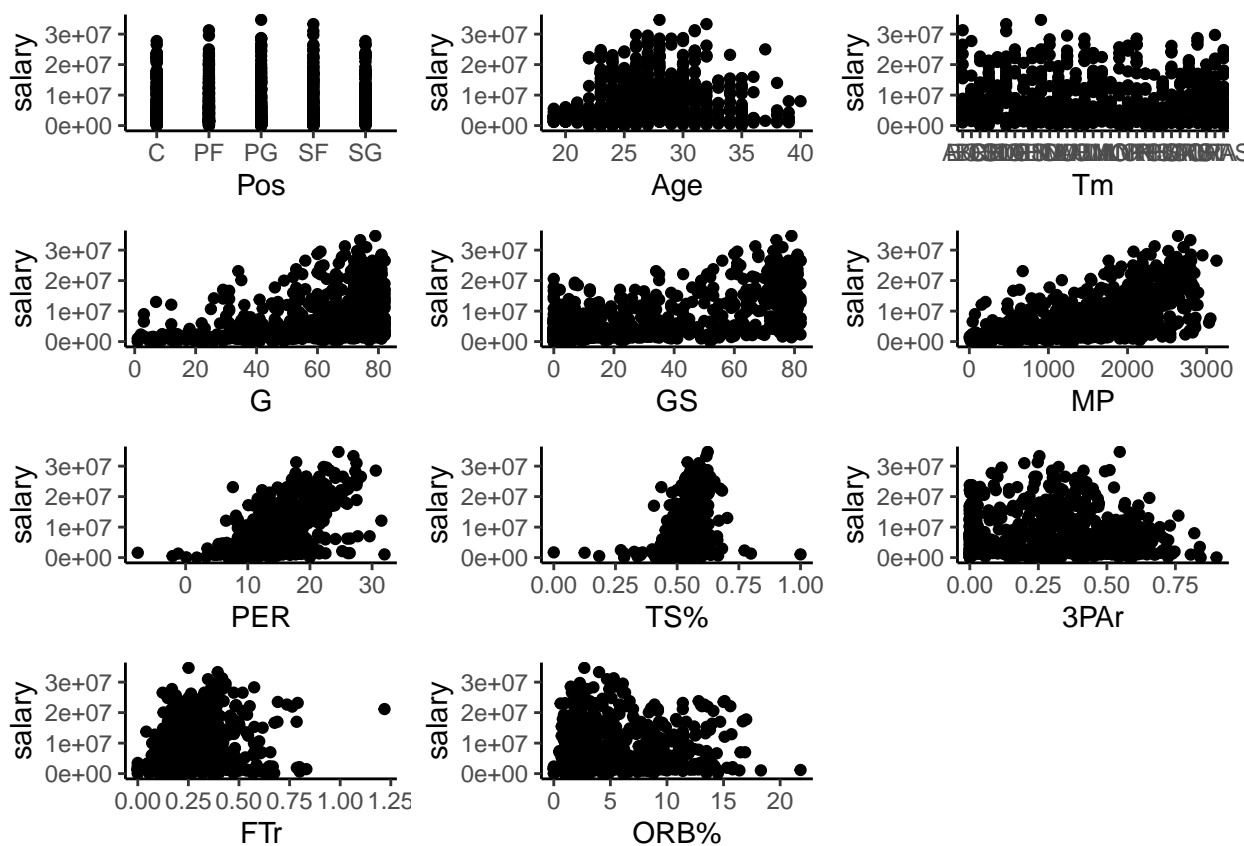
```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
## combine

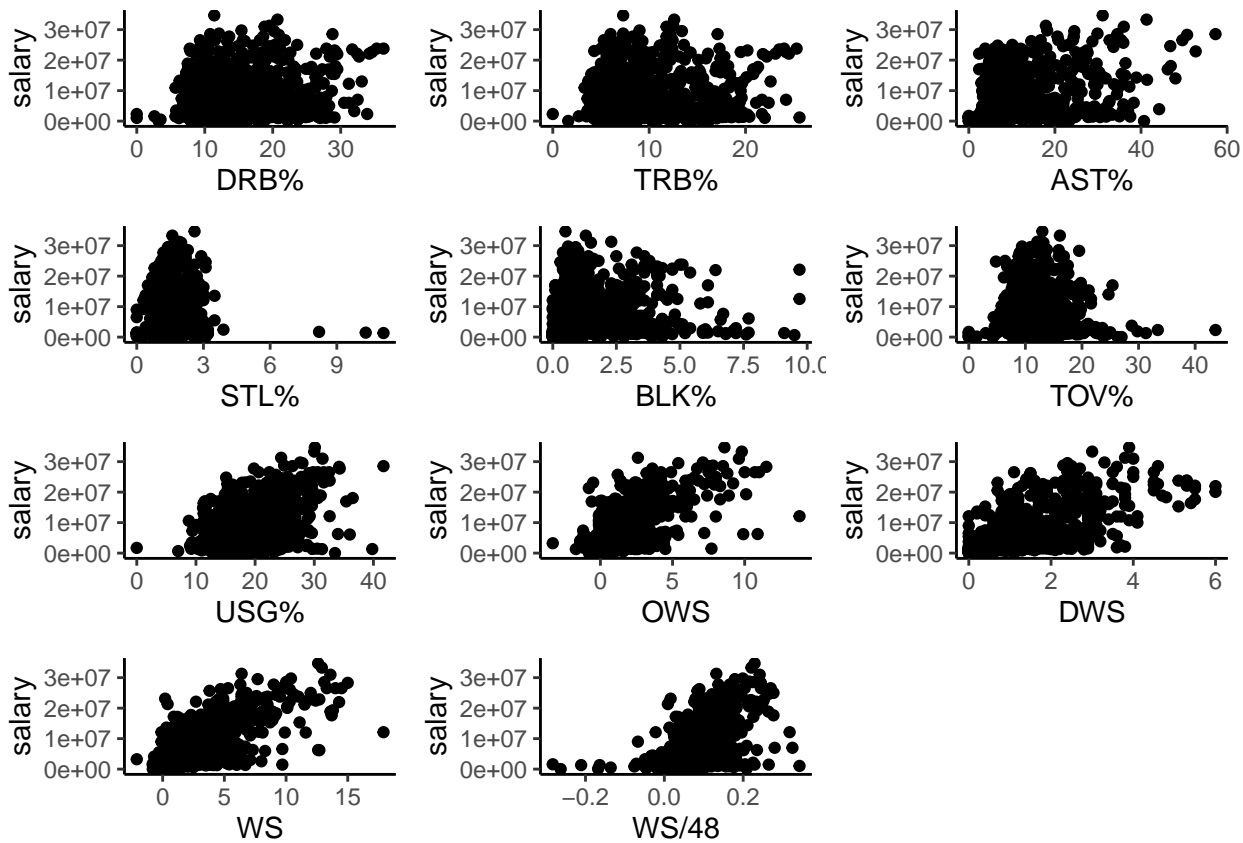
p = list()
p <- NULL
val <- 0
```

```
d <- df_final[,4:ncol(df_final)]
for(j in 1:5)
{
  for(i in 1:11) {
    name = names(d[i+val])
    p[[i]] = plot(as.character(name))
  }
  val = i+val
do.call(grid.arrange,p)
p <- NULL
}
```

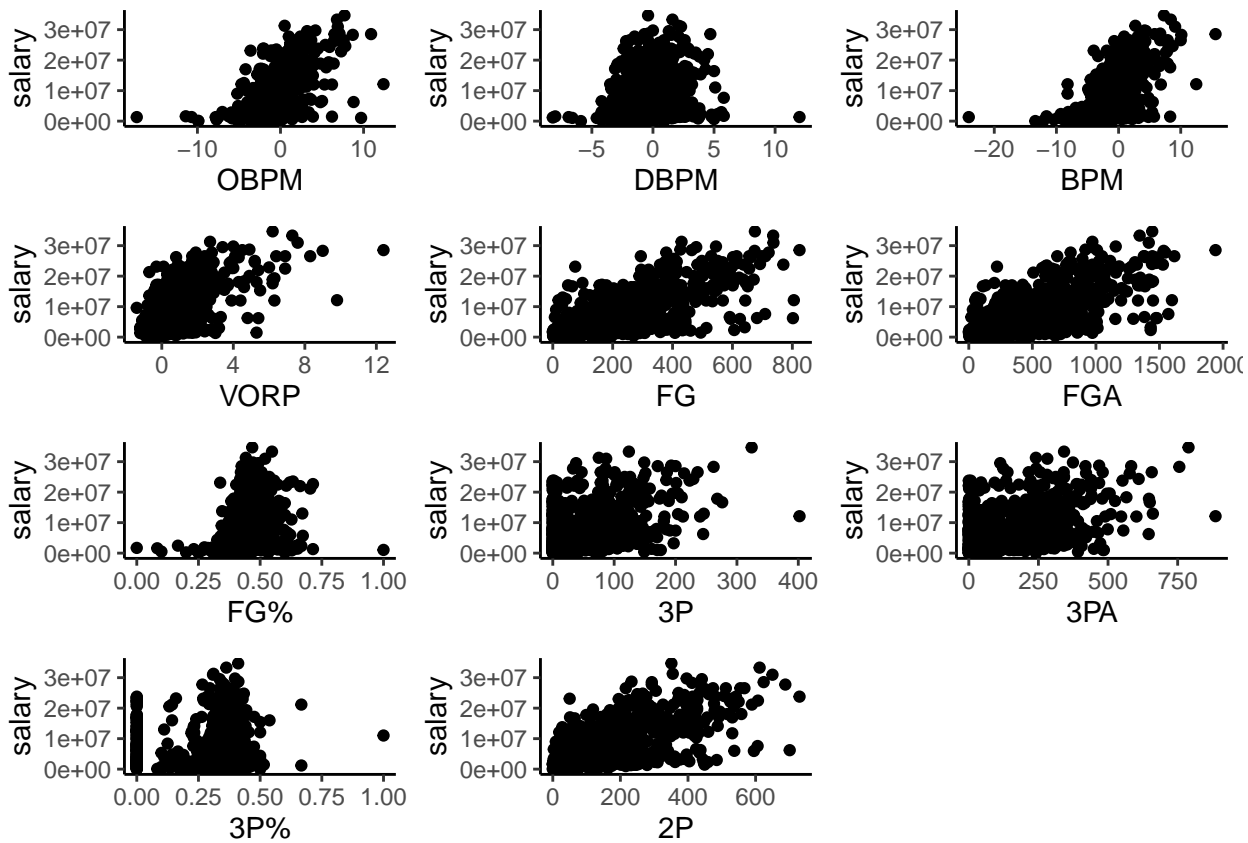
```
## [1] "Pos"
## [1] "Age"
## [1] "Tm"
## [1] "G"
## [1] "GS"
## [1] "MP"
## [1] "PER"
## [1] "TS%"
## [1] "3PAr"
## [1] "FTr"
## [1] "ORB%"
```



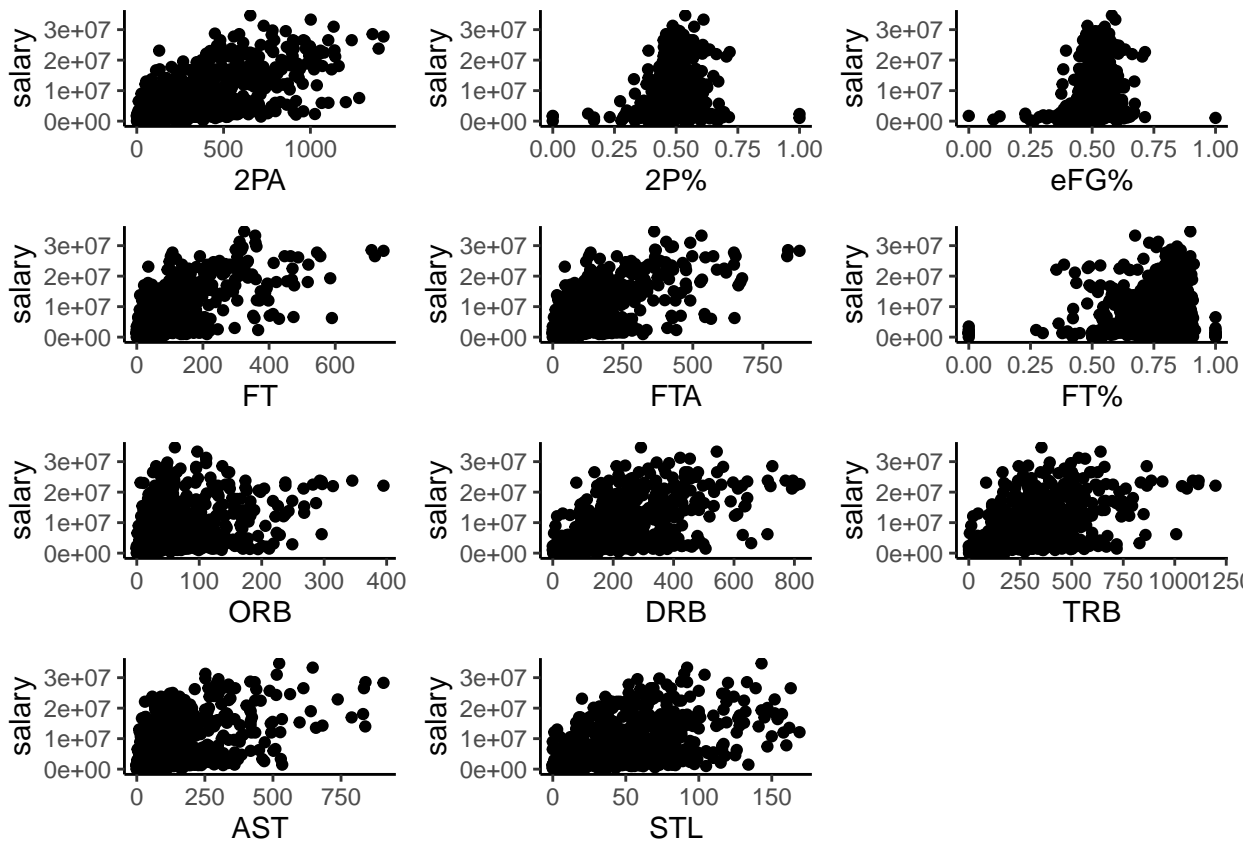
```
## [1] "DRB%"
## [1] "TRB%"
## [1] "AST%"
## [1] "STL%"
## [1] "BLK%"
## [1] "TOV%"
## [1] "USG%"
## [1] "OWS"
## [1] "DWS"
## [1] "WS"
## [1] "WS/48"
```



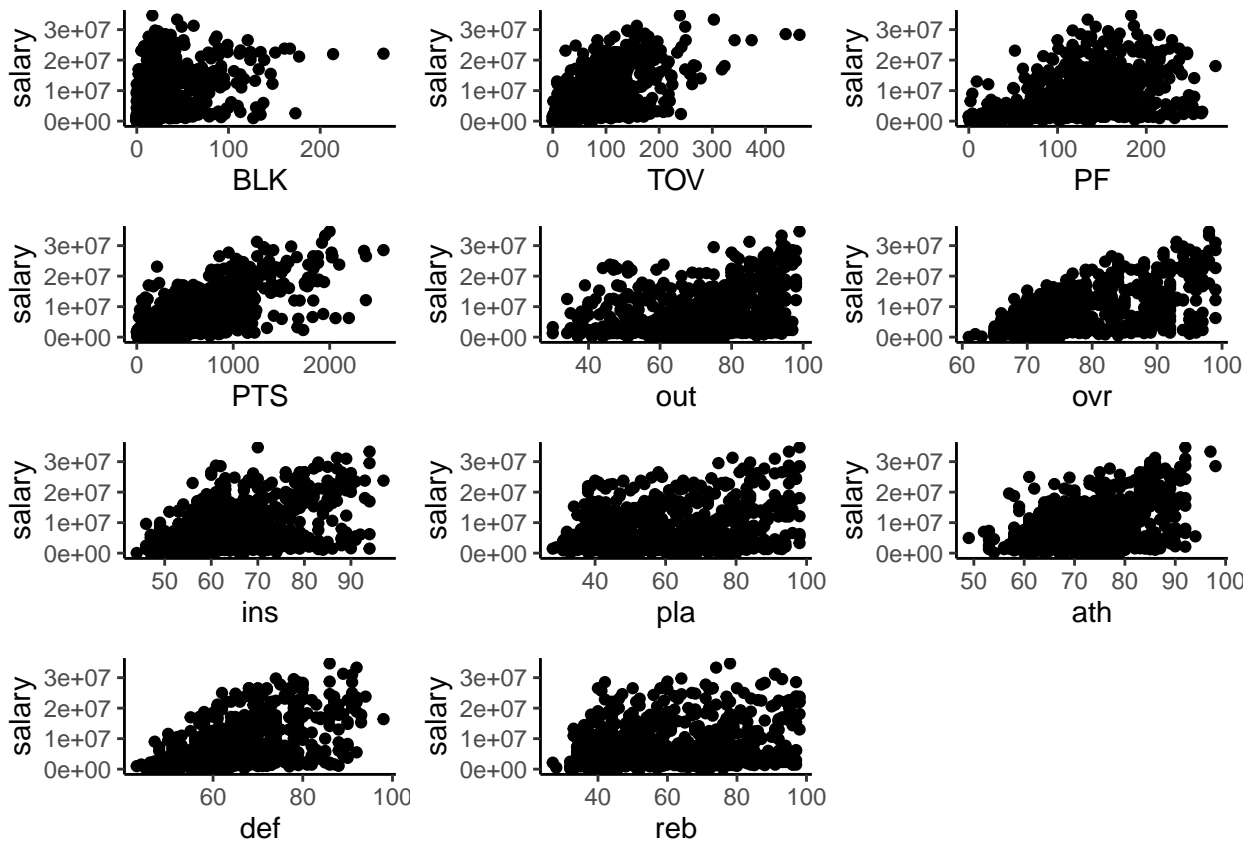
```
## [1] "OBPM"
## [1] "DBPM"
## [1] "BPM"
## [1] "VORP"
## [1] "FG"
## [1] "FGA"
## [1] "FG%"
## [1] "3P"
## [1] "3PA"
## [1] "3P%"
## [1] "2P"
```



```
## [1] "2PA"
## [1] "2P%"
## [1] "eFG%"
## [1] "FT"
## [1] "FTA"
## [1] "FT%"
## [1] "ORB"
## [1] "DRB"
## [1] "TRB"
## [1] "AST"
## [1] "STL"
```



```
## [1] "BLK"
## [1] "TOV"
## [1] "PF"
## [1] "PTS"
## [1] "out"
## [1] "ovr"
## [1] "ins"
## [1] "pla"
## [1] "ath"
## [1] "def"
## [1] "reb"
```



VARIABLE SELECTION

Helper Functions

```
get_salary_formula <- function(x_vars){
  return(as.formula(sprintf('salary ~ `%s`',paste(x_vars,collapse='` + `'))))}

```

Primary Dataset Variable Selection Using Automated F-Test-Based Backward Selection

```
library(rms)

## Loading required package: Hmisc
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:dplyr':
##
##   src, summarize
## The following object is masked from 'package:rvest':
##
##   html
## The following objects are masked from 'package:base':
##
##   format.pval, units
## Loading required package: SparseM

```



```
##
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':
##
##      backsolve

p_x_vars <- names(df_p_final)[!(names(df_p_final))%in%c('salary','name','2P','2PA','PTS','TRB')]
# 2P, 2PA, PTS, and TRB were causing singularity in predictor matrix, so they were dropped
p_formula <- get_salary_formula(p_x_vars)
p_formula
```

```
## salary ~ year + Pos + Age + Tm + G + GS + MP + PER + `TS` +
##      `3PA` + FTr + `ORB` + `DRB` + `TRB` + `AST` + `STL` +
##      `BLK` + `TOV` + `USG` + OWS + DWS + WS + `WS/48` + OBPM +
##      DBPM + BPM + VORP + FG + FGA + `FG` + `3P` + `3PA` + `3P` +
##      `2P` + `eFG` + FT + FTA + `FT` + ORB + DRB + AST + STL +
##      BLK + TOV + PF
## <environment: 0x55edd47d97d8>
```

```
p_selection_model <- ols(p_formula, data = df_p_final)
p_selection_model
```

```
## Linear Regression Model
##
##      ols(formula = p_formula, data = df_p_final)
##
##
##              Model Likelihood      Discrimination
##              Ratio Test              Indexes
## Obs              734      LR chi2      771.48      R2              0.650
## sigma4554312.9448      d.f.              77      R2 adj      0.609
## d.f.              656      Pr(> chi2) 0.0000      g      6422499.821
##
## Residuals
##
##      Min      1Q      Median      3Q      Max
## -14921310 -2699391 -225890  2596986 14905073
##
##
##      Coef      S.E.      t      Pr(>|t|)
## Intercept  9214166.5550  5544796.3006  1.66 0.0970
## year=2017  1446953.2357  464566.5387  3.11 0.0019
## Pos=PF     -419863.3840  680204.5900 -0.62 0.5373
## Pos=PG     -4142194.2617 1141749.4202 -3.63 0.0003
## Pos=SF     -929354.8944  913025.4517 -1.02 0.3091
## Pos=SG     -2319108.4079  995521.8816 -2.33 0.0201
## Age        214746.8625   46356.5201  4.63 <0.0001
## Tm=BOS     -1755692.1070 1530900.6289 -1.15 0.2519
## Tm=BRK     -1672476.4192 1826108.1461 -0.92 0.3601
## Tm=CHI     -1435775.3454 1599108.0198 -0.90 0.3696
## Tm=CHO     -1121286.8171 1626374.7318 -0.69 0.4908
## Tm=CLE     2011951.4170 1657897.6039  1.21 0.2254
## Tm=DAL     -452096.9019 1665070.3130 -0.27 0.7861
## Tm=DEN     -2285430.0846 1795602.7183 -1.27 0.2035
## Tm=DET     -1329591.8752 1675441.1485 -0.79 0.4277
## Tm=GSW     -1115910.2109 1553594.3395 -0.72 0.4728
## Tm=HOU     -1705158.6310 1739505.9995 -0.98 0.3273
## Tm=IND     -1977666.2135 1585568.1044 -1.25 0.2127
## Tm=LAC      966243.9607 1595703.0275  0.61 0.5450
## Tm=LAL     -399415.1886 2044572.1263 -0.20 0.8452
## Tm=MEM      512292.1746 1693690.2585  0.30 0.7624
## Tm=MIA     -1302523.1923 1567836.0016 -0.83 0.4064
## Tm=MIL      522625.3410 1723741.5914  0.30 0.7618
## Tm=MIN     -2153395.7770 1875319.7471 -1.15 0.2513
## Tm=NOP      732905.5603 1734514.9265  0.42 0.6728
## Tm=NYK     -1374276.3665 1756556.8104 -0.78 0.4343
```

```

## Tm=OKC      1053757.3280  1670618.6121  0.63 0.5284
## Tm=ORL      -807923.6520  1708700.6813 -0.47 0.6365
## Tm=PHI     -3530252.2850  1767380.0526 -2.00 0.0462
## Tm=PHO       38231.7585  1826307.9806  0.02 0.9833
## Tm=POR      2336575.3834  1750402.3841  1.33 0.1824
## Tm=SAC     -1229779.8996  1741639.8592 -0.71 0.4804
## Tm=SAS     -2719268.7329  1584189.3239 -1.72 0.0865
## Tm=TOR       272743.0838  1661348.2487  0.16 0.8696
## Tm=TOT     -2001328.9485  1423648.2302 -1.41 0.1603
## Tm=UTA     -1306977.2553  1558770.6600 -0.84 0.4021
## Tm=WAS       805652.6230  1682466.2477  0.48 0.6322
## G           -85878.9727   22558.5550 -3.81 0.0002
## GS          21880.4935   11615.8265  1.88 0.0601
## MP          2565.8600    1772.3533  1.45 0.1482
## PER        -20969.8371   432167.2140 -0.05 0.9613
## TS%        -6669563.4611 21665362.9662 -0.31 0.7583
## 3PAr       -9208140.9546  6583976.8260 -1.40 0.1624
## FTr        -864832.0012  2966216.2684 -0.29 0.7707
## ORB%       -13935.3085   983455.7931 -0.01 0.9887
## DRB%        96905.9868   952108.9228  0.10 0.9190
## TRB%       -127898.5705  1925475.0388 -0.07 0.9471
## AST%        22358.3863    87919.7488  0.25 0.7993
## STL%       -318597.4194   549093.9678 -0.58 0.5620
## BLK%       -125003.0175   446685.1093 -0.28 0.7797
## TOV%        162938.3018    93647.8751  1.74 0.0823
## USG%        63188.3455   191214.7960  0.33 0.7412
## OWS        3802326.4106  3685604.0741  1.03 0.3026
## DWS        5753339.0399  3706359.0546  1.55 0.1211
## WS         -2766186.6134  3665208.6031 -0.75 0.4507
## WS/48      -13514187.4727 23306267.6705 -0.58 0.5622
## OBPM       -3349707.6554  3739313.3692 -0.90 0.3707
## DBPM       -4411997.3281  3702135.0896 -1.19 0.2338
## BPM        4395848.8350  3705300.2579  1.19 0.2359
## VORP       -1466784.7748   578669.7910 -2.53 0.0115
## FG          5321.2958    29759.5462  0.18 0.8581
## FGA        -1427.3974    14748.7000 -0.10 0.9229
## FG%       -19482411.5673 37525528.8747 -0.52 0.6038
## 3P         -27458.0568    38356.6988 -0.72 0.4743
## 3PA        16857.2778    15702.1863  1.07 0.2834
## 3P%       -1458432.2260  2062109.4362 -0.71 0.4797
## 2P%       -2848969.5726  6705704.4713 -0.42 0.6711
## eFG%       14601808.7383 36443979.8601  0.40 0.6888
## FT         -31094.2371    24687.5980 -1.26 0.2083
## FTA        28981.2745     15243.1741  1.90 0.0577
## FT%       1239694.3273  2270353.8892  0.55 0.5852
## ORB        -9758.0069    13596.9216 -0.72 0.4732
## DRB         471.5758     6390.2429  0.07 0.9412
## AST        4909.9933     9000.0989  0.55 0.5856
## STL       -20833.1875    18431.4023 -1.13 0.2588
## BLK        10471.4317    18122.7556  0.58 0.5636
## TOV       -9249.0388    22702.3733 -0.41 0.6838
## PF        -22655.3333     9017.1945 -2.51 0.0122
##

```

```

p_selected <- fastbw(p_selection_model, rule = "p", sls = 0.1)
p_selected

```

```

##
## Deleted Chi-Sq d.f. P      Residual d.f. P      AIC      R2
## ORB%      0.00   1   0.9887  0.00      1   0.9887  -2.00 0.650
## PER       0.00   1   0.9604  0.00      2   0.9987  -4.00 0.650
## DRB       0.01   1   0.9428  0.01      3   0.9998  -5.99 0.650
## FGA       0.01   1   0.9278  0.02      4   1.0000  -7.98 0.650
## AST%      0.06   1   0.8010  0.08      5   0.9999  -9.92 0.650

```

| | | | | | | | | | |
|----|-------|------|---|--------|-------|----|--------|--------|-------|
| ## | TS% | 0.10 | 1 | 0.7542 | 0.18 | 6 | 0.9999 | -11.82 | 0.650 |
| ## | eFG% | 0.07 | 1 | 0.7896 | 0.25 | 7 | 0.9999 | -13.75 | 0.650 |
| ## | BLK% | 0.10 | 1 | 0.7517 | 0.35 | 8 | 1.0000 | -15.65 | 0.650 |
| ## | FG | 0.16 | 1 | 0.6933 | 0.50 | 9 | 1.0000 | -17.50 | 0.650 |
| ## | TOV | 0.13 | 1 | 0.7199 | 0.63 | 10 | 1.0000 | -19.37 | 0.650 |
| ## | FT% | 0.17 | 1 | 0.6811 | 0.80 | 11 | 1.0000 | -21.20 | 0.650 |
| ## | BLK | 0.24 | 1 | 0.6272 | 1.04 | 12 | 1.0000 | -22.96 | 0.650 |
| ## | FTr | 0.42 | 1 | 0.5165 | 1.46 | 13 | 1.0000 | -24.54 | 0.650 |
| ## | 3P% | 0.39 | 1 | 0.5339 | 1.85 | 14 | 0.9999 | -26.15 | 0.649 |
| ## | 2P% | 0.44 | 1 | 0.5091 | 2.28 | 15 | 0.9999 | -27.72 | 0.649 |
| ## | WS | 0.60 | 1 | 0.4402 | 2.88 | 16 | 0.9999 | -29.12 | 0.649 |
| ## | OBPM | 0.67 | 1 | 0.4114 | 3.55 | 17 | 0.9998 | -30.45 | 0.649 |
| ## | 3P | 0.57 | 1 | 0.4496 | 4.12 | 18 | 0.9997 | -31.88 | 0.648 |
| ## | ORB | 0.50 | 1 | 0.4804 | 4.62 | 19 | 0.9997 | -33.38 | 0.648 |
| ## | USG% | 0.71 | 1 | 0.3990 | 5.33 | 20 | 0.9995 | -34.67 | 0.648 |
| ## | STL% | 0.57 | 1 | 0.4492 | 5.91 | 21 | 0.9995 | -36.09 | 0.647 |
| ## | AST | 2.68 | 1 | 0.1014 | 8.59 | 22 | 0.9952 | -35.41 | 0.646 |
| ## | GS | 3.18 | 1 | 0.0744 | 11.77 | 23 | 0.9738 | -34.23 | 0.644 |
| ## | 3PA | 3.68 | 1 | 0.0551 | 15.45 | 24 | 0.9068 | -32.55 | 0.642 |
| ## | FT | 2.19 | 1 | 0.1388 | 17.64 | 25 | 0.8572 | -32.36 | 0.641 |
| ## | TRB% | 3.54 | 1 | 0.0600 | 21.18 | 26 | 0.7326 | -30.82 | 0.639 |
| ## | DRB% | 0.84 | 1 | 0.3588 | 22.02 | 27 | 0.7363 | -31.98 | 0.639 |
| ## | VORP | 5.25 | 1 | 0.0219 | 27.28 | 28 | 0.5033 | -28.72 | 0.636 |
| ## | OWS | 1.82 | 1 | 0.1768 | 29.10 | 29 | 0.4598 | -28.90 | 0.635 |
| ## | WS/48 | 1.90 | 1 | 0.1676 | 31.00 | 30 | 0.4152 | -29.00 | 0.634 |
| ## | FTA | 7.36 | 1 | 0.0067 | 38.37 | 31 | 0.1701 | -23.63 | 0.630 |

##

Approximate Estimates after Deleting Factors

##

| ## | | Coef | S.E. | Wald Z | P |
|----|-----------|----------|-----------|-----------|-----------|
| ## | Intercept | 11368576 | 2727851.6 | 4.167593 | 3.078e-05 |
| ## | year=2017 | 1666466 | 345793.0 | 4.819259 | 1.441e-06 |
| ## | Pos=PF | -734026 | 603754.2 | -1.215770 | 2.241e-01 |
| ## | Pos=PG | -3600195 | 816386.1 | -4.409917 | 1.034e-05 |
| ## | Pos=SF | -1084503 | 743912.2 | -1.457838 | 1.449e-01 |
| ## | Pos=SG | -2215167 | 806338.7 | -2.747192 | 6.011e-03 |
| ## | Age | 205814 | 43171.4 | 4.767363 | 1.867e-06 |
| ## | Tm=BOS | -1480253 | 1463348.8 | -1.011552 | 3.118e-01 |
| ## | Tm=BRK | -1661045 | 1578683.2 | -1.052171 | 2.927e-01 |
| ## | Tm=CHI | -1808958 | 1462290.5 | -1.237071 | 2.161e-01 |
| ## | Tm=CHO | -1126049 | 1502063.7 | -0.749668 | 4.535e-01 |
| ## | Tm=CLE | 2243122 | 1517714.6 | 1.477961 | 1.394e-01 |
| ## | Tm=DAL | -268113 | 1494602.6 | -0.179387 | 8.576e-01 |
| ## | Tm=DEN | -2191865 | 1538554.2 | -1.424626 | 1.543e-01 |
| ## | Tm=DET | -1498243 | 1452736.7 | -1.031325 | 3.024e-01 |
| ## | Tm=GSW | -1076531 | 1409553.6 | -0.763739 | 4.450e-01 |
| ## | Tm=HOU | -849539 | 1534211.5 | -0.553730 | 5.798e-01 |
| ## | Tm=IND | -2325625 | 1490515.5 | -1.560282 | 1.187e-01 |
| ## | Tm=LAC | 2107487 | 1474254.0 | 1.429528 | 1.529e-01 |
| ## | Tm=LAL | -123513 | 1655380.4 | -0.074613 | 9.405e-01 |
| ## | Tm=MEM | 86986 | 1480249.0 | 0.058764 | 9.531e-01 |
| ## | Tm=MIA | -1252501 | 1460806.8 | -0.857403 | 3.912e-01 |
| ## | Tm=MIL | 600669 | 1498849.8 | 0.400753 | 6.886e-01 |
| ## | Tm=MIN | -2103225 | 1568247.4 | -1.341131 | 1.799e-01 |
| ## | Tm=NOP | 800044 | 1522970.8 | 0.525318 | 5.994e-01 |
| ## | Tm=NYK | -1464675 | 1521011.3 | -0.962961 | 3.356e-01 |
| ## | Tm=OKC | 797103 | 1470721.8 | 0.541981 | 5.878e-01 |
| ## | Tm=ORL | -674383 | 1506074.6 | -0.447776 | 6.543e-01 |
| ## | Tm=PHI | -2929741 | 1560719.2 | -1.877174 | 6.049e-02 |
| ## | Tm=PHO | -52994 | 1524861.2 | -0.034754 | 9.723e-01 |
| ## | Tm=POR | 2480883 | 1518392.5 | 1.633888 | 1.023e-01 |
| ## | Tm=SAC | -974468 | 1535754.8 | -0.634521 | 5.257e-01 |
| ## | Tm=SAS | -3174467 | 1476189.9 | -2.150446 | 3.152e-02 |
| ## | Tm=TOR | 4237 | 1431884.8 | 0.002959 | 9.976e-01 |

```
## Tm=TOT      -1863692 1261176.0 -1.477741 1.395e-01
## Tm=UTA      -1507688 1429298.9 -1.054844 2.915e-01
## Tm=WAS      1044388 1560997.7  0.669051 5.035e-01
## G           -103906   17813.9 -5.832870 5.448e-09
## MP           6512      701.8  9.279091 0.000e+00
## 3PAr        -8430336 1415510.9 -5.955684 2.590e-09
## TOV%         183538   50384.3  3.642767 2.697e-04
## DWS          3025582 403476.0  7.498790 6.439e-14
## DBPM         -1567736 185014.1 -8.473603 0.000e+00
## BPM          1029230 117101.2  8.789239 0.000e+00
## FG%         -20767380 3845676.5 -5.400189 6.657e-08
## STL          -42860   11479.1 -3.733711 1.887e-04
## PF           -26674    6769.4 -3.940413 8.134e-05
##
## Factors in Final Model
##
## [1] year Pos Age Tm G MP 3PAr TOV% DWS DBPM BPM FG% STL PF
```

Checking for Multicollinearity Among Optimal Subset of Primary Variables.

```
p_subset_formula <- get_salary_formula(p_selected[['names.kept']])
p_subset_formula
```

```
## salary ~ year + Pos + Age + Tm + G + MP + `3PAr` + `TOV%` + DWS +
##      DBPM + BPM + `FG%` + STL + PF
## <environment: 0x55edcc71ded0>
```

```
p_subset_lm <- lm(p_subset_formula , data=df_p_final)
summary(p_subset_lm)
```

```
##
## Call:
## lm(formula = p_subset_formula, data = df_p_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13774314 -2907843  -191667   2718447  16789754
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.137e+07  2.742e+06   4.145 3.82e-05 ***
## year2017     1.666e+06  3.476e+05   4.794 2.01e-06 ***
## PosPF        -7.340e+05  6.070e+05  -1.209 0.226963
## PosPG        -3.600e+06  8.208e+05  -4.386 1.33e-05 ***
## PosSF        -1.085e+06  7.479e+05  -1.450 0.147492
## PosSG        -2.215e+06  8.107e+05  -2.733 0.006446 **
## Age           2.058e+05  4.340e+04   4.742 2.57e-06 ***
## TmBOS         -1.480e+06  1.471e+06  -1.006 0.314688
## TmBRK         -1.661e+06  1.587e+06  -1.047 0.295665
## TmCHI         -1.809e+06  1.470e+06  -1.230 0.218935
## TmCHO         -1.126e+06  1.510e+06  -0.746 0.456116
## TmCLE         2.243e+06  1.526e+06   1.470 0.141993
## TmDAL        -2.681e+05  1.503e+06  -0.178 0.858436
## TmDEN        -2.192e+06  1.547e+06  -1.417 0.156923
## TmDET        -1.498e+06  1.461e+06  -1.026 0.305329
## TmGSW        -1.077e+06  1.417e+06  -0.760 0.447709
## TmHOU        -8.495e+05  1.542e+06  -0.551 0.581961
## TmIND        -2.326e+06  1.498e+06  -1.552 0.121127
## TmLAC         2.107e+06  1.482e+06   1.422 0.155503
## TmLAL        -1.235e+05  1.664e+06  -0.074 0.940860
## TmMEM         8.699e+04  1.488e+06   0.058 0.953406
## TmMIA        -1.253e+06  1.469e+06  -0.853 0.394045
## TmMIL         6.007e+05  1.507e+06   0.399 0.690296
```

```
## TmMIN      -2.103e+06  1.577e+06  -1.334  0.182648
## TmNOP       8.000e+05  1.531e+06   0.523  0.601475
## TmNYK      -1.465e+06  1.529e+06  -0.958  0.338482
## TmOKC       7.971e+05  1.479e+06   0.539  0.589995
## TmORL      -6.744e+05  1.514e+06  -0.445  0.656176
## TmPHI      -2.930e+06  1.569e+06  -1.867  0.062301 .
## TmPHO      -5.299e+04  1.533e+06  -0.035  0.972434
## TmPOR       2.481e+06  1.527e+06   1.625  0.104580
## TmSAC      -9.745e+05  1.544e+06  -0.631  0.528156
## TmSAS      -3.174e+06  1.484e+06  -2.139  0.032787 *
## TmTOR       4.237e+03  1.440e+06   0.003  0.997652
## TmTOT      -1.864e+06  1.268e+06  -1.470  0.142052
## TmUTA      -1.508e+06  1.437e+06  -1.049  0.294440
## TmWAS       1.044e+06  1.569e+06   0.665  0.505959
## G          -1.039e+05  1.791e+04  -5.802  1.00e-08 ***
## MP          6.512e+03  7.056e+02   9.230  < 2e-16 ***
## `3Par`     -8.430e+06  1.423e+06  -5.924  4.97e-09 ***
## `TOV`      1.835e+05  5.065e+04   3.623  0.000312 ***
## DWS        3.026e+06  4.056e+05   7.459  2.64e-13 ***
## DBPM      -1.568e+06  1.860e+05  -8.429  < 2e-16 ***
## BPM        1.029e+06  1.177e+05   8.742  < 2e-16 ***
## `FG`      -2.077e+07  3.866e+06  -5.371  1.07e-07 ***
## STL        -4.286e+04  1.154e+04  -3.714  0.000221 ***
## PF         -2.667e+04  6.806e+03  -3.919  9.76e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4579000 on 687 degrees of freedom
## Multiple R-squared:  0.63, Adjusted R-squared:  0.6052
## F-statistic: 25.43 on 46 and 687 DF, p-value: < 2.2e-16
```

```
sort(vif(p_subset_lm),decreasing=T) # All variables have low VIF values. So no multicollinearity.
```

```
##          MP          DWS          PF          STL          BPM          TmTOT          G          DBPM
## 10.699640  7.253925  6.003790  5.286187  5.265620  4.855967  4.658775  4.551077
##      PosSG      PosPG      `FG`      PosSF      `3Par`      TmPHO      TmTOR      TmDEN
##  3.722505  3.600599  3.090799  3.072027  3.035783  2.707105  2.570736  2.542619
##      TmMIN      TmGSW      TmPOR      TmMEM      TmSAC      TmMIL      TmNYK      TmUTA
##  2.530405  2.491177  2.476417  2.452434  2.426636  2.413081  2.380267  2.378431
##      TmDET      TmSAS      TmORL      TmOKC      TmLAL      TmHOU      TmBOS      TmDAL
##  2.362118  2.340670  2.333747  2.323361  2.319740  2.314928  2.300124  2.298330
##      TmCHI      TmMIA      TmPHI      TmNOP      TmLAC      TmBRK      TmIND      TmCLE
##  2.296799  2.292140  2.284736  2.281131  2.236174  2.223862  2.184943  2.160561
##      TmCHO      PosPF      TmWAS      `TOV`      Age      year2017
##  2.116231  2.066017  2.062758  1.690431  1.229562  1.057815
```

```
p_vars_final <- p_selected[['names.kept']]
```

Complete Dataset Variable Selection Using Automated F-Test-Based Backward Selection

```
library(rms)
c_x_vars <- names(df_final)[!(names(df_final)%in%c('salary','name','2P','2PA','PTS','TRB'))]
# 2P, 2PA, PTS, and TRB were causing singularity in predictor matrix, so they were dropped
c_formula <- get_salary_formula(c_x_vars)
c_formula

## salary ~ year + Pos + Age + Tm + G + GS + MP + PER + `TS` +
##      `3Par` + FTr + `ORB` + `DRB` + `TRB` + `AST` + `STL` +
##      `BLK` + `TOV` + `USG` + OWS + DWS + WS + `WS/48` + OBPM +
##      DBPM + BPM + VORP + FG + FGA + `FG` + `3P` + `3PA` + `3P` +
##      `2P` + `eFG` + FT + FTA + `FT` + ORB + DRB + AST + STL +
##      BLK + TOV + PF + out + ovr + ins + pla + ath + def + reb
## <environment: 0x55edccac9210>
```

```
c_selection_model <- ols(c_formula, data = df_final)
c_selection_model
```

```
## Linear Regression Model
```

```
##
##   ols(formula = c_formula, data = df_final)
##
```

| | | Model Likelihood | Discrimination |
|----------|--------------|-------------------|----------------|
| | | Ratio Test | Indexes |
| ## Obs | 734 | LR chi2 784.09 | R2 0.656 |
| ## sigma | 4539655.3086 | d.f. 84 | R2 adj 0.612 |
| ## d.f. | 649 | Pr(> chi2) 0.0000 | g 6444818.561 |

```
## Residuals
```

| | Min | 1Q | Median | 3Q | Max |
|----|-----------|----------|---------|---------|----------|
| ## | -16194213 | -2558542 | -107917 | 2485207 | 14679029 |

| | Coef | S.E. | t | Pr(> t) |
|--------------|---------------|---------------|-------|----------|
| ## Intercept | 5069879.8497 | 6057385.9045 | 0.84 | 0.4029 |
| ## year=2017 | 1026799.6630 | 537415.8908 | 1.91 | 0.0565 |
| ## Pos=PF | -213699.9327 | 719393.8901 | -0.30 | 0.7665 |
| ## Pos=PG | -4530000.1865 | 1375907.7253 | -3.29 | 0.0010 |
| ## Pos=SF | -1246715.1812 | 999443.4478 | -1.25 | 0.2127 |
| ## Pos=SG | -2628516.2955 | 1170261.5843 | -2.25 | 0.0250 |
| ## Age | 180118.7983 | 50586.3694 | 3.56 | 0.0004 |
| ## Tm=BOS | -1910782.5143 | 1531454.8079 | -1.25 | 0.2126 |
| ## Tm=BRK | -1027333.4462 | 1836543.0295 | -0.56 | 0.5761 |
| ## Tm=CHI | -1235154.2953 | 1600571.2118 | -0.77 | 0.4406 |
| ## Tm=CHO | -806715.1214 | 1626793.9177 | -0.50 | 0.6201 |
| ## Tm=CLE | 1810376.1242 | 1662862.9070 | 1.09 | 0.2767 |
| ## Tm=DAL | -213499.3871 | 1662666.2103 | -0.13 | 0.8979 |
| ## Tm=DEN | -2028846.3612 | 1799714.1896 | -1.13 | 0.2600 |
| ## Tm=DET | -967487.2096 | 1680411.3219 | -0.58 | 0.5650 |
| ## Tm=GSW | -1286489.9973 | 1561334.9816 | -0.82 | 0.4103 |
| ## Tm=HOU | -1700951.4787 | 1738177.4563 | -0.98 | 0.3282 |
| ## Tm=IND | -1710858.8654 | 1588124.1583 | -1.08 | 0.2818 |
| ## Tm=LAC | 882690.7239 | 1592572.0266 | 0.55 | 0.5796 |
| ## Tm=LAL | -236822.0964 | 2042542.9447 | -0.12 | 0.9077 |
| ## Tm=MEM | 539004.4487 | 1704502.4187 | 0.32 | 0.7519 |
| ## Tm=MIA | -1148356.1482 | 1566535.6889 | -0.73 | 0.4638 |
| ## Tm=MIL | 807317.2762 | 1736914.8761 | 0.46 | 0.6422 |
| ## Tm=MIN | -1797604.5336 | 1884855.1503 | -0.95 | 0.3406 |
| ## Tm=NOP | 969831.5406 | 1734862.6445 | 0.56 | 0.5763 |
| ## Tm=NYK | -1194019.2090 | 1762895.0678 | -0.68 | 0.4985 |
| ## Tm=OKC | 1040172.6929 | 1674304.6286 | 0.62 | 0.5346 |
| ## Tm=ORL | -429140.9323 | 1708364.4490 | -0.25 | 0.8017 |
| ## Tm=PHI | -3092421.7525 | 1769717.6303 | -1.75 | 0.0810 |
| ## Tm=PHO | 338524.2954 | 1828707.3826 | 0.19 | 0.8532 |
| ## Tm=POR | 2466807.4132 | 1752140.0259 | 1.41 | 0.1596 |
| ## Tm=SAC | -948962.4587 | 1744399.9545 | -0.54 | 0.5866 |
| ## Tm=SAS | -2775949.5750 | 1593231.6903 | -1.74 | 0.0819 |
| ## Tm=TOR | 231609.2909 | 1664445.9920 | 0.14 | 0.8894 |
| ## Tm=TOT | -1881972.5803 | 1423107.3755 | -1.32 | 0.1865 |
| ## Tm=UTA | -1363039.6807 | 1559125.1493 | -0.87 | 0.3823 |
| ## Tm=WAS | 1089527.4447 | 1683625.8370 | 0.65 | 0.5178 |
| ## G | -75883.1333 | 22855.8379 | -3.32 | 0.0010 |
| ## GS | 20244.9840 | 11717.5256 | 1.73 | 0.0845 |
| ## MP | 2128.0078 | 1795.7054 | 1.19 | 0.2364 |
| ## PER | -21337.9181 | 432048.8045 | -0.05 | 0.9606 |
| ## TS% | 1614363.8063 | 21889647.7330 | 0.07 | 0.9412 |
| ## 3PAr | -8710974.9721 | 6574253.9610 | -1.33 | 0.1856 |

```
## FTr      -1751621.8308  2995131.4794 -0.58 0.5589
## ORB%      93447.0686   985377.2109  0.09 0.9245
## DRB%     144385.7456   952758.9161  0.15 0.8796
## TRB%    -300388.0615  1927609.0706 -0.16 0.8762
## AST%      9230.9182    89334.7433  0.10 0.9177
## STL%    -295561.7893   552323.8476 -0.54 0.5927
## BLK%    -254371.4054   452379.7970 -0.56 0.5741
## TOV%     168439.9776    94724.2638  1.78 0.0758
## USG%      98834.2503   192779.1510  0.51 0.6083
## OWS     3726467.4784   3697059.0193  1.01 0.3139
## DWS     5425708.1459   3707651.9702  1.46 0.1438
## WS     -2635473.5557   3669699.5315 -0.72 0.4729
## WS/48    -8402747.4398 23359333.8287 -0.36 0.7192
## OBPM    -3207274.0621   3733338.3518 -0.86 0.3906
## DBPM    -4030858.2537   3699995.5729 -1.09 0.2764
## BPM     4079645.8399   3700140.0605  1.10 0.2706
## VORP    -1330323.9000   589958.2154 -2.25 0.0245
## FG       2515.6487     30121.5891  0.08 0.9335
## FGA      375.5791      14884.0044  0.03 0.9799
## FG%    -25561757.2251 37765742.3383 -0.68 0.4987
## 3P      -35049.0329     38850.1281 -0.90 0.3673
## 3PA     18880.1008     15817.1429  1.19 0.2331
## 3P%    -824870.5240   2122420.2492 -0.39 0.6977
## 2P%    -2356650.3731   6794461.5338 -0.35 0.7288
## eFG%    12420433.1872 36498836.7788  0.34 0.7337
## FT      -30536.8475     24891.4453 -1.23 0.2203
## FTA     26552.1472     15332.8802  1.73 0.0838
## FT%    1326996.2442   2266847.9637  0.59 0.5585
## ORB     -11686.6123     13698.9782 -0.85 0.3939
## DRB      3919.5191      6521.5827  0.60 0.5480
## AST      5525.7960      9171.0101  0.60 0.5470
## STL    -29141.3105     18660.1985 -1.56 0.1189
## BLK      7554.9460     18352.7056  0.41 0.6807
## TOV    -10045.4418     23113.9737 -0.43 0.6640
## PF     -21759.1628      9011.0403 -2.41 0.0160
## out    -66470.5101     36533.9396 -1.82 0.0693
## ovr    103934.6911     81796.4626  1.27 0.2043
## ins      4894.2371     44068.3014  0.11 0.9116
## pla     -374.5551     28163.8693 -0.01 0.9894
## ath     -788.0691     44103.6760 -0.02 0.9857
## def     44044.5138     35411.9386  1.24 0.2140
## reb    -43928.0359     24579.9507 -1.79 0.0744
##
```

```
c_selected <- fastbw(c_selection_model, rule = "p", sls = 0.1)
c_selected
```

```
##
## Deleted Chi-Sq d.f. P      Residual d.f. P      AIC      R2
## pla      0.00   1   0.9894  0.00      1   0.9894  -2.00 0.656
## ath      0.00   1   0.9842  0.00      2   0.9997  -4.00 0.656
## FGA      0.00   1   0.9787  0.00      3   1.0000  -6.00 0.656
## PER      0.00   1   0.9531  0.00      4   1.0000  -8.00 0.656
## TS%      0.01   1   0.9424  0.01      5   1.0000  -9.99 0.656
## AST%     0.01   1   0.9265  0.02      6   1.0000 -11.98 0.656
## ORB%     0.01   1   0.9221  0.03      7   1.0000 -13.97 0.656
## ins      0.01   1   0.9197  0.04      8   1.0000 -15.96 0.656
## DRB%     0.08   1   0.7748  0.12      9   1.0000 -17.88 0.656
## 2P%      0.09   1   0.7626  0.21     10   1.0000 -19.79 0.656
## 3P%      0.14   1   0.7068  0.35     11   1.0000 -21.65 0.656
## TRB%     0.16   1   0.6934  0.51     12   1.0000 -23.49 0.656
## BLK      0.28   1   0.5956  0.79     13   1.0000 -25.21 0.656
## STL%     0.27   1   0.6022  1.06     14   1.0000 -26.94 0.656
## BLK%     0.24   1   0.6270  1.30     15   1.0000 -28.70 0.656
```


| | | | | | | | | | |
|----|-------|------|---|--------|-------|----|--------|--------|-------|
| ## | USG% | 0.25 | 1 | 0.6199 | 1.54 | 16 | 1.0000 | -30.46 | 0.656 |
| ## | TOV | 0.37 | 1 | 0.5426 | 1.91 | 17 | 1.0000 | -32.09 | 0.655 |
| ## | FTr | 0.35 | 1 | 0.5553 | 2.26 | 18 | 1.0000 | -33.74 | 0.655 |
| ## | OBPM | 0.54 | 1 | 0.4623 | 2.80 | 19 | 1.0000 | -35.20 | 0.655 |
| ## | WS | 0.60 | 1 | 0.4368 | 3.41 | 20 | 1.0000 | -36.59 | 0.655 |
| ## | eFG% | 0.93 | 1 | 0.3352 | 4.34 | 21 | 1.0000 | -37.66 | 0.654 |
| ## | 3P | 0.62 | 1 | 0.4323 | 4.95 | 22 | 0.9999 | -39.05 | 0.654 |
| ## | FG | 0.96 | 1 | 0.3260 | 5.92 | 23 | 0.9999 | -40.08 | 0.653 |
| ## | FT% | 0.92 | 1 | 0.3365 | 6.84 | 24 | 0.9998 | -41.16 | 0.653 |
| ## | AST | 1.35 | 1 | 0.2454 | 8.19 | 25 | 0.9994 | -41.81 | 0.652 |
| ## | 3PA | 0.80 | 1 | 0.3718 | 8.99 | 26 | 0.9992 | -43.01 | 0.652 |
| ## | def | 1.30 | 1 | 0.2550 | 10.28 | 27 | 0.9985 | -43.72 | 0.651 |
| ## | DRB | 1.50 | 1 | 0.2202 | 11.79 | 28 | 0.9969 | -44.21 | 0.650 |
| ## | reb | 2.45 | 1 | 0.1177 | 14.23 | 29 | 0.9901 | -43.77 | 0.649 |
| ## | GS | 2.41 | 1 | 0.1203 | 16.65 | 30 | 0.9766 | -43.35 | 0.648 |
| ## | FT | 2.62 | 1 | 0.1053 | 19.27 | 31 | 0.9502 | -42.73 | 0.646 |
| ## | FTA | 2.53 | 1 | 0.1116 | 21.80 | 32 | 0.9126 | -42.20 | 0.645 |
| ## | VORP | 5.69 | 1 | 0.0170 | 27.49 | 33 | 0.7377 | -38.51 | 0.642 |
| ## | WS/48 | 2.96 | 1 | 0.0854 | 30.45 | 34 | 0.6422 | -37.55 | 0.640 |
| ## | OWS | 4.36 | 1 | 0.0367 | 34.82 | 35 | 0.4770 | -35.18 | 0.638 |
| ## | ORB | 3.60 | 1 | 0.0577 | 38.42 | 36 | 0.3605 | -33.58 | 0.636 |
| ## | out | 5.31 | 1 | 0.0212 | 43.73 | 37 | 0.2072 | -30.27 | 0.633 |

##

Approximate Estimates after Deleting Factors

##

| ## | | Coef | S.E. | Wald Z | P |
|----|-----------|----------|-----------|-----------|-----------|
| ## | Intercept | 5896725 | 3504215.6 | 1.682752 | 9.242e-02 |
| ## | year=2017 | 1343200 | 368589.3 | 3.644166 | 2.683e-04 |
| ## | Pos=PF | -619750 | 603579.1 | -1.026791 | 3.045e-01 |
| ## | Pos=PG | -3484554 | 815098.4 | -4.275009 | 1.911e-05 |
| ## | Pos=SF | -1058211 | 741594.1 | -1.426941 | 1.536e-01 |
| ## | Pos=SG | -2244312 | 803829.8 | -2.792023 | 5.238e-03 |
| ## | Age | 184012 | 43924.5 | 4.189272 | 2.799e-05 |
| ## | Tm=BOS | -1612686 | 1459619.9 | -1.104867 | 2.692e-01 |
| ## | Tm=BRK | -1619365 | 1573692.4 | -1.029023 | 3.035e-01 |
| ## | Tm=CHI | -1698077 | 1458272.4 | -1.164444 | 2.442e-01 |
| ## | Tm=CHO | -963081 | 1498676.2 | -0.642621 | 5.205e-01 |
| ## | Tm=CLE | 2014658 | 1515642.6 | 1.329243 | 1.838e-01 |
| ## | Tm=DAL | -272497 | 1489793.4 | -0.182909 | 8.549e-01 |
| ## | Tm=DEN | -2077768 | 1534295.0 | -1.354217 | 1.757e-01 |
| ## | Tm=DET | -1128787 | 1455732.3 | -0.775408 | 4.381e-01 |
| ## | Tm=GSW | -1231953 | 1406419.3 | -0.875950 | 3.811e-01 |
| ## | Tm=HOU | -921033 | 1529546.5 | -0.602161 | 5.471e-01 |
| ## | Tm=IND | -2243262 | 1486090.9 | -1.509506 | 1.312e-01 |
| ## | Tm=LAC | 1929302 | 1471271.2 | 1.311316 | 1.898e-01 |
| ## | Tm=LAL | -200703 | 1650347.3 | -0.121612 | 9.032e-01 |
| ## | Tm=MEM | 7518 | 1475834.1 | 0.005094 | 9.959e-01 |
| ## | Tm=MIA | -1150693 | 1456686.0 | -0.789939 | 4.296e-01 |
| ## | Tm=MIL | 512355 | 1494451.8 | 0.342838 | 7.317e-01 |
| ## | Tm=MIN | -2095279 | 1563203.5 | -1.340375 | 1.801e-01 |
| ## | Tm=NOP | 885845 | 1518464.9 | 0.583382 | 5.596e-01 |
| ## | Tm=NYK | -1276740 | 1518015.8 | -0.841058 | 4.003e-01 |
| ## | Tm=OKC | 806826 | 1465993.7 | 0.550361 | 5.821e-01 |
| ## | Tm=ORL | -638599 | 1501297.0 | -0.425365 | 6.706e-01 |
| ## | Tm=PHI | -2930377 | 1555696.2 | -1.883643 | 5.961e-02 |
| ## | Tm=PHO | -69435 | 1519968.0 | -0.045682 | 9.636e-01 |
| ## | Tm=POR | 2375810 | 1514100.8 | 1.569123 | 1.166e-01 |
| ## | Tm=SAC | -834386 | 1531857.8 | -0.544689 | 5.860e-01 |
| ## | Tm=SAS | -3172968 | 1471439.1 | -2.156371 | 3.105e-02 |
| ## | Tm=TOR | 58125 | 1427442.5 | 0.040720 | 9.675e-01 |
| ## | Tm=TOT | -1857670 | 1257119.4 | -1.477719 | 1.395e-01 |
| ## | Tm=UTA | -1322482 | 1426662.0 | -0.926976 | 3.539e-01 |
| ## | Tm=WAS | 1247451 | 1558134.6 | 0.800605 | 4.234e-01 |
| ## | G | -97610 | 17937.8 | -5.441588 | 5.281e-08 |


```
## MP          6109      718.2  8.505608 0.000e+00
## 3PAr        -7476606 1462612.3 -5.111817 3.191e-07
## TOV%        165806   50730.4  3.268380 1.082e-03
## DWS         2773984  414821.8  6.687170 2.275e-11
## DBPM        -1404939 195794.0 -7.175601 7.199e-13
## BPM         936082   122639.7  7.632782 2.298e-14
## FG%        -18924506 3904922.1 -4.846321 1.258e-06
## STL         -39837    11507.1 -3.461954 5.363e-04
## PF          -25940     6754.1 -3.840600 1.227e-04
## ovr         70843     28618.4  2.475425 1.331e-02
##
## Factors in Final Model
##
## [1] year Pos Age Tm G MP 3PAr TOV% DWS DBPM BPM FG% STL PF ovr
```

Checking for Multicollinearity Among Optimal Subset of Complete Variables.

```
c_subset_formula <- get_salary_formula(c_selected[['names.kept']])
c_subset_formula
```

```
## salary ~ year + Pos + Age + Tm + G + MP + `3PAr` + `TOV%` + DWS +
## DBPM + BPM + `FG%` + STL + PF + ovr
## <environment: 0x55edcc202478>
```

```
c_subset_lm <- lm(c_subset_formula , data=df_final)
summary(c_subset_lm)
```

```
##
## Call:
## lm(formula = c_subset_formula, data = df_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13829069 -2864315  -174920   2678634  15997459
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.897e+06  3.521e+06   1.675 0.094478 .
## year2017     1.343e+06  3.704e+05   3.626 0.000309 ***
## PosPF        -6.197e+05  6.065e+05  -1.022 0.307241
## PosPG        -3.485e+06  8.191e+05  -4.254 2.39e-05 ***
## PosSF        -1.058e+06  7.452e+05  -1.420 0.156065
## PosSG        -2.244e+06  8.078e+05  -2.778 0.005612 **
## Age          1.840e+05  4.414e+04   4.169 3.45e-05 ***
## TmBOS        -1.613e+06  1.467e+06  -1.099 0.271942
## TmBRK        -1.619e+06  1.581e+06  -1.024 0.306192
## TmCHI        -1.698e+06  1.465e+06  -1.159 0.246952
## TmCHO        -9.631e+05  1.506e+06  -0.639 0.522717
## TmCLE        2.015e+06  1.523e+06   1.323 0.186353
## TmDAL        -2.725e+05  1.497e+06  -0.182 0.855622
## TmDEN        -2.078e+06  1.542e+06  -1.348 0.178226
## TmDET        -1.129e+06  1.463e+06  -0.772 0.440599
## TmGSW        -1.232e+06  1.413e+06  -0.872 0.383687
## TmHOU        -9.210e+05  1.537e+06  -0.599 0.549219
## TmIND        -2.243e+06  1.493e+06  -1.502 0.133518
## TmLAC        1.929e+06  1.478e+06   1.305 0.192355
## TmLAL        -2.007e+05  1.658e+06  -0.121 0.903711
## TmMEM        7.518e+03  1.483e+06   0.005 0.995957
## TmMIA        -1.151e+06  1.464e+06  -0.786 0.432086
## TmMIL        5.124e+05  1.502e+06   0.341 0.733081
## TmMIN        -2.095e+06  1.571e+06  -1.334 0.182697
## TmNOP        8.858e+05  1.526e+06   0.581 0.561741
## TmNYK        -1.277e+06  1.525e+06  -0.837 0.402906
```

```
## TmOKC      8.068e+05  1.473e+06   0.548 0.584090
## TmORL     -6.386e+05  1.509e+06  -0.423 0.672215
## TmPHI     -2.930e+06  1.563e+06  -1.874 0.061291 .
## TmPHO     -6.943e+04  1.527e+06  -0.045 0.963755
## TmPOR      2.376e+06  1.522e+06   1.561 0.118872
## TmSAC     -8.344e+05  1.539e+06  -0.542 0.587970
## TmSAS     -3.173e+06  1.479e+06  -2.146 0.032234 *
## TmTOR      5.813e+04  1.434e+06   0.041 0.967689
## TmTOT     -1.858e+06  1.263e+06  -1.471 0.141879
## TmUTA     -1.322e+06  1.434e+06  -0.922 0.356613
## TmWAS      1.247e+06  1.566e+06   0.797 0.425898
## G         -9.761e+04  1.803e+04  -5.415 8.48e-08 ***
## MP         6.109e+03  7.218e+02   8.464 < 2e-16 ***
## `3PAr`    -7.477e+06  1.470e+06  -5.087 4.70e-07 ***
## `TOV%`     1.658e+05  5.098e+04   3.252 0.001200 **
## DWS        2.774e+06  4.169e+05   6.655 5.81e-11 ***
## DBPM       -1.405e+06  1.968e+05  -7.141 2.37e-12 ***
## BPM        9.361e+05  1.232e+05   7.596 1.01e-13 ***
## `FG%`     -1.892e+07  3.924e+06  -4.823 1.75e-06 ***
## STL        -3.984e+04  1.156e+04  -3.445 0.000606 ***
## PF         -2.594e+04  6.787e+03  -3.822 0.000144 ***
## ovr        7.084e+04  2.876e+04   2.463 0.014008 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4562000 on 686 degrees of freedom
## Multiple R-squared:  0.6332, Adjusted R-squared:  0.6081
## F-statistic: 25.2 on 47 and 686 DF,  p-value: < 2.2e-16
```

```
sort(vif(c_subset_lm),decreasing=T) # All variables have low VIF values. So no multicollinearity.
```

```
##          MP          DWS          PF          BPM          STL          DBPM          TmTOT          G
## 11.279200  7.717215  6.015393  5.812854  5.346383  5.129833  4.855985  4.754360
##      PosSG      PosPG      `3PAr`      `FG%`      PosSF      TmPHO      TmTOR      TmDEN
##  3.723304  3.612465  3.262140  3.207377  3.072658  2.707156  2.571335  2.544916
##      ovr      TmMIN      TmGSW      TmPOR      TmMEM      TmSAC      TmMIL      TmDET
##  2.538978  2.530416  2.496152  2.478364  2.453595  2.429952  2.414457  2.387211
##      TmNYK      TmUTA      TmSAS      TmORL      TmOKC      TmLAL      TmHOU      TmBOS
##  2.386236  2.384990  2.340670  2.333964  2.323378  2.320569  2.315754  2.303219
##      TmCHI      TmDAL      TmMIA      TmPHI      TmNOP      TmLAC      TmBRK      TmIND
##  2.298968  2.298333  2.293969  2.284736  2.282320  2.241539  2.224116  2.186038
##      TmCLE      TmCHO      PosPF      TmWAS      `TOV%`      Age      year2017
##  2.168603  2.120323  2.078174  2.068491  1.724820  1.281067  1.209658
```

```
c_vars_final <- c_seleced[['names.kept']]
```

Subset Primary and Complete Dataframes to Include Only Name, Salary, and Selected Variables

```
p_vars_subset <- c('name','salary',p_vars_final)
df_p_subset_final <- df_p_final[,p_vars_subset]
c_vars_subset <- c('name','salary',c_vars_final)
df_c_subset_final <- df_final[,c_vars_subset]
```

Split Train-Test

```
library(caret)
set.seed(7)
```

Primary Dataset

```
train_rows <- createDataPartition(y=df_p_subset_final[, 'salary'], list=FALSE, p=.8)
p_train_df <- df_p_subset_final[train_rows,]
p_test_df <- df_p_subset_final[-train_rows,]
stopifnot(nrow(p_train_df) + nrow(p_test_df) == nrow(df_p_subset_final))
nrow(p_train_df)

## [1] 590

nrow(p_test_df)

## [1] 144

names(p_train_df)

## [1] "name" "salary" "year" "Pos" "Age" "Tm" "G" "MP"
## [9] "3PAr" "TOV%" "DWS" "DBPM" "BPM" "FG%" "STL" "PF"

head(p_train_df)

##           name salary year Pos Age  Tm  G   MP  3PAr TOV% DWS DBPM  BPM  FG%
## 1 aaron brooks 2700000 2016  PG  31 CHI 69 1108 0.394 14.2 0.7 -2.8 -3.3 0.401
## 2 aaron brooks 2116955 2017  PG  32 IND 65  894 0.427 17.2 0.5 -2.6 -4.6 0.403
## 3 aaron gordon 4351320 2016  PF  20 ORL 78 1863 0.245  9.0 2.2  1.2  1.8 0.473
## 4 aaron gordon 5504420 2017  SF  21 ORL 80 2298 0.309  8.5 1.7 -0.4 -0.7 0.454
## 5 adreian payne 2022240 2016  PF  24 MIN 52  486 0.221 18.7 0.4 -0.2 -6.1 0.366
## 6   aj hammons 1312611 2017   C  24 DAL 22  163 0.238 16.4 0.2  1.9 -5.6 0.405
##   STL  PF
## 1   30 132
## 2   25  93
## 3   59 153
## 4   64 172
## 5   16  77
## 6    1  21

write.csv(p_train_df, '../data/train_test/primary/train.csv')
write.csv(p_test_df, '../data/train_test/primary/test.csv')
```

Complete Dataset

```
library(caret)
set.seed(7)
train_rows <- createDataPartition(y=df_c_subset_final[, 'salary'], list=FALSE, p=.8)
c_train_df <- df_c_subset_final[train_rows,]
c_test_df <- df_c_subset_final[-train_rows,]
stopifnot(nrow(c_train_df) + nrow(c_test_df) == nrow(df_c_subset_final))
nrow(c_train_df)

## [1] 590

nrow(c_test_df)

## [1] 144

names(c_train_df)

## [1] "name" "salary" "year" "Pos" "Age" "Tm" "G" "MP"
## [9] "3PAr" "TOV%" "DWS" "DBPM" "BPM" "FG%" "STL" "PF"
## [17] "ovr"

head(c_train_df)

##           name salary year Pos Age  Tm  G   MP  3PAr TOV% DWS DBPM  BPM  FG%
## 1 aaron brooks 2700000 2016  PG  31 CHI 69 1108 0.394 14.2 0.7 -2.8 -3.3 0.401
## 2 aaron brooks 2116955 2017  PG  32 IND 65  894 0.427 17.2 0.5 -2.6 -4.6 0.403
## 3 aaron gordon 4351320 2016  PF  20 ORL 78 1863 0.245  9.0 2.2  1.2  1.8 0.473
```

```
## 4 aaron gordon 5504420 2017 SF 21 ORL 80 2298 0.309 8.5 1.7 -0.4 -0.7 0.454
## 5 adreian payne 2022240 2016 PF 24 MIN 52 486 0.221 18.7 0.4 -0.2 -6.1 0.366
## 6 aj hammons 1312611 2017 C 24 DAL 22 163 0.238 16.4 0.2 1.9 -5.6 0.405
## STL PF ovr
## 1 30 132 75
## 2 25 93 85
## 3 59 153 90
## 4 64 172 92
## 5 16 77 69
## 6 1 21 66
```

```
write.csv(c_train_df,'../data/train_test/complete/train.csv')
write.csv(c_test_df,'../data/train_test/complete/test.csv')
```