

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://mtdb.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://mtdb.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[,-c('name_p','salary')]
colnames(df_primary)[1:3] <- c('year','name_p','salary')
df_primary <- df_primary[!is.na(df_primary[['salary']]),] # drop rows with no salaryes

original_p <- df_primary
# Checking for missing values in primary dataset
library(Amelia)
missCounts <- sapply(df_primary,function(x) sum(is.na(x)))
missCounts
```

##	year	name_p	salary	Pos	Age	Tm	G	GS	MP	PER	TS%
##	0	0	0	0	0	0	0	0	0	0	24
##	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	OWS
##	30	30	0	0	0	0	0	0	19	0	0
##	DWS	WS	WS/48	OBPM	DBPM	BPM	VORP	FG	FGA	FG%	3P

```

##      0      0      0      0      0      0      0      0      0      30      0
##  3PA   3P%    2P   2PA   2P%   eFG%    FT   FTA   FT%   ORB   DRB
##      0  1562      0      0     40     30      0      0   233      0      0
##  TRB   AST   STL   BLK   TOV    PF    PTS
##      0      0      0      0      0      0      0

```

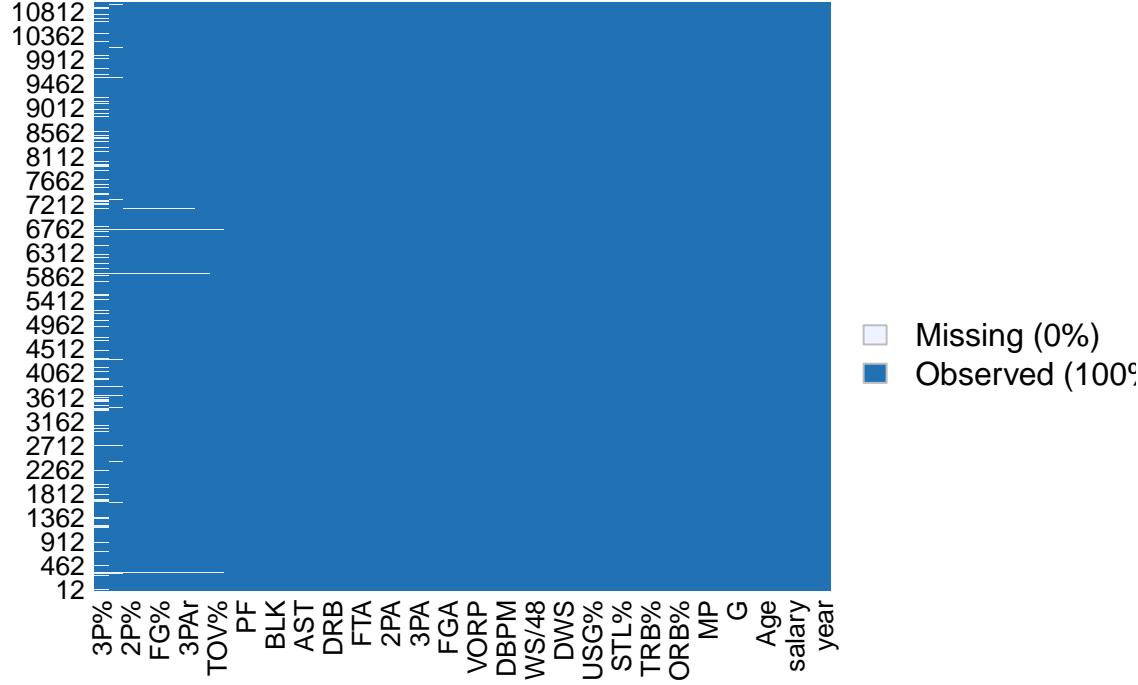
```
missmap(df_primary, main = "Missing values ")
```

```

## Warning: Unknown or uninitialised column: 'arguments'.
## Warning: Unknown or uninitialised column: 'arguments'.
## Warning: Unknown or uninitialised column: 'imputations'.

```

Missing values



```
dev.copy(png,"../figures/Missing values/Missing_value_before_primary.png")
```

```

## png
## 3
dev.off()

## pdf
## 2
sum(missCounts)

## [1] 1998
nrow(df_primary)

## [1] 10977
# Replacing the missng values with mean

df_primary$`3P%`[is.na(df_primary$`3P%`)] <- mean(df_primary$`3P%` , na.rm = T)
df_primary$`FT%`[is.na(df_primary$`FT%`)] <- mean(df_primary$`FT%` , na.rm = T)
df_primary$`TOV%`[is.na(df_primary$`TOV%`)] <- mean(df_primary$`TOV%` , na.rm = T)
df_primary$`FG%`[is.na(df_primary$`FG%`)] <- mean(df_primary$`FG%` , na.rm = T)
df_primary$`2P%`[is.na(df_primary$`2P%`)] <- mean(df_primary$`2P%` , na.rm = T)
df_primary$`eFG%`[is.na(df_primary$`eFG%`)] <- mean(df_primary$`eFG%` , na.rm = T)
df_primary$`TS%`[is.na(df_primary$`TS%`)] <- mean(df_primary$`TS%` , na.rm = T)
df_primary$`3Par`[is.na(df_primary$`3Par`)] <- mean(df_primary$`3Par` , na.rm = T)
df_primary$`FTr`[is.na(df_primary$`FTr`)] <- mean(df_primary$`FTr` , na.rm = T)

```

```

missCounts <- sapply(df_primary, function(x) sum(is.na(x)))
missCounts

##   year name_p salary   Pos   Age   Tm    G   GS   MP   PER   TS%
##   0     0      0     0     0     0     0     0     0     0     0     0
## 3Par   FTr   ORB%  DRB%  TRB%  AST%  STL%  BLK%  TOV%  USG%  OWS
##   0     0      0     0     0     0     0     0     0     0     0     0
##  DWS   WS   WS/48  OBPM  DBPM  BPM   VORP   FG   FGA   FG%   3P
##   0     0      0     0     0     0     0     0     0     0     0     0
## 3PA   3P%   2P   2PA   2P%  eFG%   FT   FTA   FT%  ORB   DRB
##   0     0      0     0     0     0     0     0     0     0     0     0
##  TRB   AST   STL   BLK   TOV   PF   PTS
##   0     0      0     0     0     0     0

```

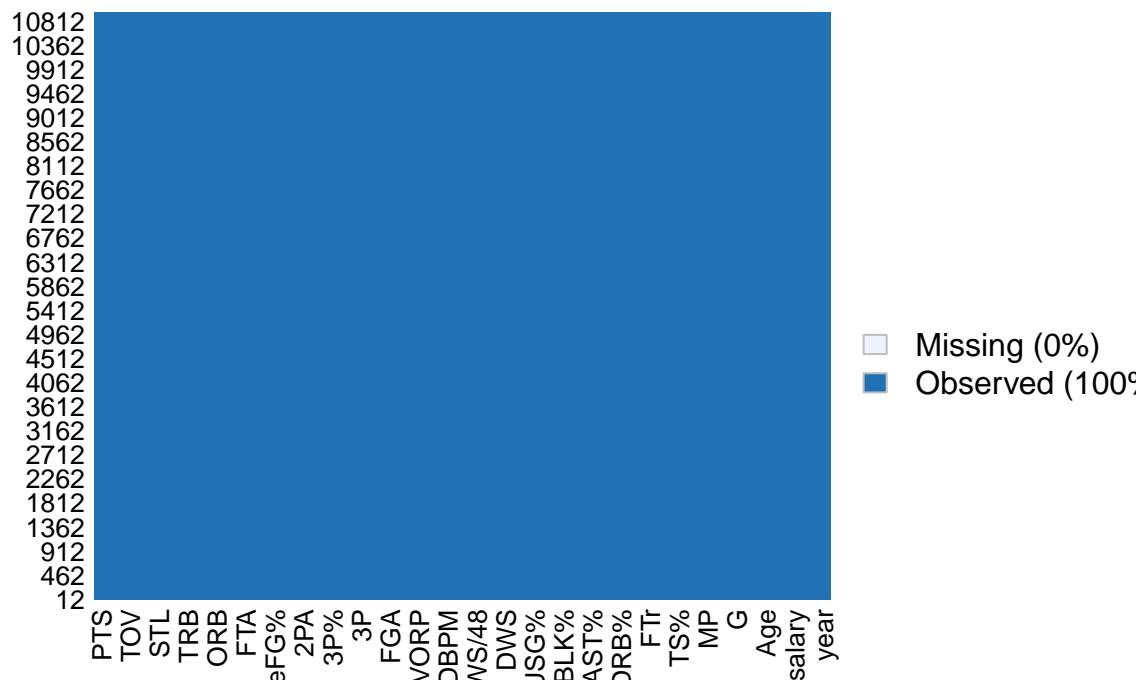
```
missmap(df_primary, main = "Missing values")
```

```
## Warning: Unknown or uninitialized column: 'arguments'.
```

```
## Warning: Unknown or uninitialized column: 'arguments'.
```

```
## Warning: Unknown or uninitialized column: 'imputations'.
```

Missing values



```
dev.copy(png, "../figures/Missing values/Missing_value_after_primary.png")
```

```
## png
## 3
dev.off()

## pdf
## 2
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> <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.:0.1400
## Median :	1197	Median : 13.30	Median :0.5380	Median :0.3110
## Mean :	1247	Mean : 13.61	Mean :0.5345	Mean :0.3053
## 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.:	0.1680	1st Qu.: 1.900	1st Qu.:10.30	1st Qu.: 6.200
## Median :	0.2410	Median : 3.300	Median :14.00	Median : 8.800
## Mean :	0.2695	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 :12.88
## 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.:0.4060
## Median :	0.2000	Median :166.0	Median : 368.0	Median :0.4420
## Mean :	0.6493	Mean :200.8	Mean : 441.5	Mean : 0.4481

```

## 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.:0.2711 1st Qu.: 43
## Median : 30.00 Median : 92.0 Median :0.3330 Median :113
## Mean : 47.83 Mean :133.8 Mean :0.3042 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.:0.4480 1st Qu.:0.4680 1st Qu.: 23.00
## Median : 235.0 Median :0.4830 Median :0.5010 Median : 59.00
## Mean : 307.8 Mean :0.4875 Mean :0.5006 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.:0.6910 1st Qu.: 13.00 1st Qu.: 62
## Median : 78.0 Median :0.7640 Median : 33.00 Median :143
## Mean :120.3 Mean :0.7463 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> <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 Jofffr~ 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>, O BPM <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)

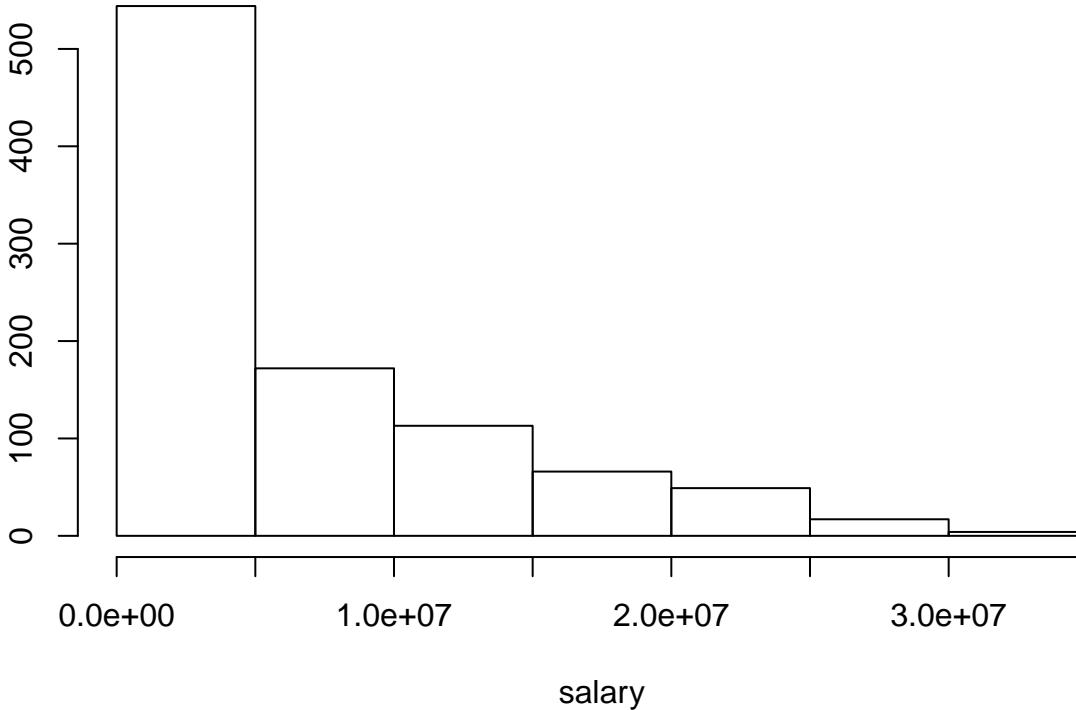
## Classes 'tbl_df', 'tbl' and 'data.frame': 965 obs. of 51 variables:
## $ year : Factor w/ 2 levels "2016","2017": 2 1 2 1 2 1 1 1 2 1 ...
## $ name_p: chr "A.J. Hammons" "Aaron Brooks" "Aaron Brooks" "Aaron Gordon" ...
## $ salary: num 1312611 2700000 2116955 4351320 5504420 ...
## $ Pos : Factor 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 ...
## $ 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

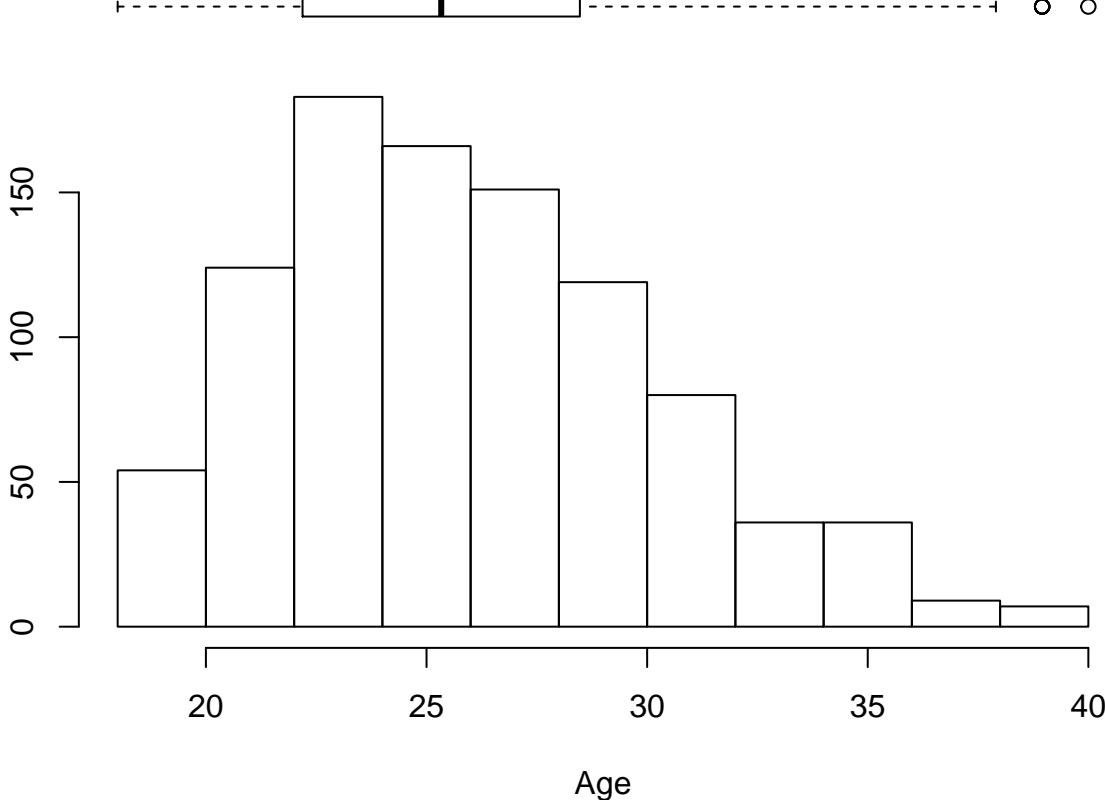
```
df_p_numeric <- Filter(is.numeric, df_primary) # numeric variables
for (col in names(df_p_numeric)){
  data <- df_p_numeric[[col]]
  layout(mat = matrix(c(1,2),2,1, byrow=TRUE), height = c(1,8))
  par(mar=c(0, 3.1, 1.1, 2.1))
  boxplot(data , horizontal=TRUE , xaxt="n", frame=F, main=sprintf('Histogram 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



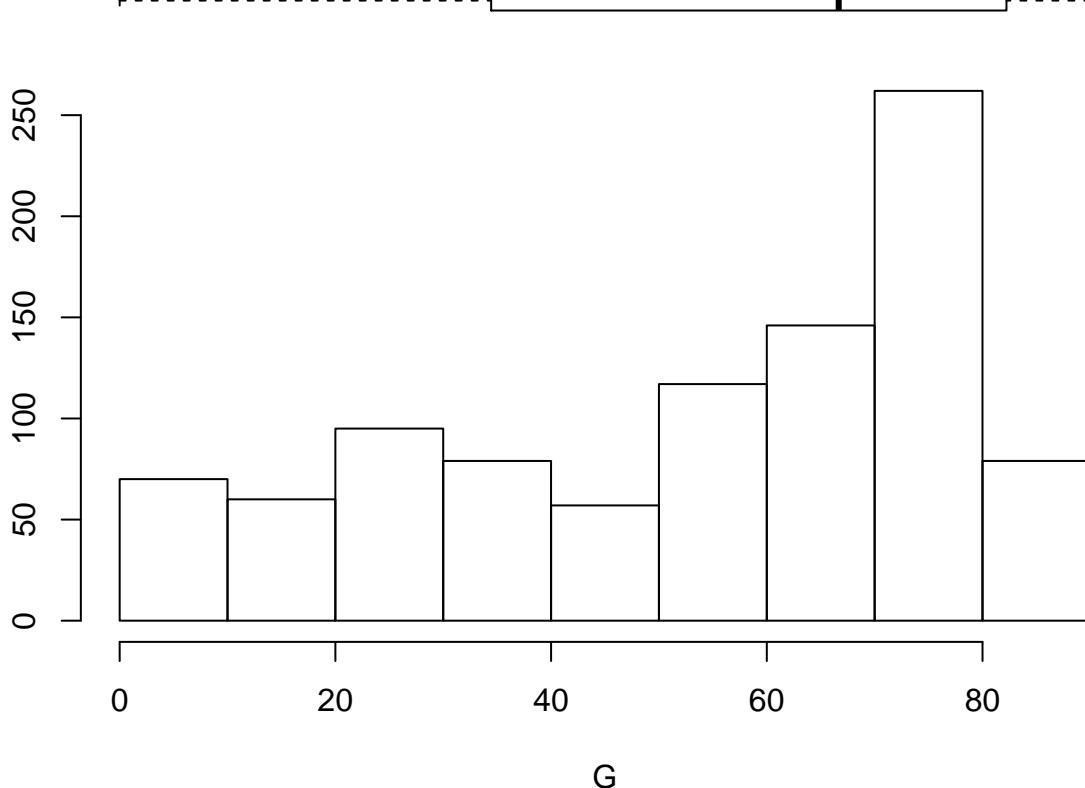
```
## Top 10 Players by salary
## # A tibble: 10 x 51
##   year name_p salary Pos     Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>  <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Steph~ 3.47e7 PG      28 GSW    79    79  2638  24.6 0.624  0.547
## 2 2017 LeBro~ 3.33e7 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



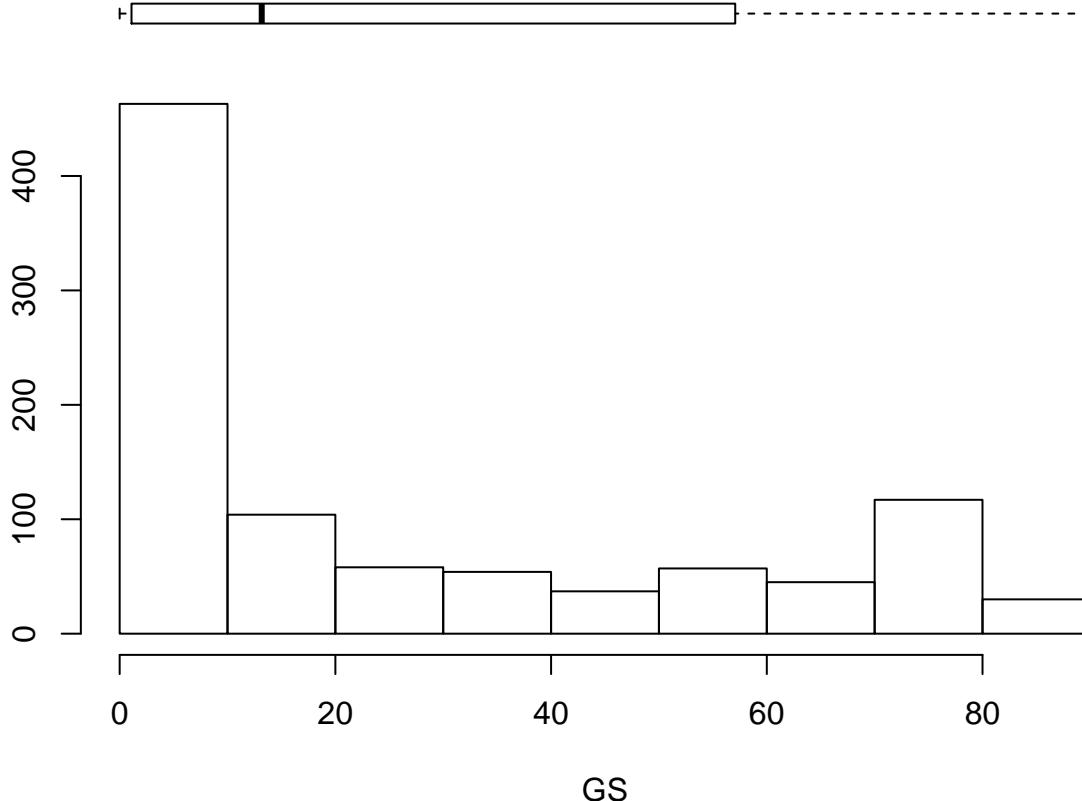
```
## Top 10 Players by Age
## # A tibble: 10 x 51
##   year name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>  <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Vince~ 8.00e6 SF      40 MEM     73   15  1799 11.7 0.542 0.604
## 2 2017 Jason~ 2.33e6 SG     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>
```

Histogram of G



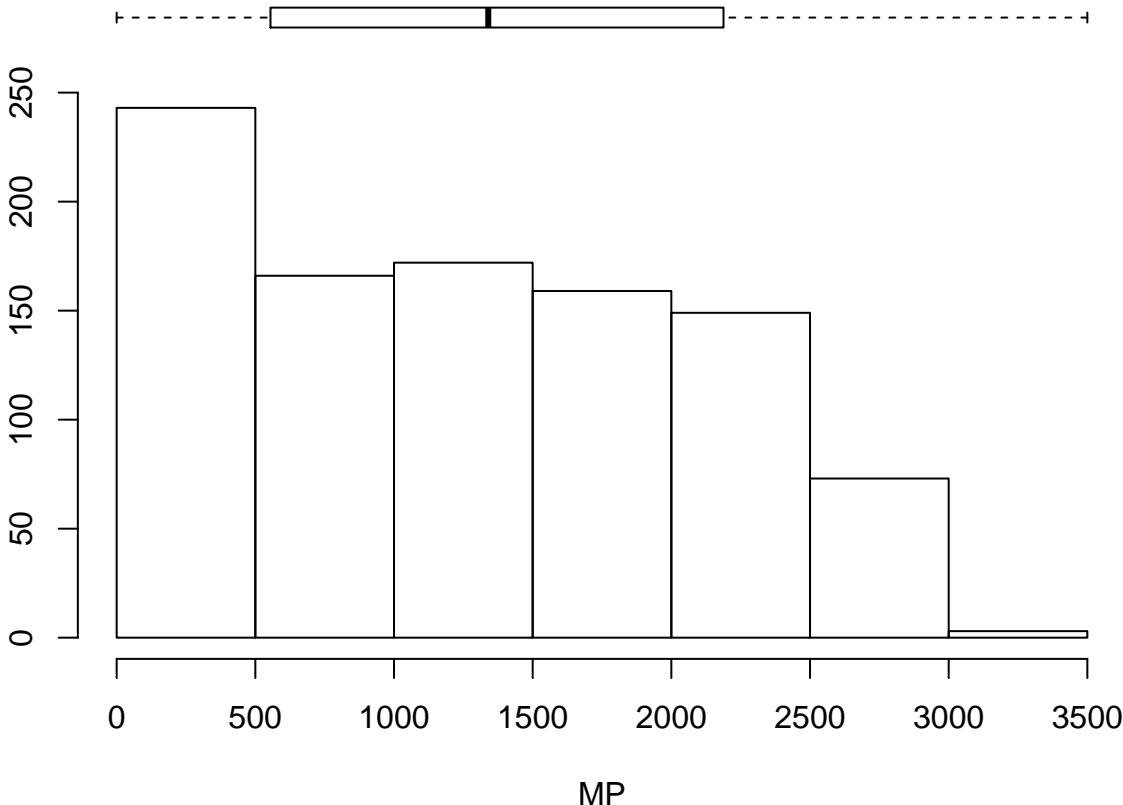
```
## Top 10 Players by G
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Al Ho- 2.65e7 C       29 ATL     82   82  2631 19.4 0.565 0.244
## 2 2016 Al-Fa- 7.68e6 SF     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>, O BPM <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



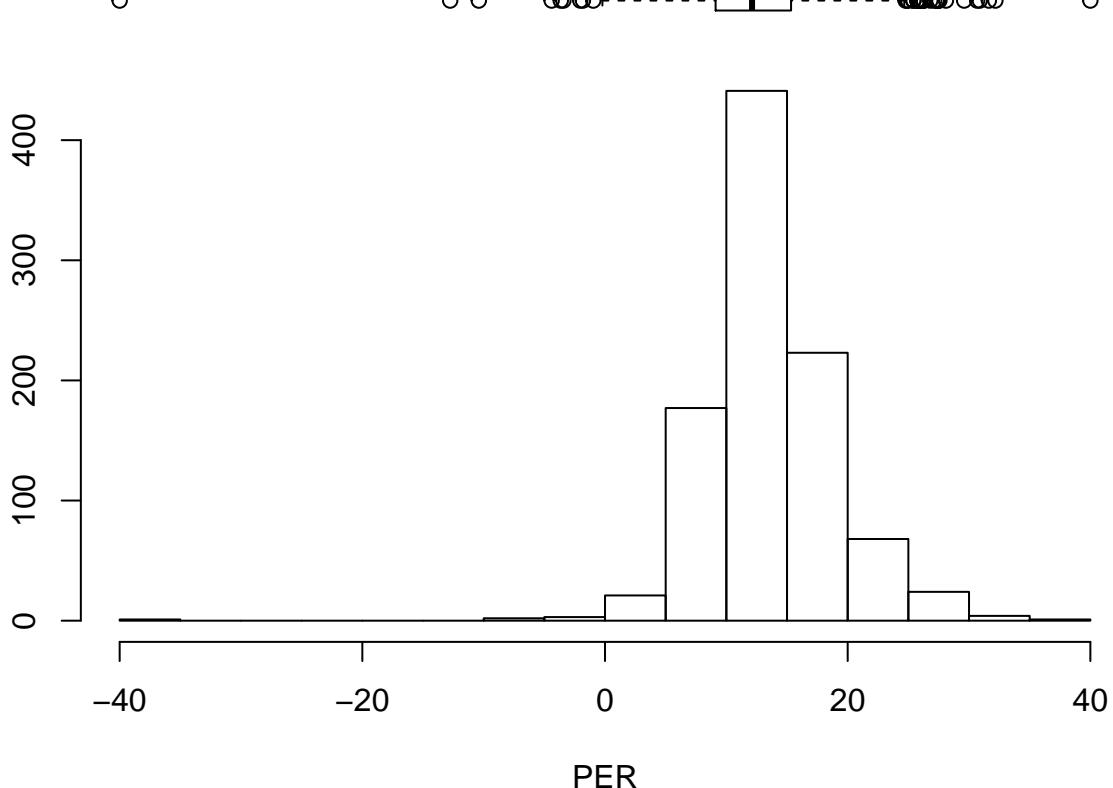
```
## Top 10 Players by GS
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Al Ho-  2.65e7 C       29 ATL     82   82  2631 19.4 0.565 0.244
## 2 2016 Al-Fa- 7.68e6 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>, O BPM <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



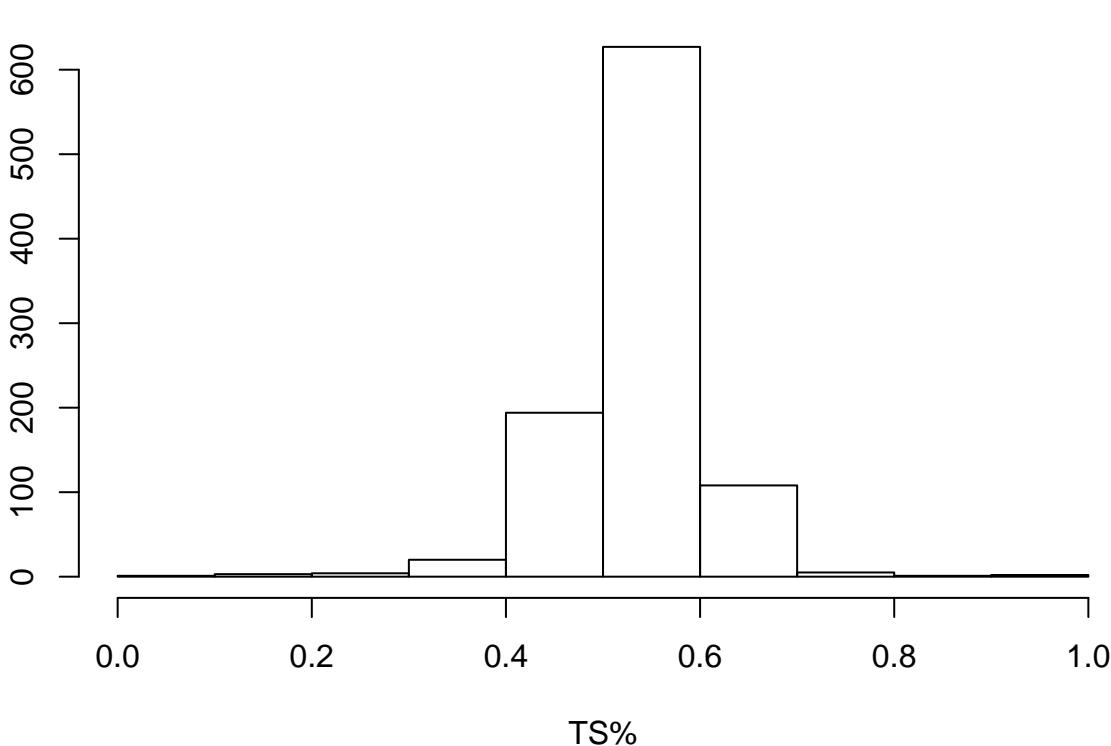
```
## Top 10 Players by MP
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 James~ 2.65e7 SG     26 HOU    82   82  3125 25.3 0.598 0.406
## 2 2017 Andre~ 7.57e6 SF     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>, O BPM <dbl>,
## # DBPM <dbl>, BPM <dbl>, VOR P <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



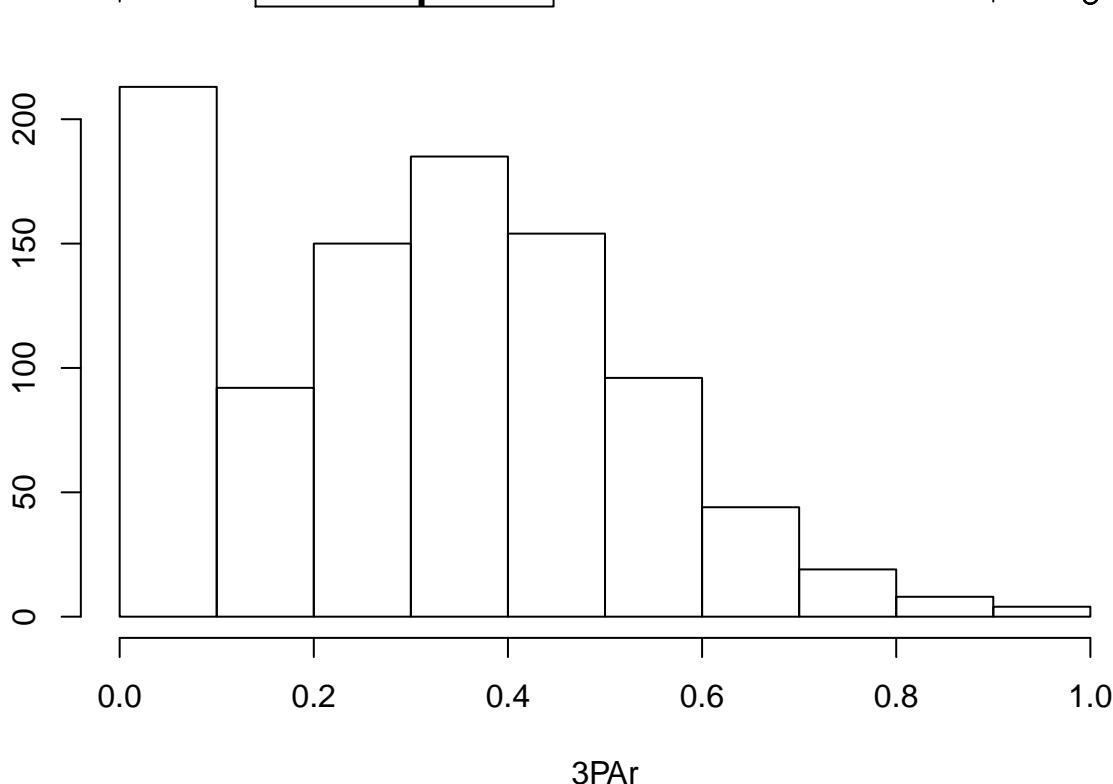
```
## Top 10 Players by PER
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS    MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Brian~ 3.28e5 PG      23 MIA     1     0     3 39.3 1     0
## 2 2016 Rakee~ 1.05e6 PF     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>, O BPM <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%



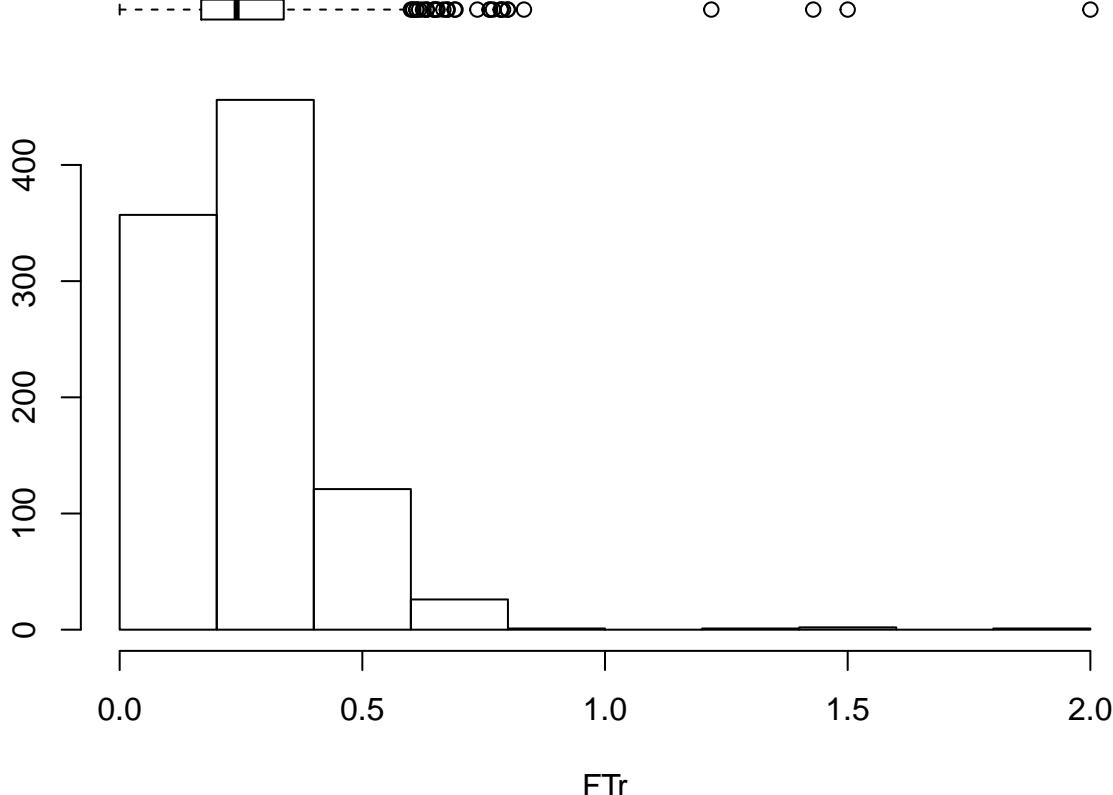
```
## Top 10 Players by TS%
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Brian~ 3.28e5 PG      23 MIA     1     0     3 39.3 1     0
## 2 2016 Rakee~ 1.05e6 PF     24 IND     1     0     6 32    1     0
## 3 2017 Wayne~ 1.31e6 SG     22 NOP     3     3     47 10    0.82  0.875
## 4 2017 China~ 1.31e6 C      20 HOU     5     1     52 12.3 0.799 0
## 5 2017 Jarre~ 2.33e6 PG     33 NOP     2     0     33 7.7   0.773 0.333
## 6 2017 Demet~ 9.29e4 PG     22 BOS     5     0     17 30.8 0.753 0.25
## 7 2016 Steve~ 1.55e6 PF     32 OKC     7     0     24 20.8 0.708 0.75
## 8 2017 Tyson~ 1.30e7 C      34 PHO     47    46    1298 16.6 0.703 0
## 9 2017 Axel ~ 2.50e4 SF     24 NOP     2     0     41 8.6   0.688 0.375
## 10 2017 Lucas~ 2.95e6 C     24 TOR     57    6    1088 15.5 0.682 0.077
## # ... 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>, O BPM <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>
## 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>, O BPM <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 FTr

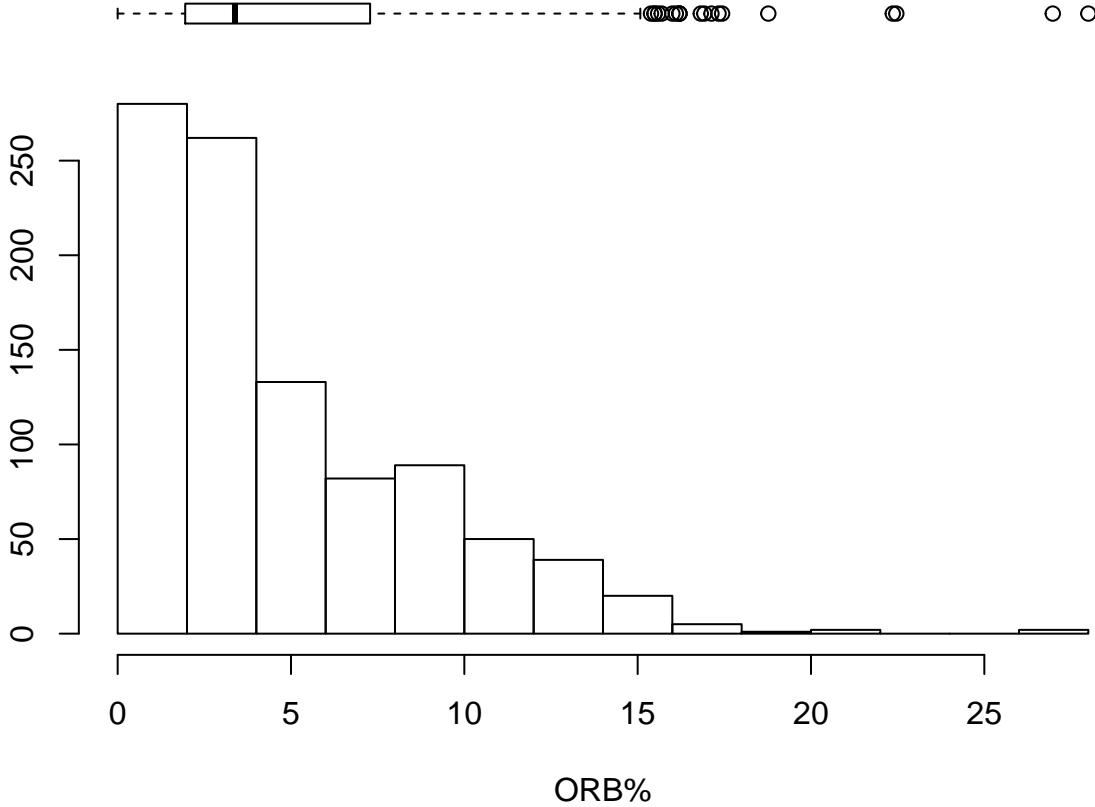


```

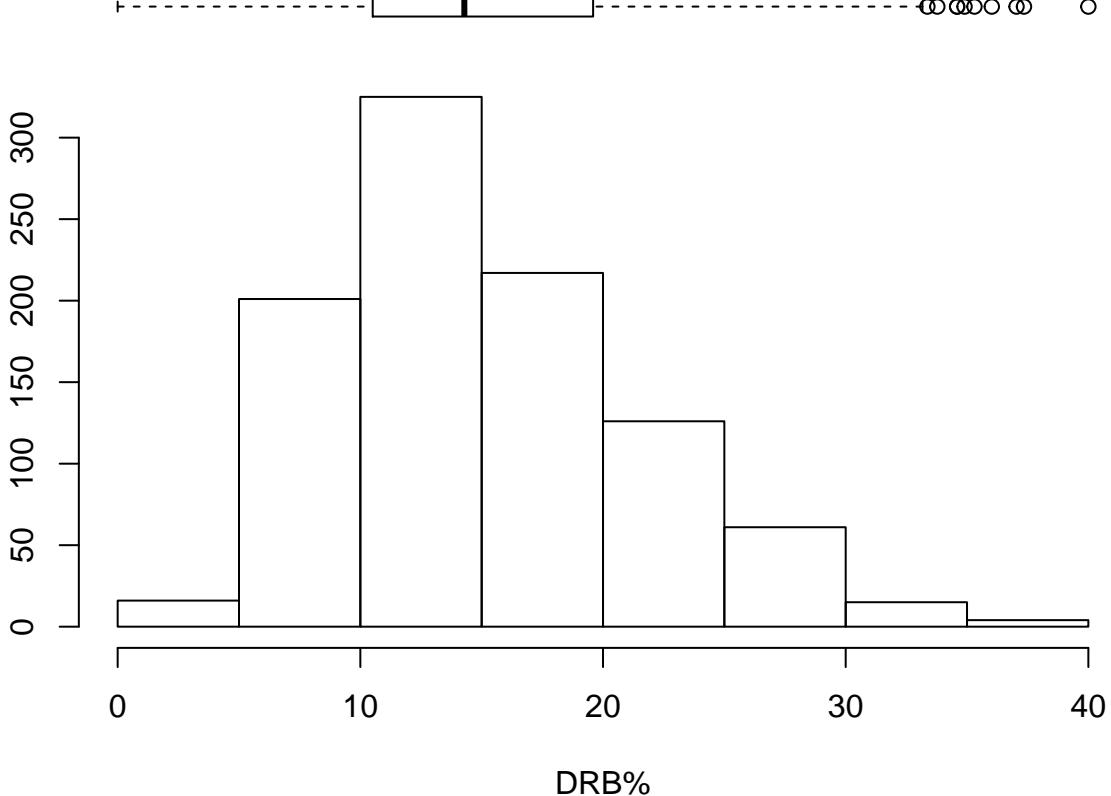
## Top 10 Players by FTr
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017  Chris~  1.47e6 PF     21 WAS     2     0     8   1.1  0.266  1
## 2 2017  Demet~  9.29e4 PG    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%

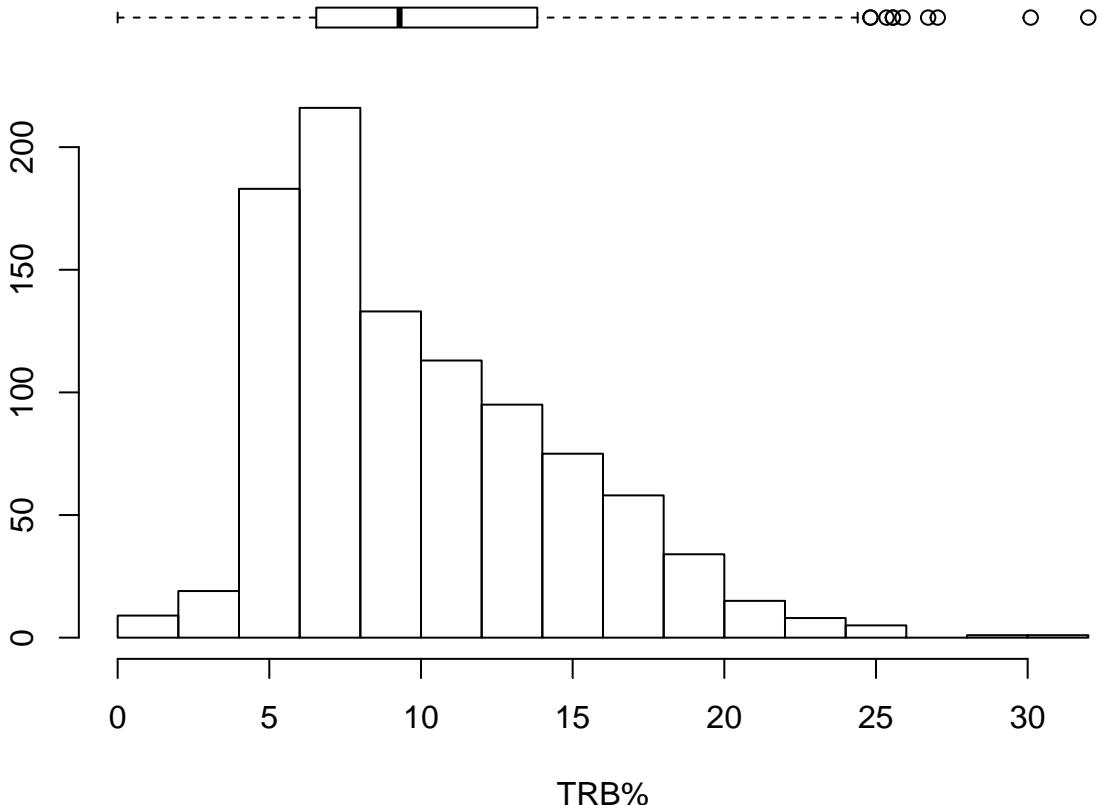


Histogram of DRB%



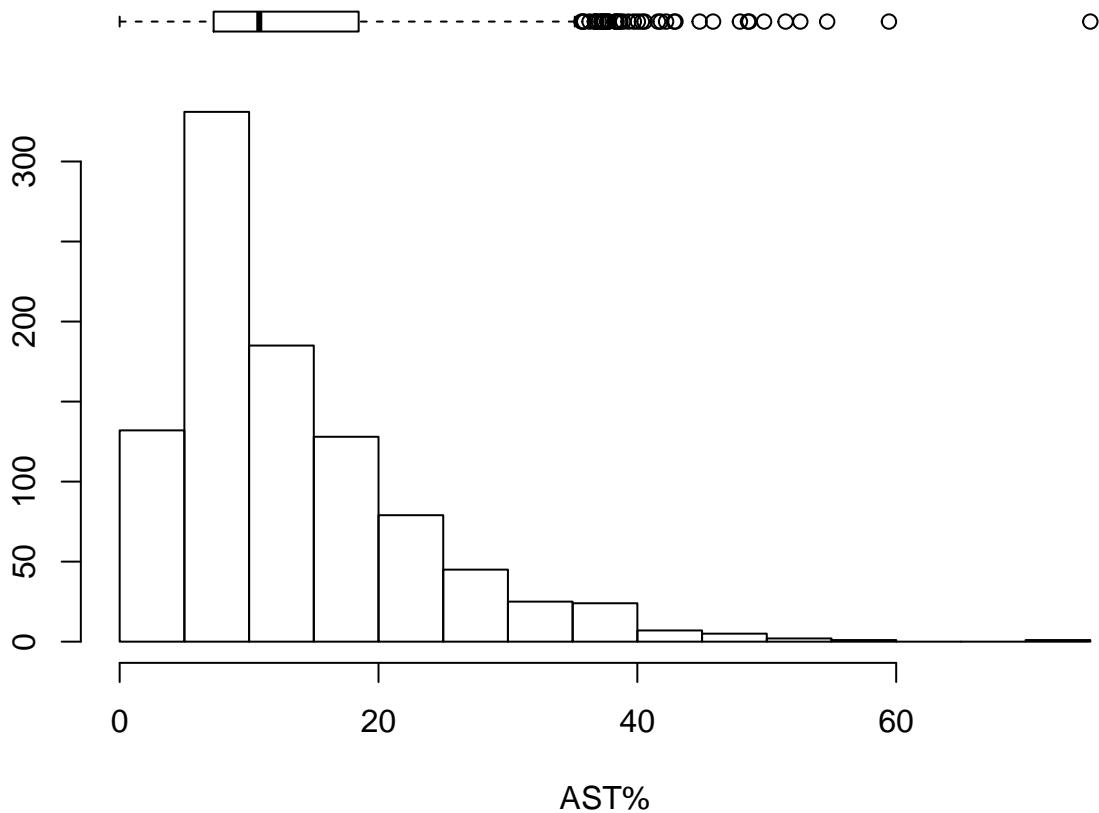
```
## Top 10 Players by DRB%
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Alan ~ 8.75e5 PF     23 PHO     10   0   68  21.1 0.481  0
## 2 2016 Brian~ 3.28e5 PG    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%



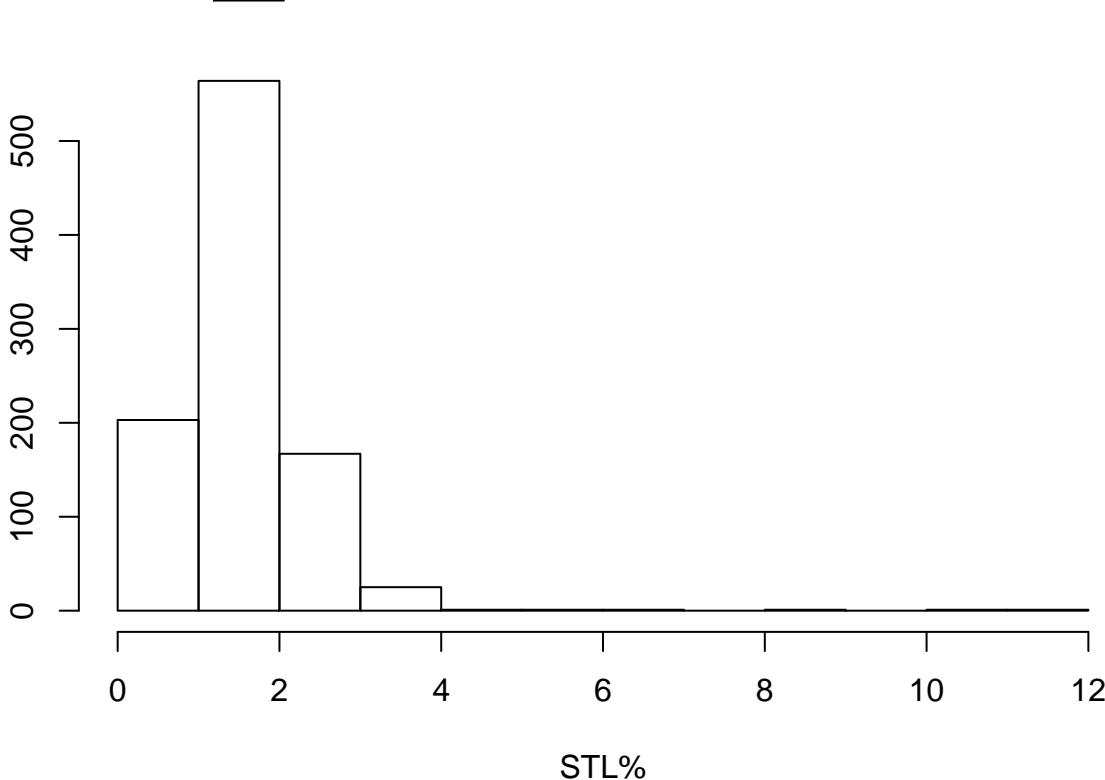
```
## Top 10 Players by TRB%
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Alan ~ 8.75e5 PF     23 PHO     10   0   68  21.1 0.481  0
## 2 2016 Jarne~ 1.50e5 C      22 MEM      2   0   4  13.6 0.515  0.191
## 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>, O BPM <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 AST%



```
## Top 10 Players by AST%
## # 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>
## 1 2016 Brian~ 3.28e5 PG      23 MIA     1    0    3 39.3 1     0
## 2 2017 Russe~ 2.85e7 PG      28 OKC    81   81 2802 30.6 0.554 0.3
## 3 2016 Chris~ 2.29e7 PG      30 LAC    74   74 2420 26.2 0.575 0.295
## 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 Rajon~ 1.40e7 PG      29 SAC    72   72 2537 16.9 0.506 0.217
## 7 2017 John ~ 1.81e7 PG      26 WAS    78   78 2836 23.2 0.541 0.19
## 8 2017 Chris~ 2.46e7 PG      31 LAC    61   61 1921 26.2 0.614 0.385
## 9 2016 John ~ 1.70e7 PG      25 WAS    77   77 2784 19.8 0.51  0.243
## 10 2017 J.J. ~ 3.90e6 PG     32 DAL   35    6 771 17.2 0.521 0.431
## # ... 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>, O BPM <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 STL%

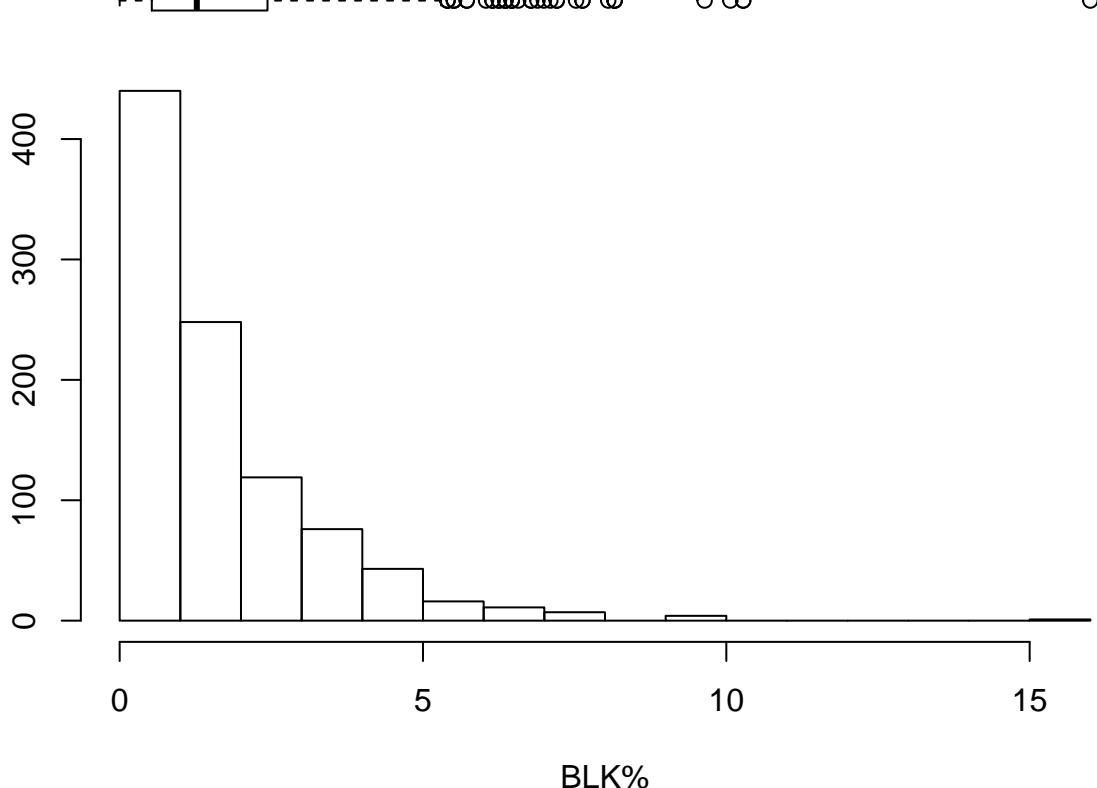


```

## 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>
## 1 2017 Brice~ 1.33e6 PF     22 LAC     3   0    9  17.2 0.286  0
## 2 2016 Jord~ 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 0.515  0.191
## 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>, O BPM <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>

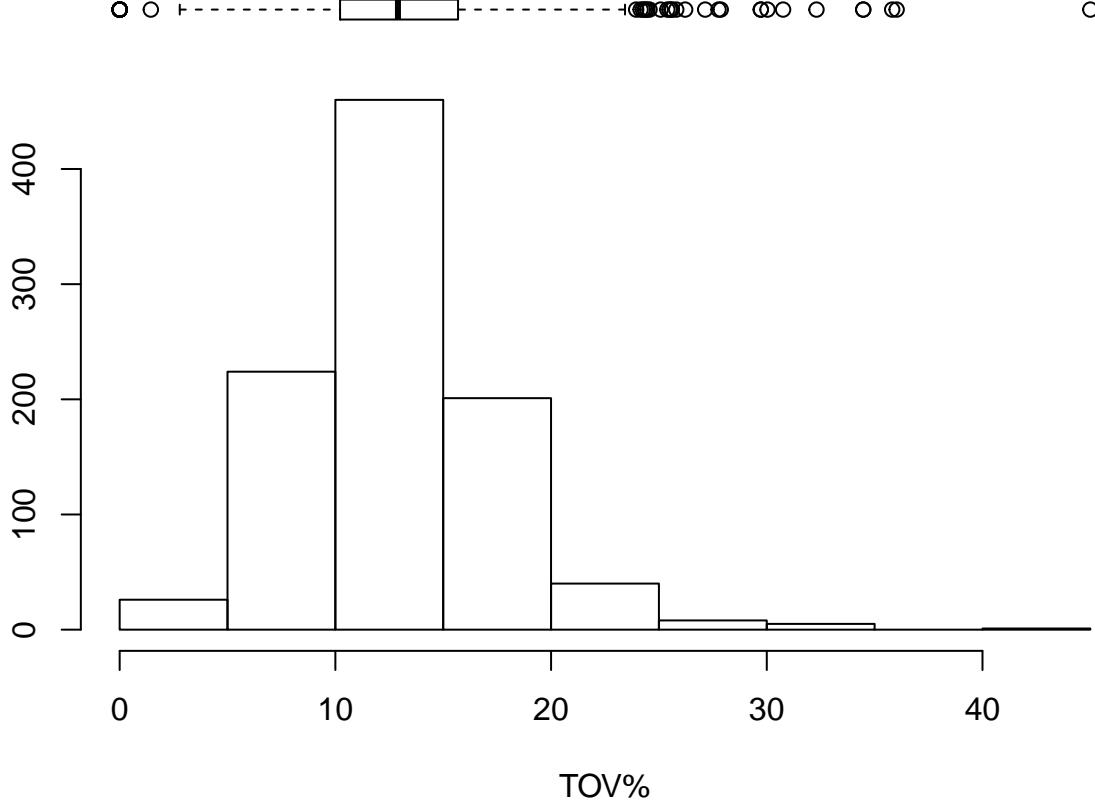
```

Histogram of BLK%



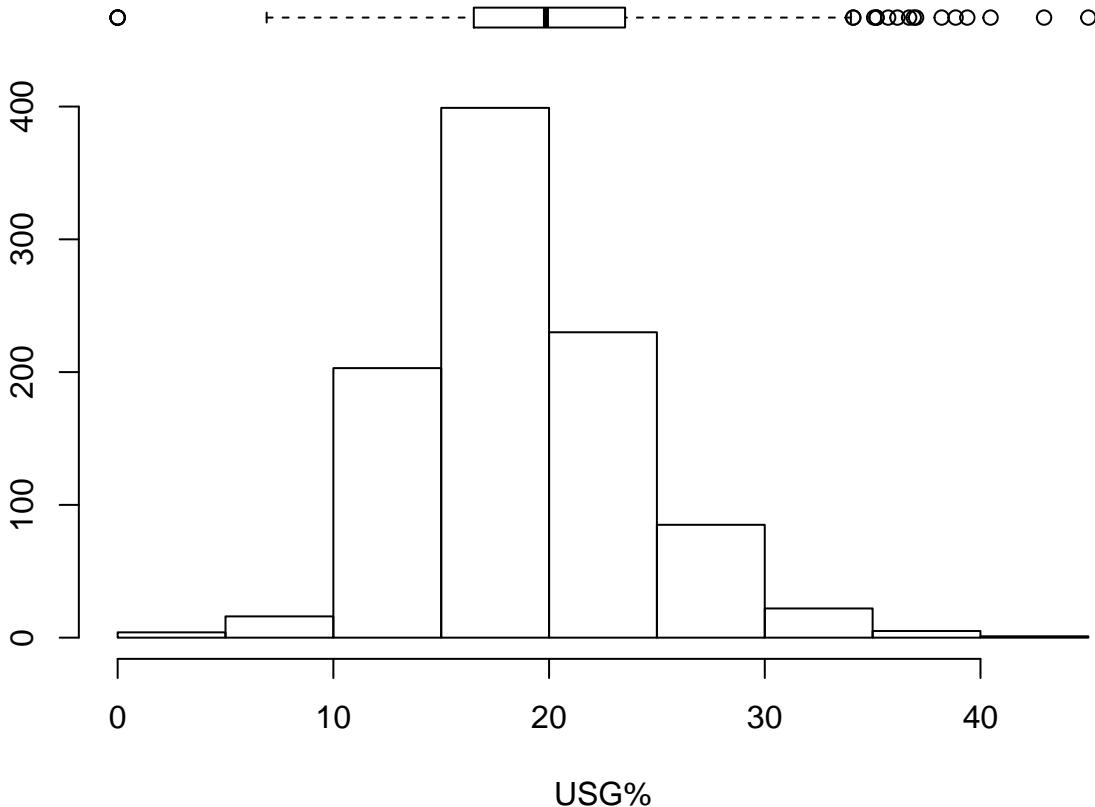
```
## Top 10 Players by BLK%
## # A tibble: 10 x 51
##   year name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr` 
##   <dbl> <chr>  <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Jordae~ 1.22e6 PF     21 BOS    16    0    57  15.3 0.398  0
## 2 2016 Hassae~ 2.21e7 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>
```

Histogram of TOV%



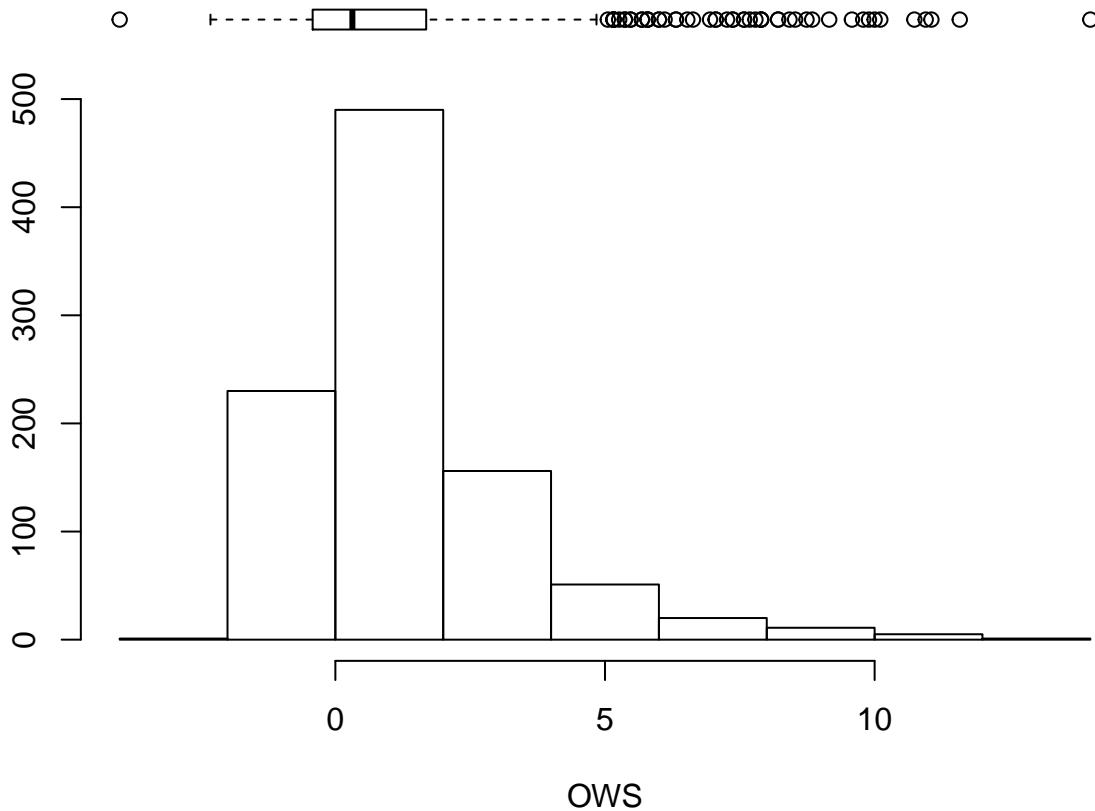
```
## Top 10 Players by TOV%
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017  Jarrett 2.33e6 PG      33 NOP     2    0    33   7.7  0.773  0.333
## 2 2016  Phil    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: 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>, O BPM <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%



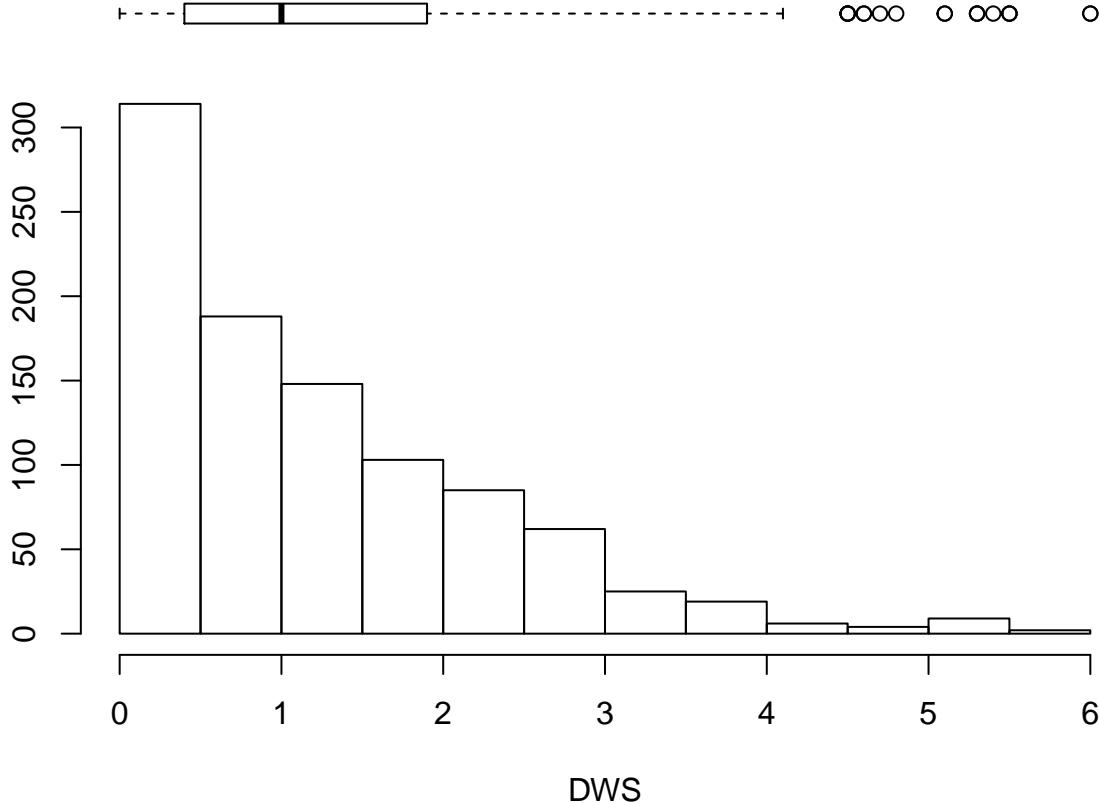
```
## Top 10 Players by USG%
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017  Russel~ 2.85e7 PG      28  OKC    81    81  2802 30.6 0.554  0.3
## 2 2017  Brice~  1.33e6 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  Isaias~ 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>, O BPM <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



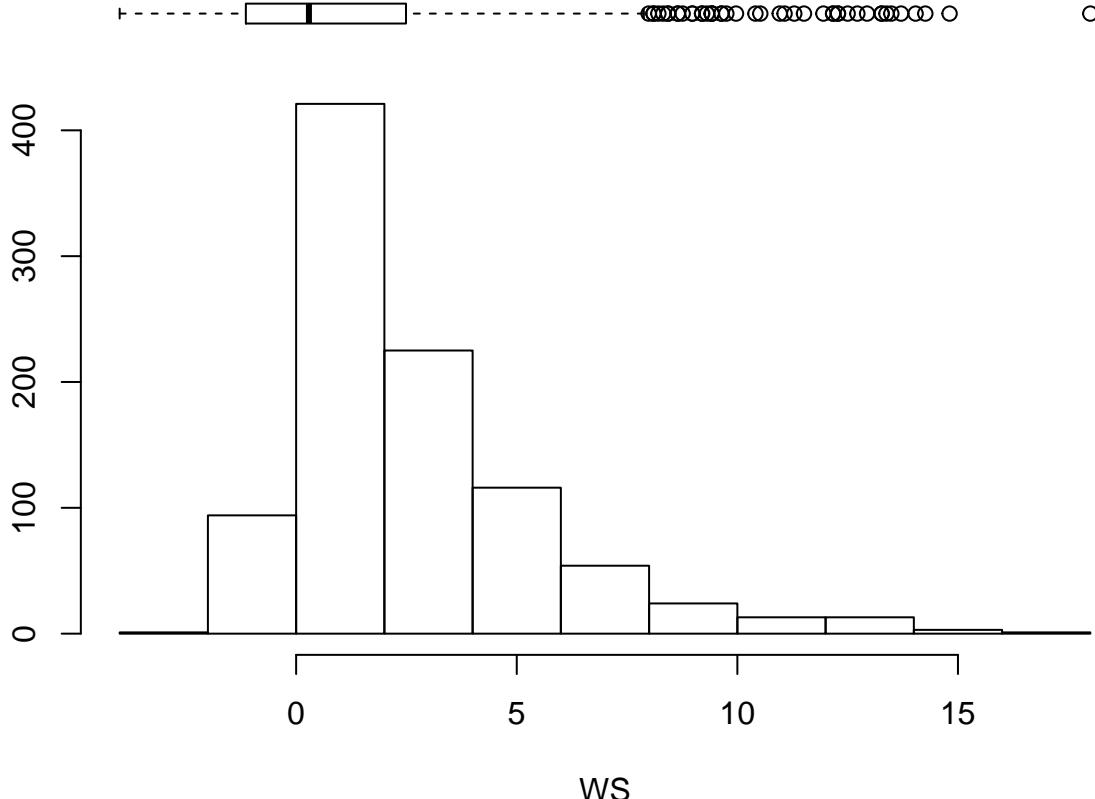
```
## Top 10 Players by OWS
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016  Steph~  1.21e7 PG     27 GSW    79    79  2700 31.5 0.669 0.554
## 2 2017  James~  2.83e7 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  Isaias~ 6.26e6 PG     27 BOS    76    76  2569 26.5 0.625 0.439
## 5 2016  James~  2.65e7 SG     26 HOU    82    82  3125 25.3 0.598 0.406
## 6 2017  Jimmy~  1.93e7 SF     27 CHI    76    75  2809 25.1 0.586 0.198
## 7 2016  Russel~ 2.65e7 PG     27 OKC    80    80  2750 27.6 0.554 0.236
## 8 2017  Karl~  6.22e6 C      21 MIN    82    82  3030 26   0.618 0.186
## 9 2017  LeBro~  3.33e7 SF     32 CLE    74    74  2794 27   0.619 0.254
## 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>, O BPM <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 DWS



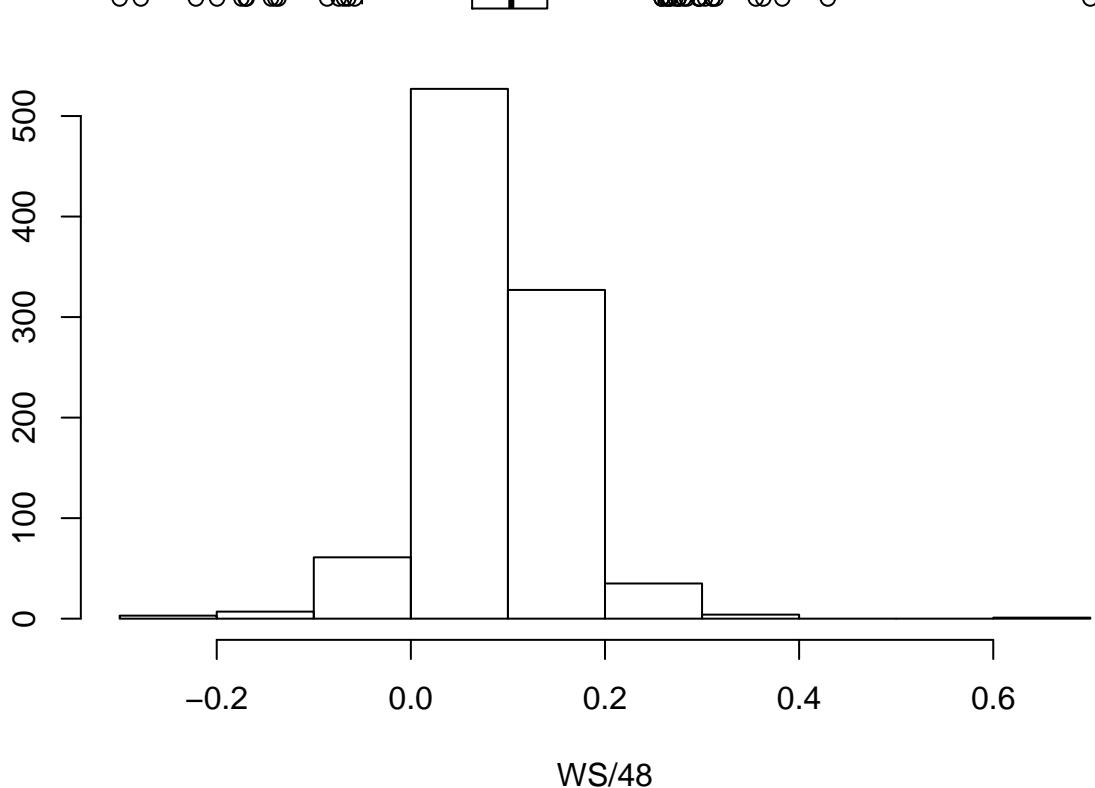
```
## Top 10 Players by DWS
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Paul ~ 2.01e7 PF     30 ATL     81   81 2647 21.3 0.556 0.218
## 2 2017 Rudy ~ 2.20e7 C      24 UTA     81   81 2744 23.3 0.682 0.002
## 3 2016 Andre~ 2.21e7 C     22 DET     81   81 2666 21.2 0.499 0.006
## 4 2016 DeAnd~ 2.12e7 C     27 LAC     77   77 2598 20.6 0.628 0.002
## 5 2016 Kawhi~ 1.76e7 SF    24 SAS     72   72 2380 26   0.616 0.267
## 6 2017 Draym~ 1.64e7 PF    26 GSW     76   76 2471 16.5 0.522 0.405
## 7 2017 Andre~ 2.38e7 C     23 DET     81   81 2409 20.9 0.518 0.008
## 8 2016 Hassa~ 2.21e7 C     26 MIA     73   43 2125 25.7 0.629 0
## 9 2017 Hassa~ 2.38e7 C     27 MIA     77   77 2513 22.6 0.579 0
## 10 2017 Antho~ 2.38e7 C    23 NOP     75   75 2708 27.5 0.579 0.088
## # ... 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>, O BPM <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



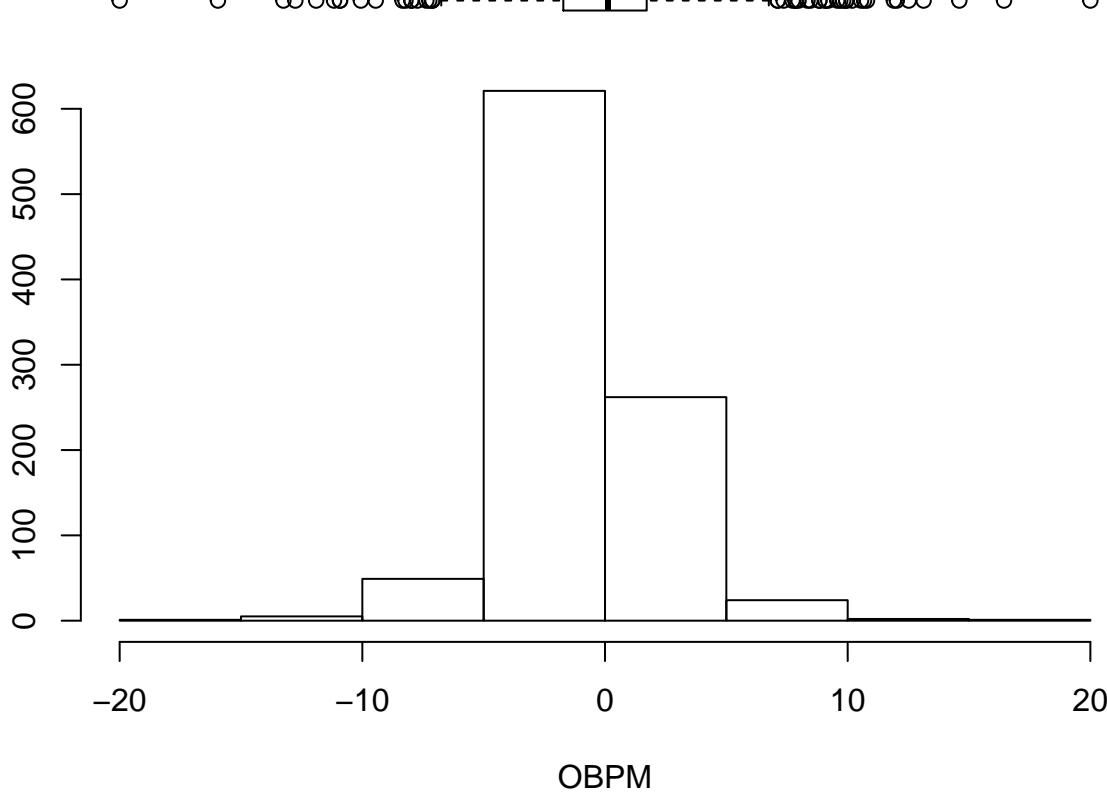
```
## Top 10 Players by WS
## # A tibble: 10 x 51
##   year name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>  <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Steph~ 1.21e7 PG     27 GSW    79   79  2700 31.5 0.669 0.554
## 2 2017 James~ 2.83e7 PG     27 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>, O BPM <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>
## 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>, O BPM <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 OBPM

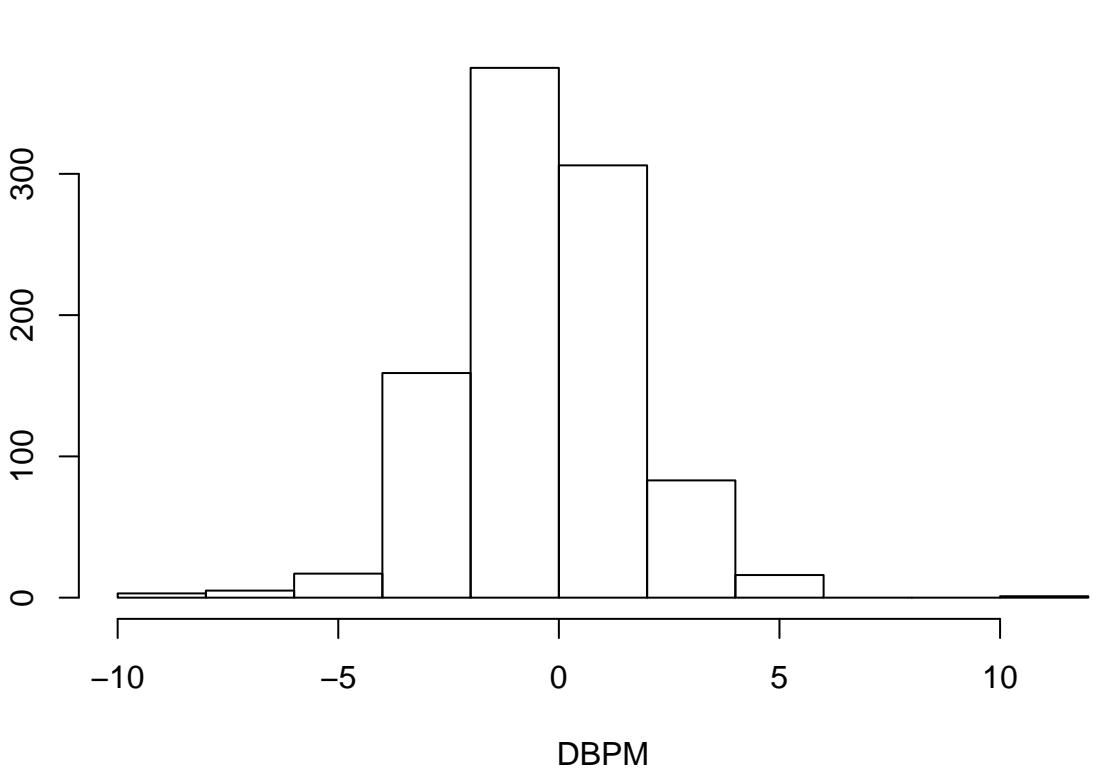


```

## Top 10 Players by OBPM
## # 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>
## 1 2016 Brian~  3.28e5 PG     23 MIA     1    0    3  39.3 1     0
## 2 2016 Steph~  1.21e7 PG     27 GSW    79   79  2700 31.5 0.669 0.554
## 3 2017 Russe~  2.85e7 PG     28 OKC    81   81  2802 30.6 0.554 0.3
## 4 2016 Rakee~  1.05e6 PF     24 IND     1    0    6  32   1     0
## 5 2017 Demet~  9.29e4 PG     22 BOS     5    0    17 30.8 0.753 0.25
## 6 2017 Isaia~  6.26e6 PG     27 BOS    76   76  2569 26.5 0.625 0.439
## 7 2017 James~  2.83e7 PG     27 HOU    81   81  2947 27.3 0.613 0.493
## 8 2017 Chris~  2.46e7 PG     31 LAC    61   61  1921 26.2 0.614 0.385
## 9 2017 Steph~  3.47e7 PG     28 GSW    79   79  2638 24.6 0.624 0.547
## 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>

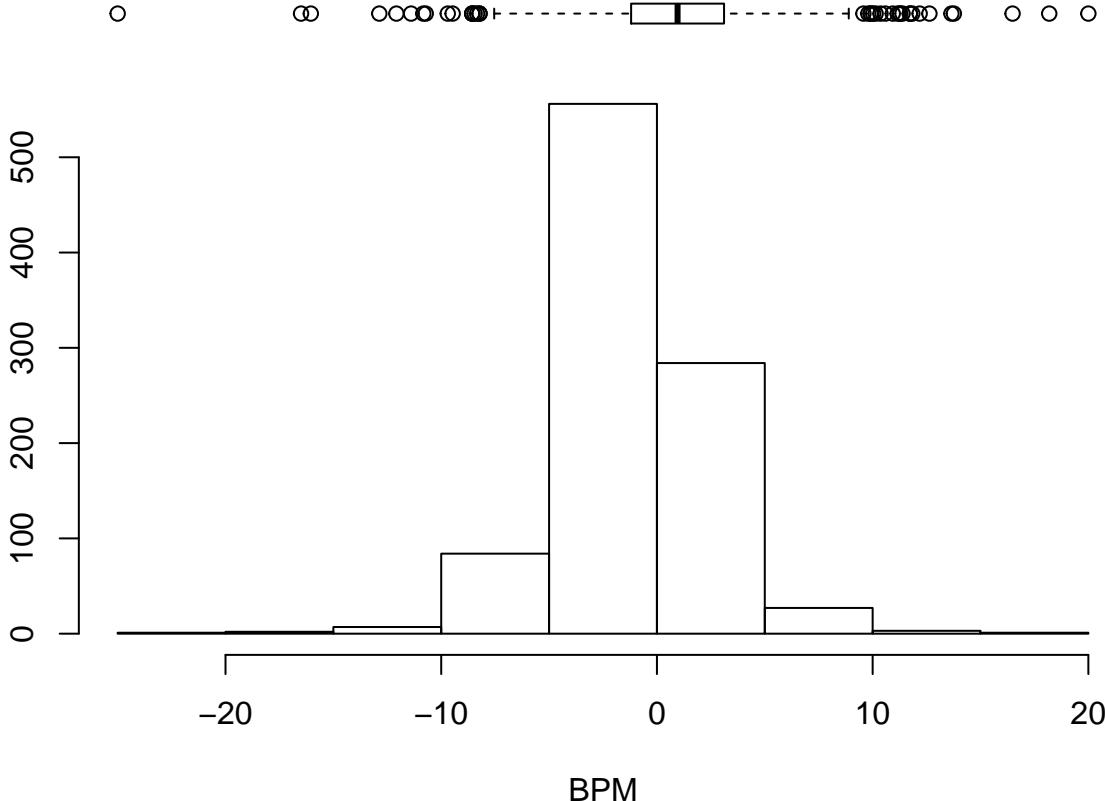
```

Histogram of DBPM



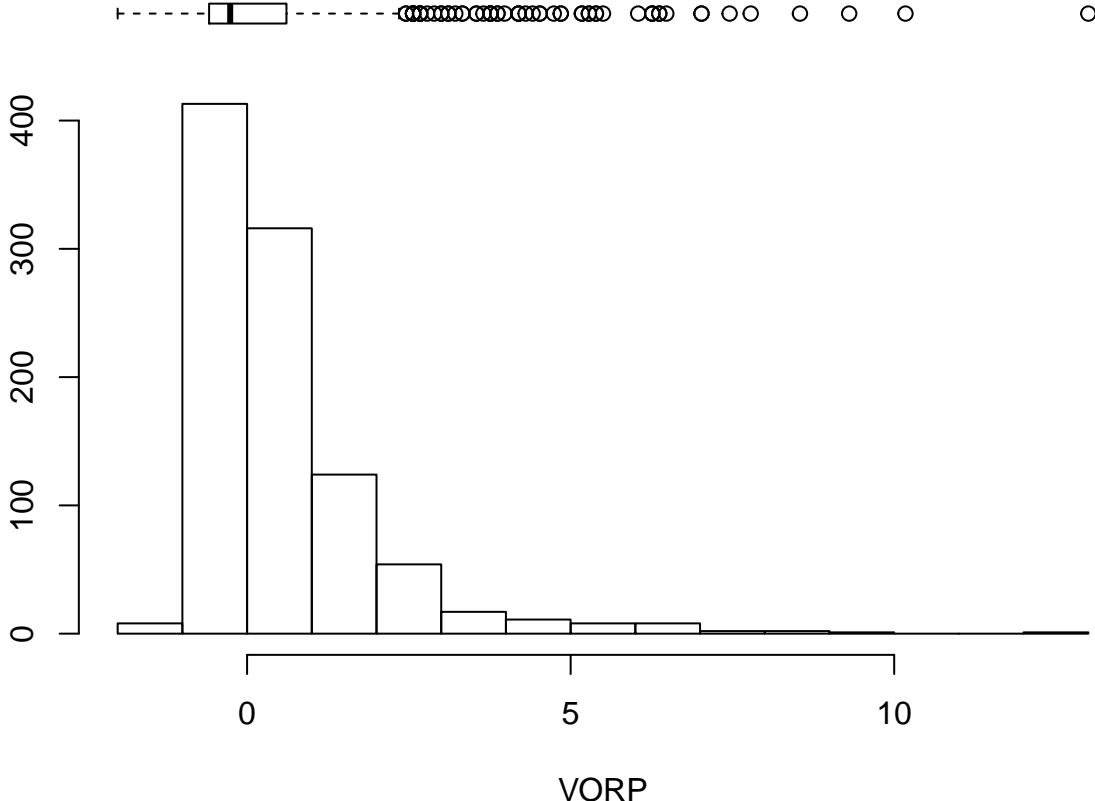
```
## Top 10 Players by DBPM
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS    MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Brice~ 1.33e6 PF     22 LAC     3     0     9 17.2 0.286  0
## 2 2016 Cole ~ 7.64e6 C      27 LAC    60     5   800 21.3 0.626  0
## 3 2016 Sam D~ 1.72e6 SF    21 HOU     3     0     6 10.8 0.515  0.191
## 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>, O BPM <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>
## 1 2017  Russel~ 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  Russel~ 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>, O BPM <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 VORP

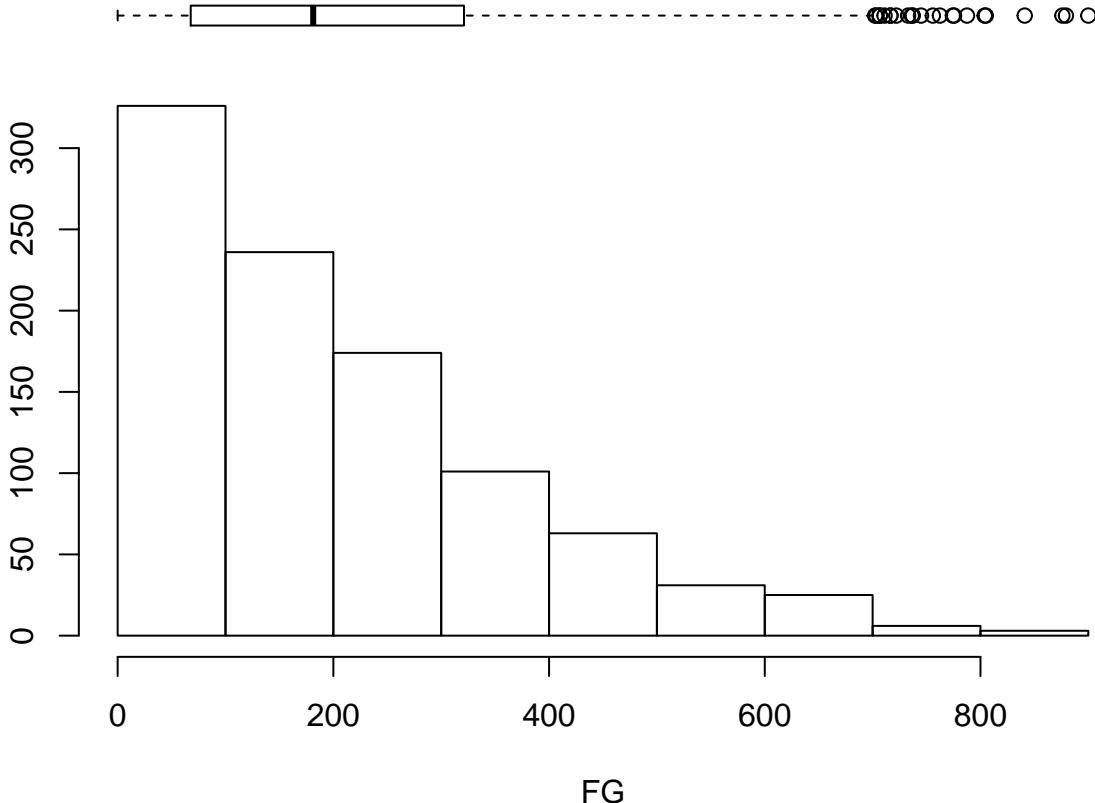


```

## Top 10 Players by VORP
## # 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>
## 1 2017  Russel~ 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  James~ 2.83e7 PG      27 HOU    81    81 2947 27.3 0.613 0.493
## 4 2016  Russel~ 2.65e7 PG      27 OKC    80    80 2750 27.6 0.554 0.236
## 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  Giann~ 2.25e7 SF      22 MIL    80    80 2845 26.1 0.599 0.143
## 8 2016  James~ 2.65e7 SG      26 HOU    82    82 3125 25.3 0.598 0.406
## 9 2016  Kevin~ 2.65e7 SF      27 OKC    72    72 2578 28.2 0.634 0.348
## 10 2017 Jimmy~ 1.93e7 SF      27 CHI    76    75 2809 25.1 0.586 0.198
## # ... 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>, O BPM <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>

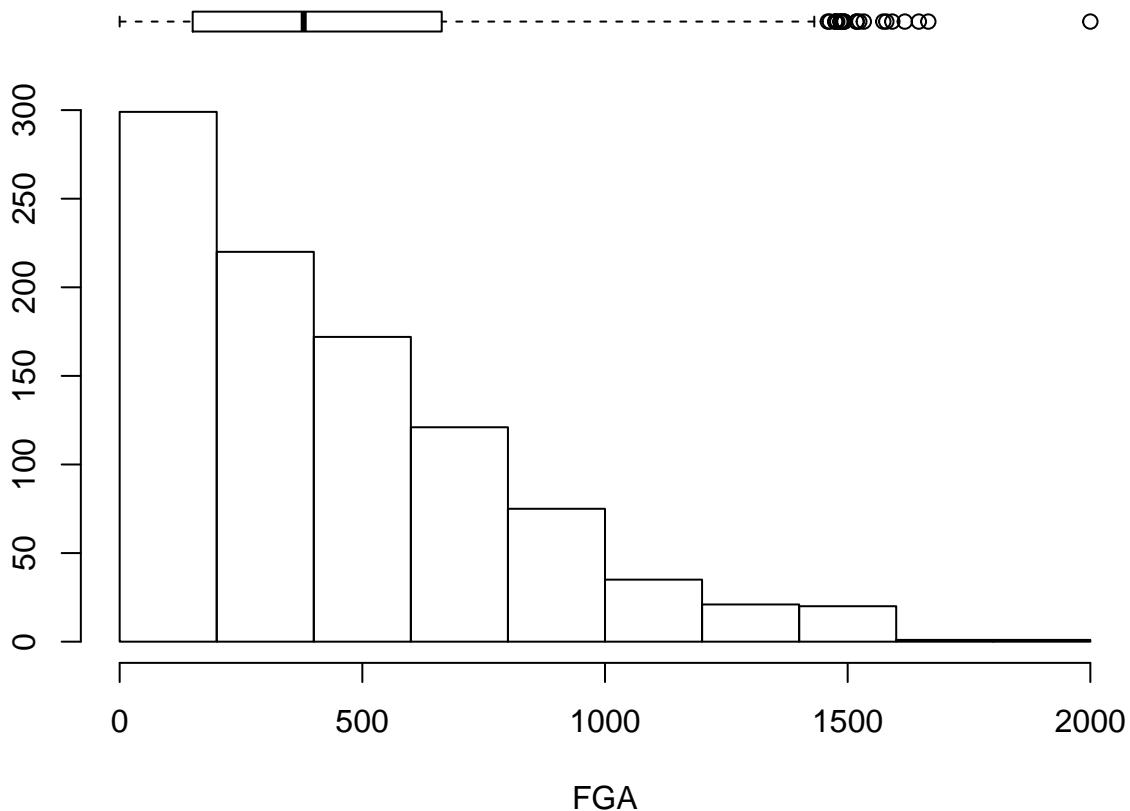
```

Histogram of FG



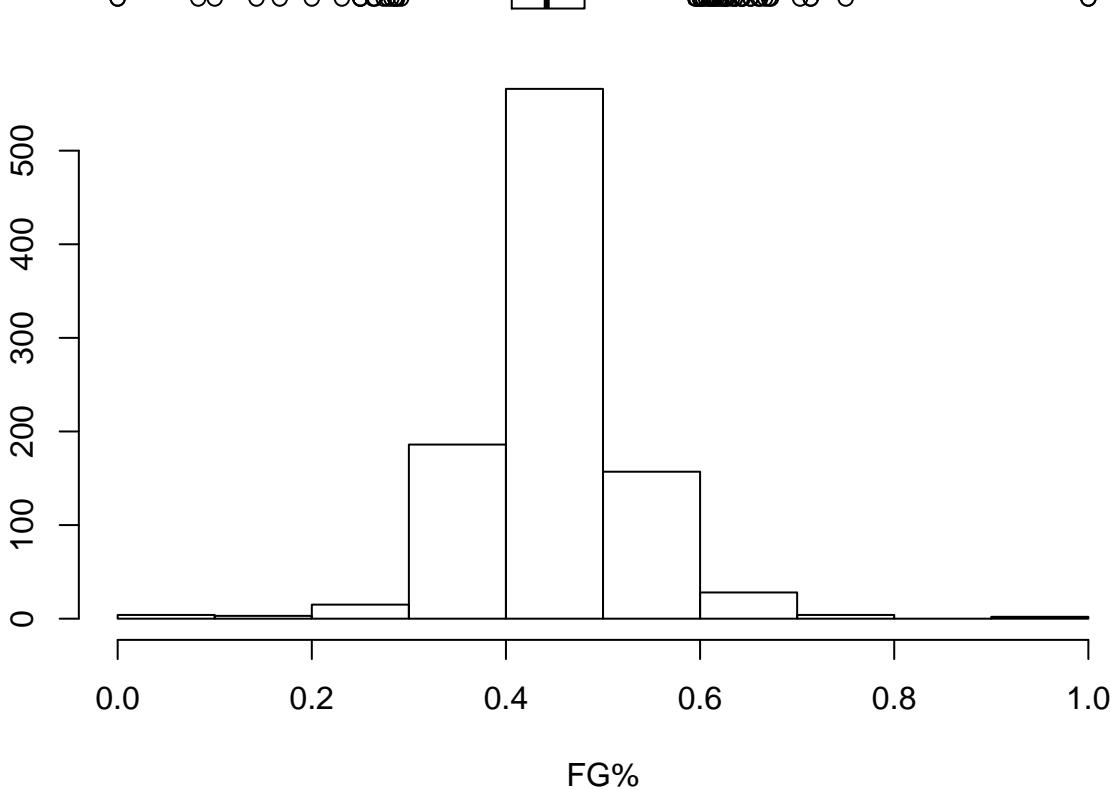
```
## Top 10 Players by FG
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS    MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017  Russel~ 2.85e7 PG      28 OKC    81    81 2802 30.6 0.554 0.3
## 2 2016  Steph~ 1.21e7 PG      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>, O BPM <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



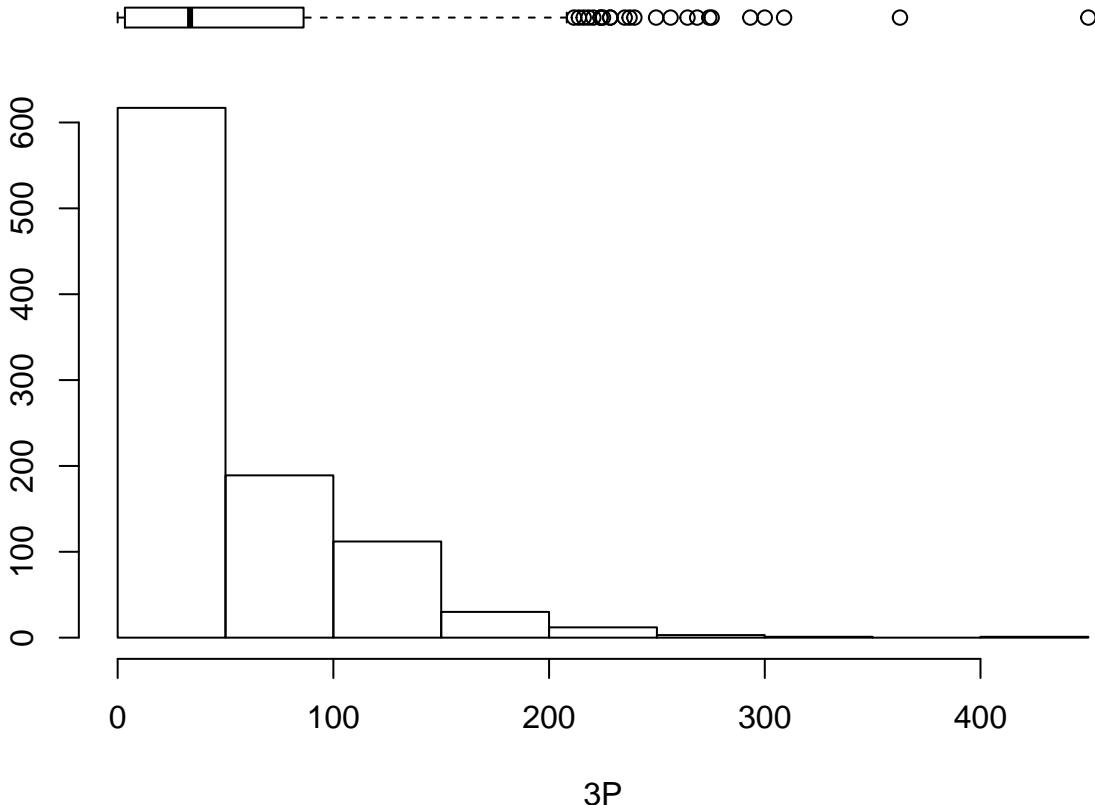
```
## Top 10 Players by FGA
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017  Russel~ 2.85e7 PG     28 OKC    81   81 2802 30.6 0.554 0.3
## 2 2016  James~   2.65e7 SG    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  Anthro~   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>, O BPM <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%



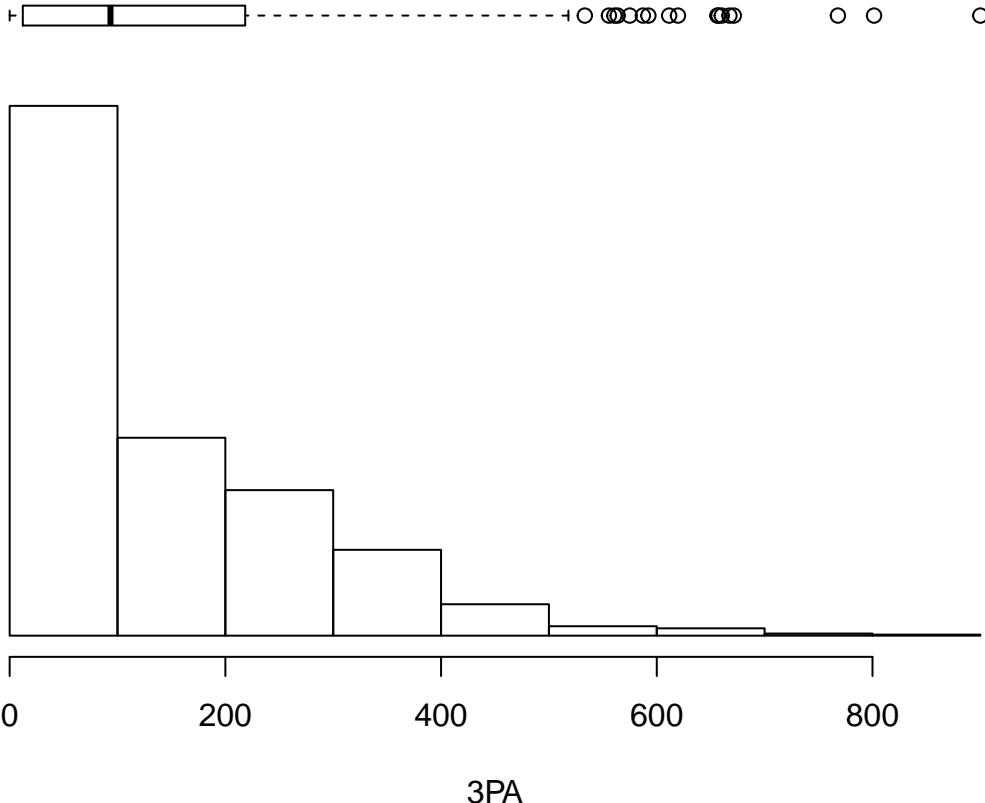
```
## Top 10 Players by FG%
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS    MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Brian~ 3.28e5 PG      23 MIA     1    0    3 39.3 1     0
## 2 2016 Rakee~ 1.05e6 PF     24 IND     1    0    6 32    1     0
## 3 2017 Demet~ 9.29e4 PG     22 BOS     5    0   17 30.8 0.753 0.25
## 4 2017 China~ 1.31e6 C      20 HOU     5    1   52 12.3 0.799 0
## 5 2017 DeAnd~ 2.26e7 C      28 LAC    81   81 2570 21.8 0.673 0.003
## 6 2016 DeAnd~ 2.12e7 C      27 LAC    77   77 2598 20.6 0.628 0.002
## 7 2016 Brand~ 5.70e6 PF     28 MEM    12    2   212 18.3 0.663 0
## 8 2017 Tyson~ 1.30e7 C      34 PHO    47   46 1298 16.6 0.703 0
## 9 2017 Jarre~ 2.33e6 PG     33 NOP     2    0   33 7.7  0.773 0.333
## 10 2017 Rudy ~ 2.20e7 C     24 UTA    81   81 2744 23.3 0.682 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>, O BPM <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



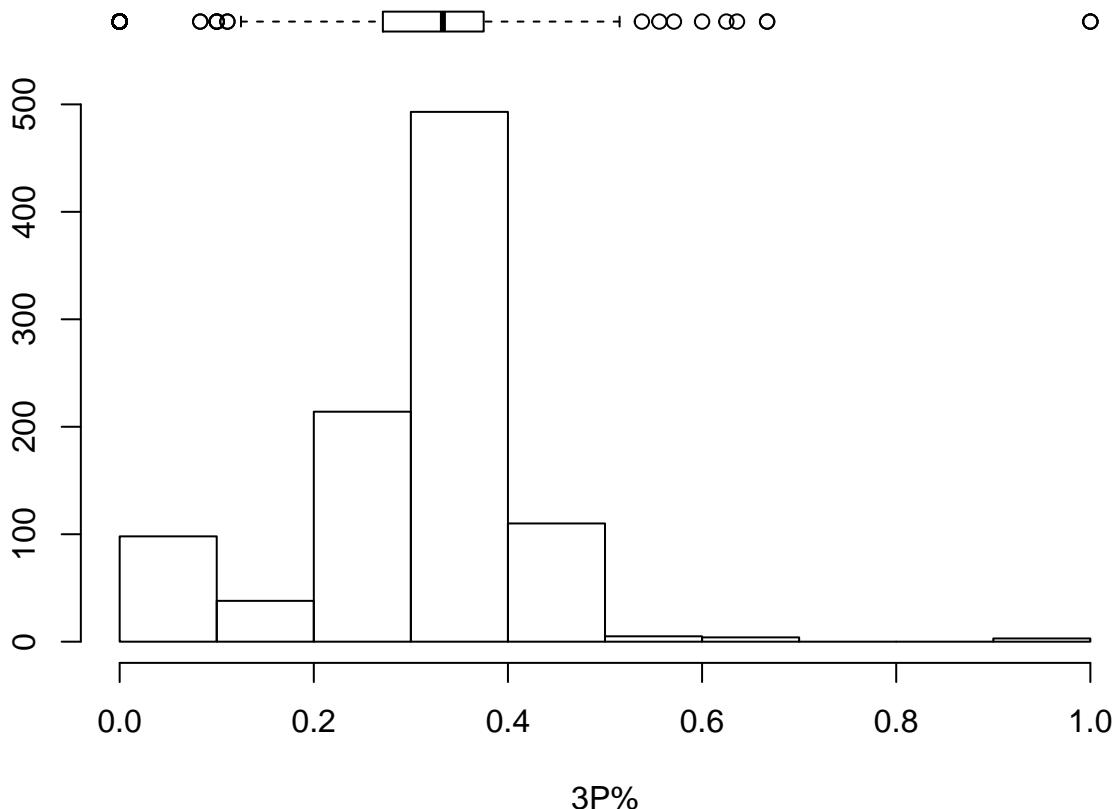
```
## Top 10 Players by 3P
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016  Steph~ 1.21e7 PG     27 GSW    79    79  2700 31.5 0.669 0.554
## 2 2017  Steph~ 3.47e7 PG     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  Isaias~ 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>, O BPM <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 3PA



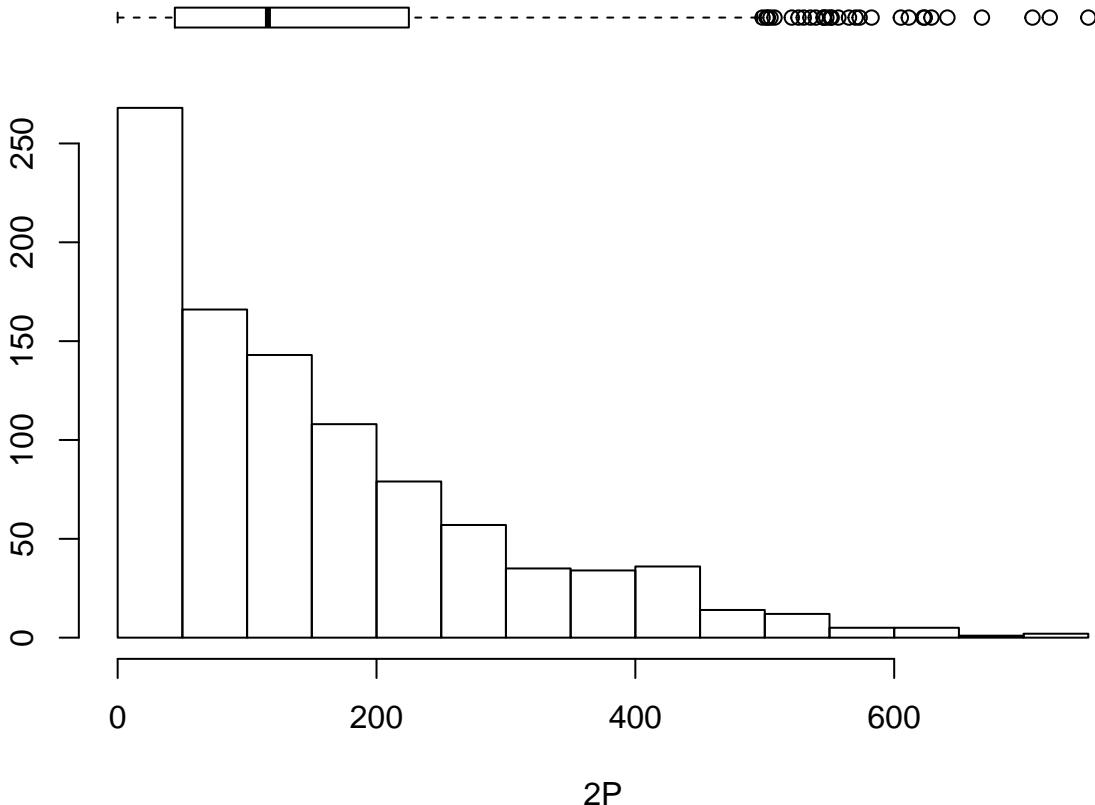
```
## Top 10 Players by 3PA
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016  Steph~ 1.21e7 PG      27 GSW    79    79 2700 31.5 0.669 0.554
## 2 2017  Steph~ 3.47e7 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>, O BPM <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%



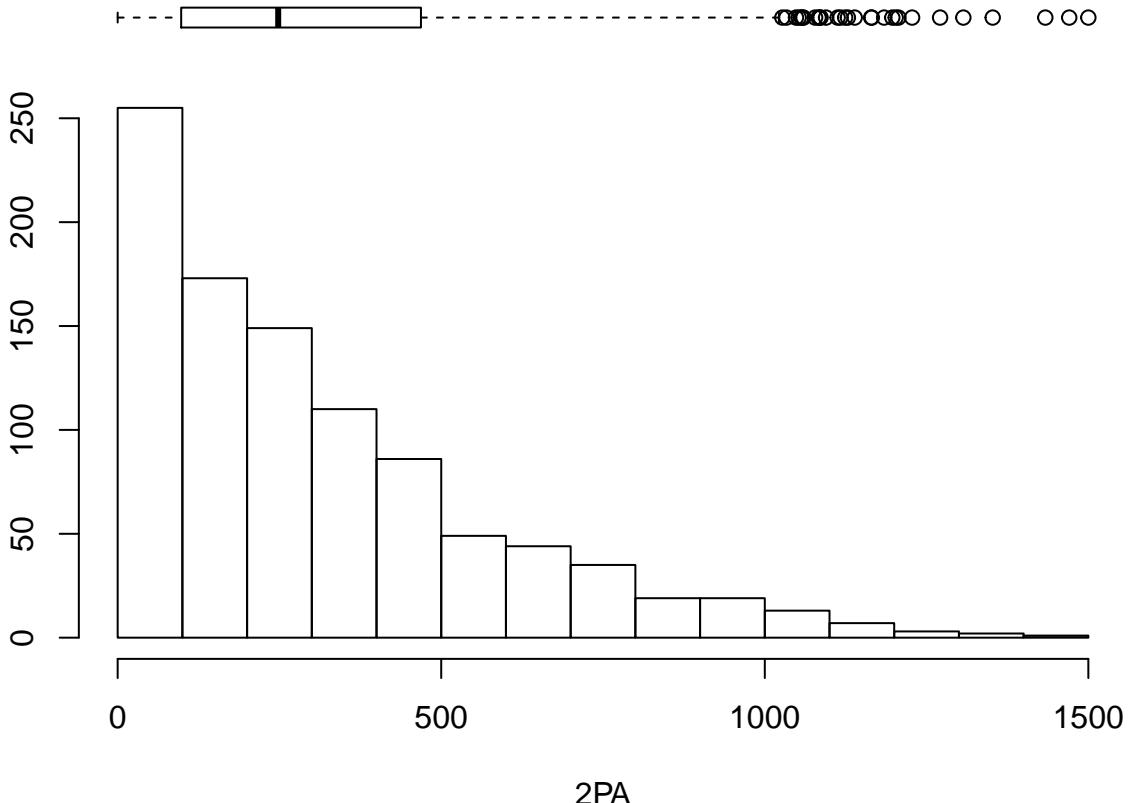
```
## Top 10 Players by 3P%
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS    MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Andre~ 1.10e7 C       31 GSW    70   66 1451 15.9 0.623 0.004
## 2 2017 Demet~ 9.29e4 PG     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 Jord~ 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>, O BPM <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



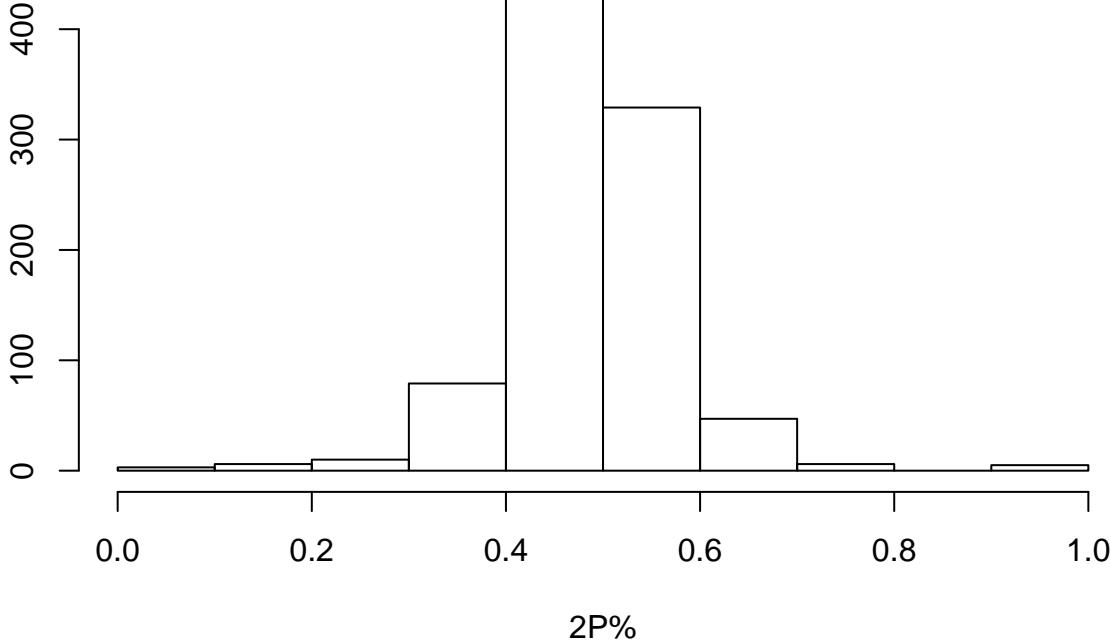
```
## Top 10 Players by 2P
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Antho~ 2.38e7 C       23 NOP     75    75 2708 27.5 0.579 0.088
## 2 2017 Karl~ 6.22e6 C       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>, O BPM <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



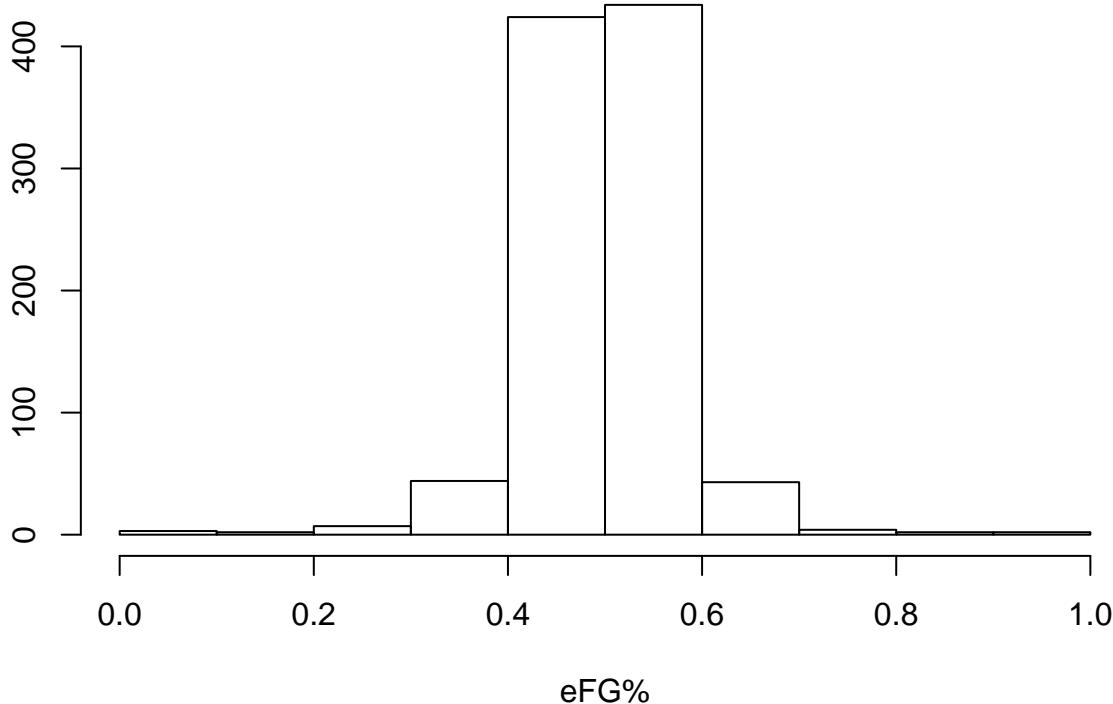
```
## Top 10 Players by 2PA
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 DeMar~ 2.77e7 SG      27 TOR     74   74 2620  24  0.552  0.08
## 2 2017 Antho~ 2.38e7 C       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>, O BPM <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%



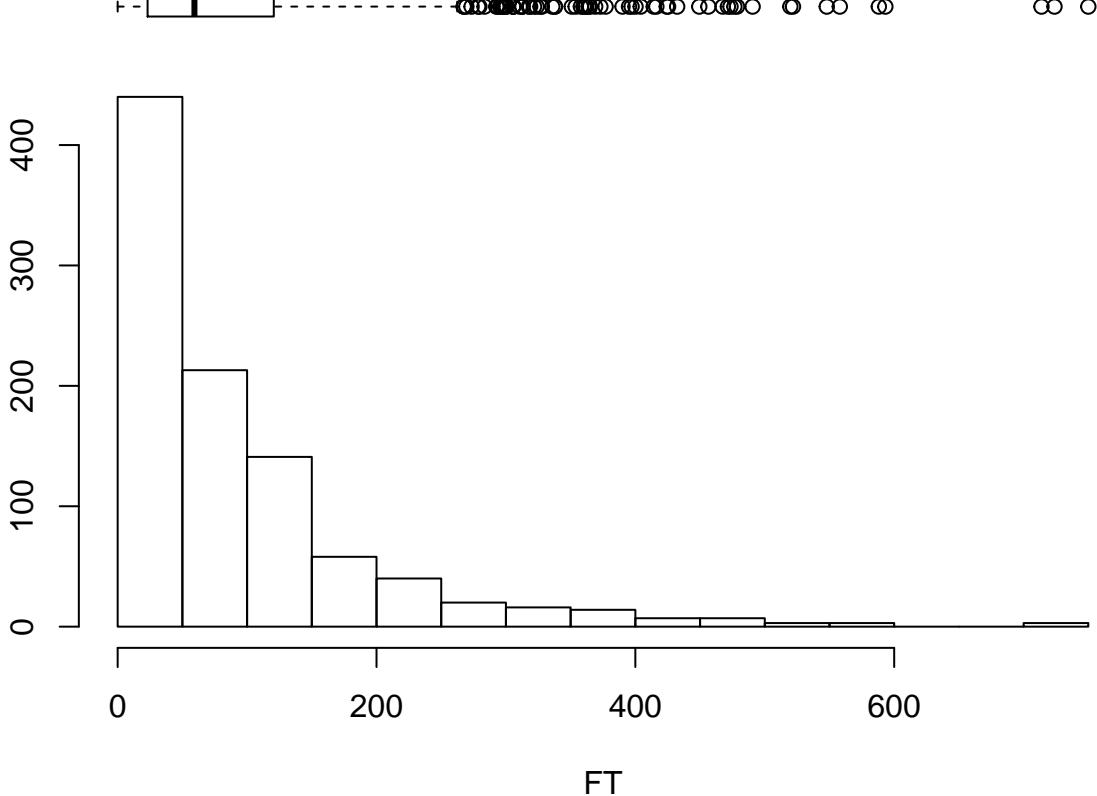
```
## Top 10 Players by 2P%
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Brian~ 3.28e5 PG     23 MIA     1     0     3 39.3 1     0
## 2 2017 Jarre~ 2.33e6 PG     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%



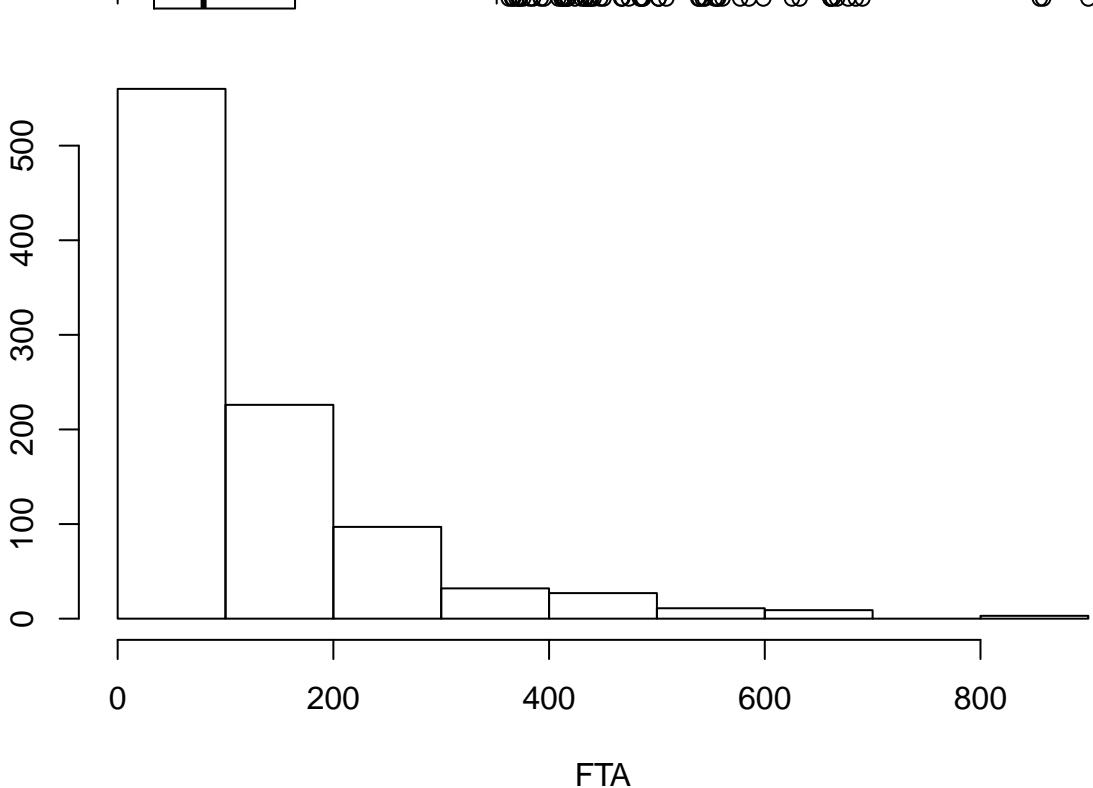
```
## Top 10 Players by eFG%
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Brian~ 3.28e5 PG      23 MIA     1    0    3 39.3 1     0
## 2 2016 Rakee~ 1.05e6 PF     24 IND     1    0    6 32    1     0
## 3 2017 Demet~ 9.29e4 PG     22 BOS     5    0   17 30.8 0.753 0.25
## 4 2017 Wayne~ 1.31e6 SG     22 NOP     3    3   47 10    0.82  0.875
## 5 2017 China~ 1.31e6 C      20 HOU     5    1   52 12.3 0.799 0
## 6 2017 DeAnd~ 2.26e7 C      28 LAC    81   81 2570 21.8 0.673 0.003
## 7 2016 Steve~ 1.55e6 PF     32 OKC     7    0   24 20.8 0.708 0.75
## 8 2016 DeAnd~ 2.12e7 C      27 LAC    77   77 2598 20.6 0.628 0.002
## 9 2017 Axel ~ 2.50e4 SF     24 NOP     2    0   41 8.6   0.688 0.375
## 10 2016 Brand~ 5.70e6 PF     28 MEM    12   2   212 18.3 0.663 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>, O BPM <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



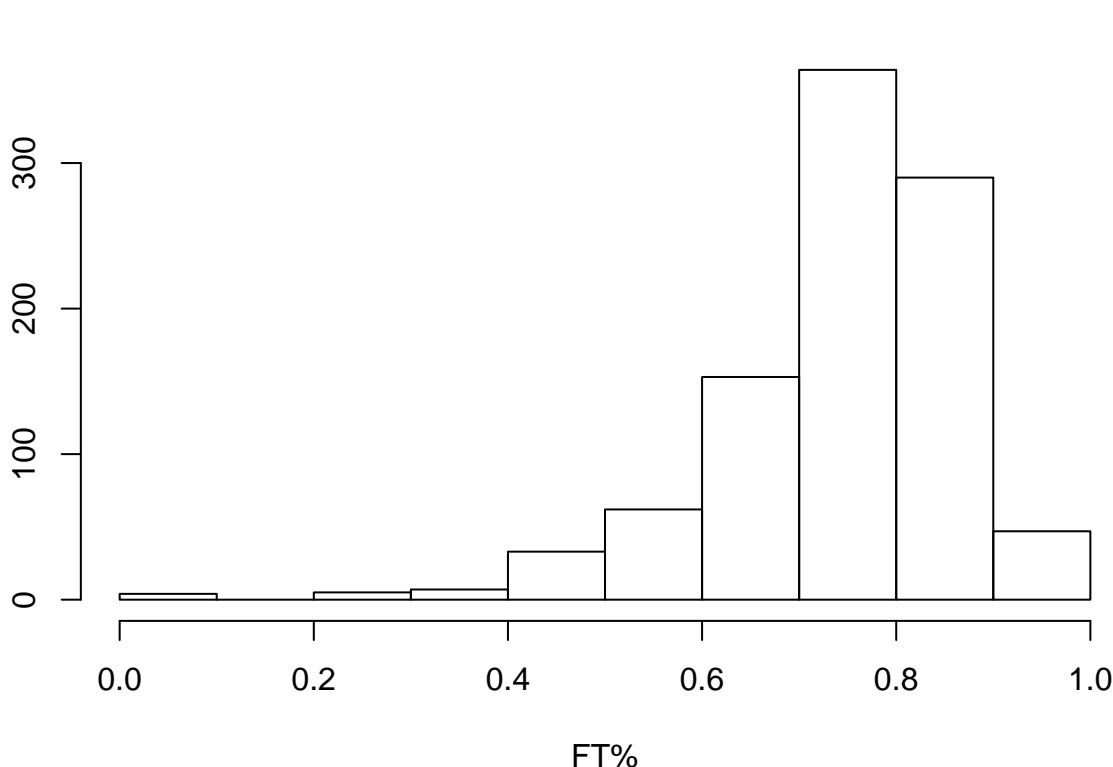
```
## Top 10 Players by FT
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 James~ 2.83e7 PG     27 HOU    81   81 2947 27.3 0.613 0.493
## 2 2016 James~ 2.65e7 SG     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



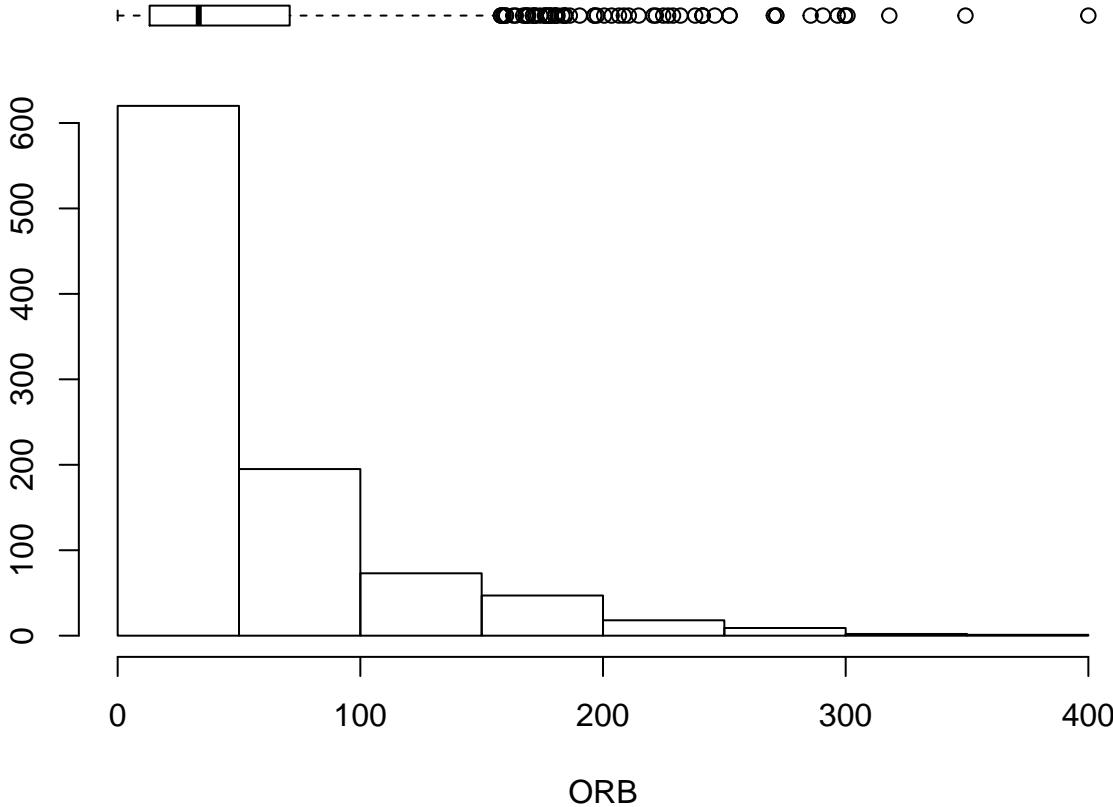
```
## Top 10 Players by FTA
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 James~ 2.83e7 PG     27 HOU    81   81 2947 27.3 0.613 0.493
## 2 2017 Russe~ 2.85e7 PG     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 Isai~ 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>, O BPM <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%



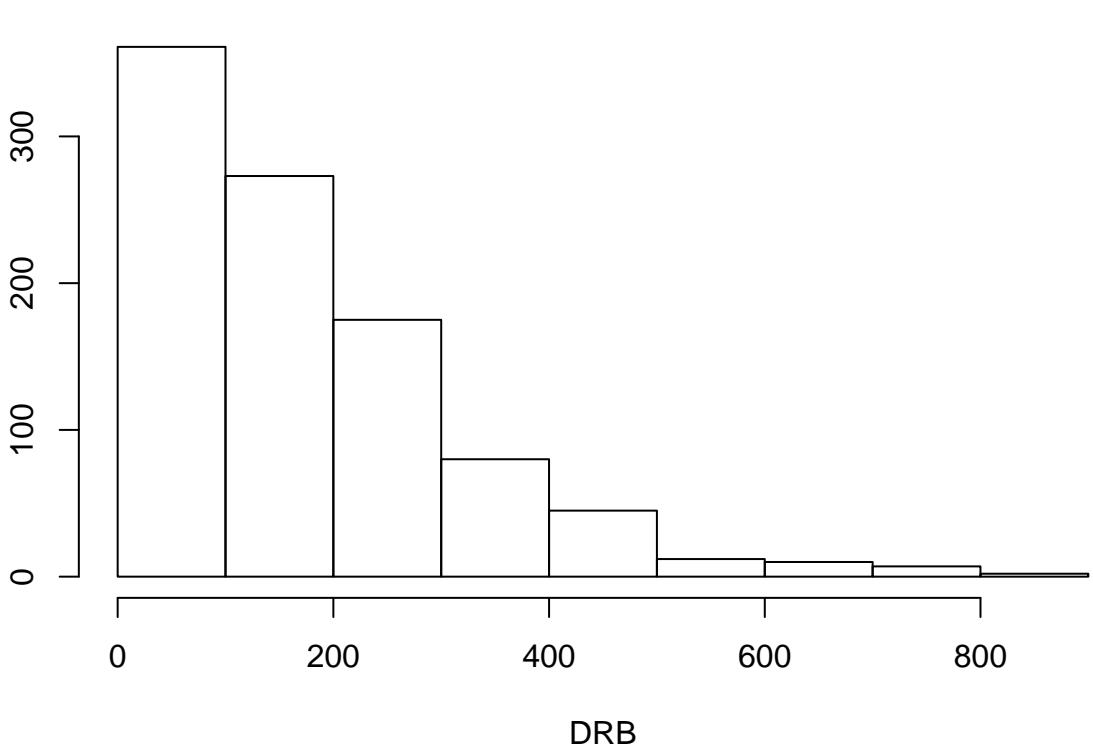
```
## Top 10 Players by FT%
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 Andre~ 2.84e6 PF     27 BRK     10   0   111   5  0.43  0.324
## 2 2016 Beno ~ 1.55e6 PG     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>, O BPM <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



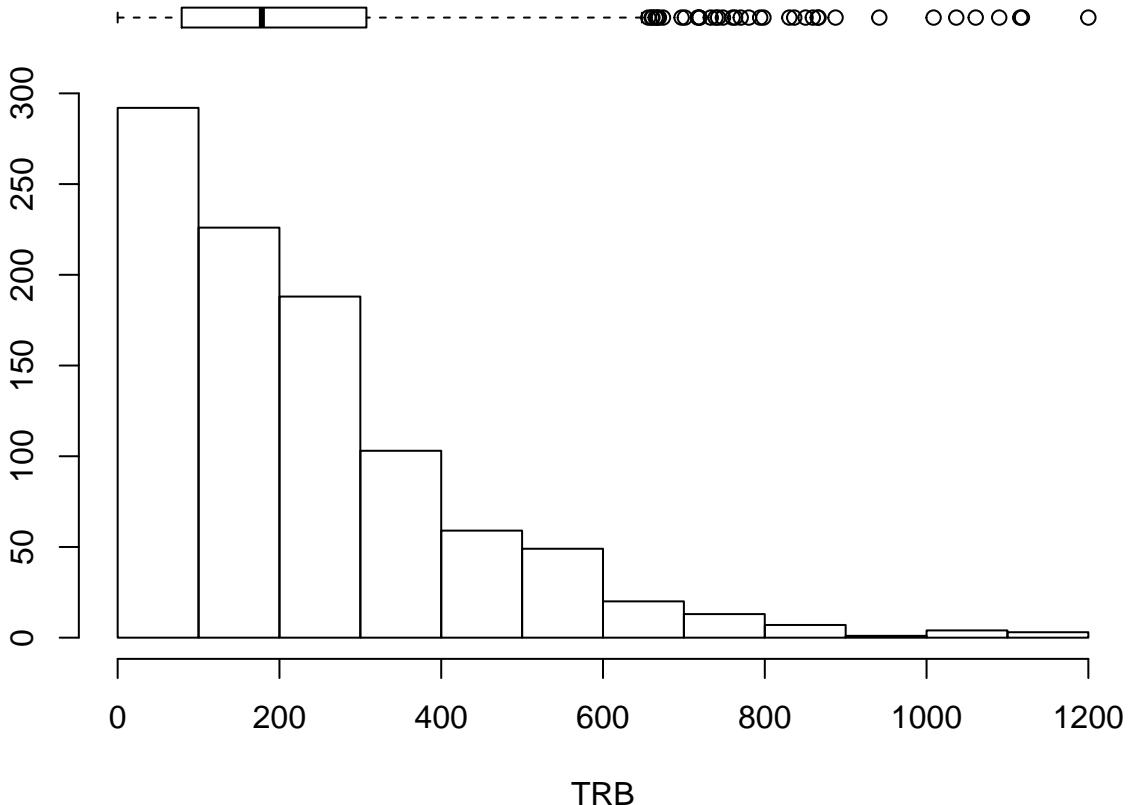
```
## Top 10 Players by ORB
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Andre~ 2.21e7 C     22 DET     81   81 2666 21.2 0.499 0.006
## 2 2017 Andre~ 2.38e7 C     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 Dwight~ 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>, O BPM <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> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 DeAndre 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 Hassan 2.38e7 C     27 MIA    77   77 2513 22.6 0.579 0
## 4 2016 DeAndre 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 Russell 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 Anthony 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 Julius 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>, O BPM <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

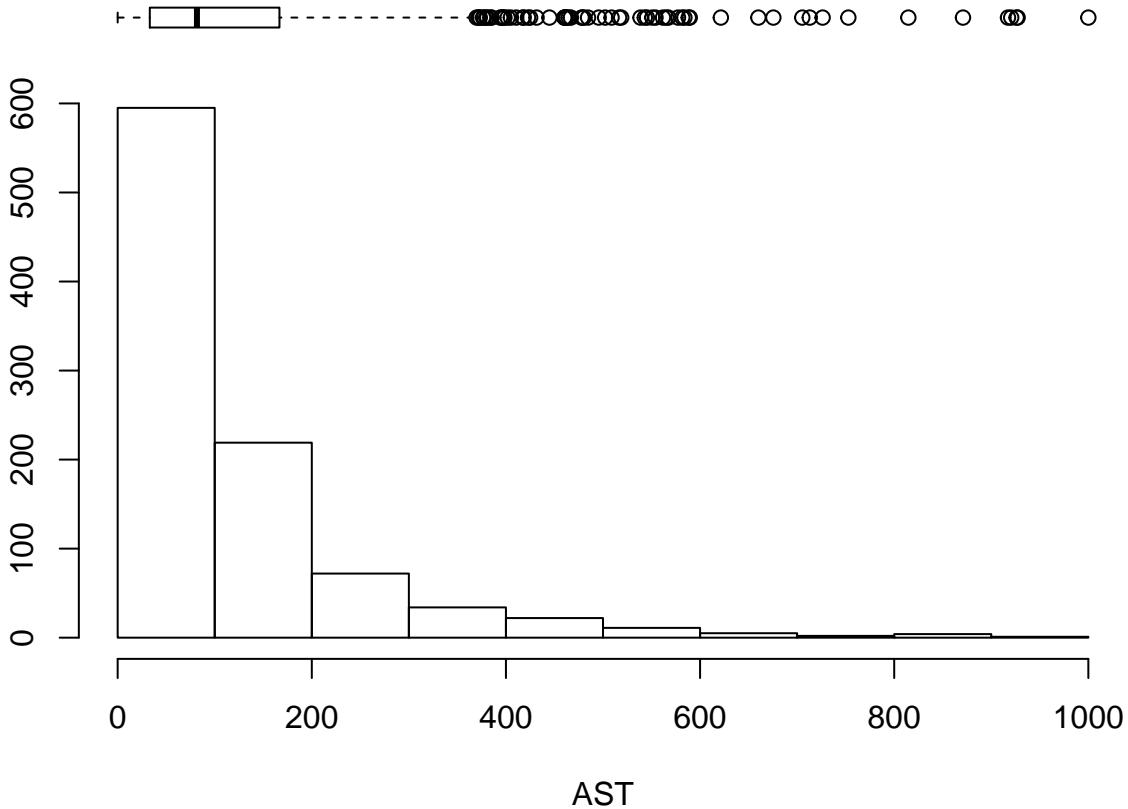


```

## 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>
## 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>, O BPM <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>

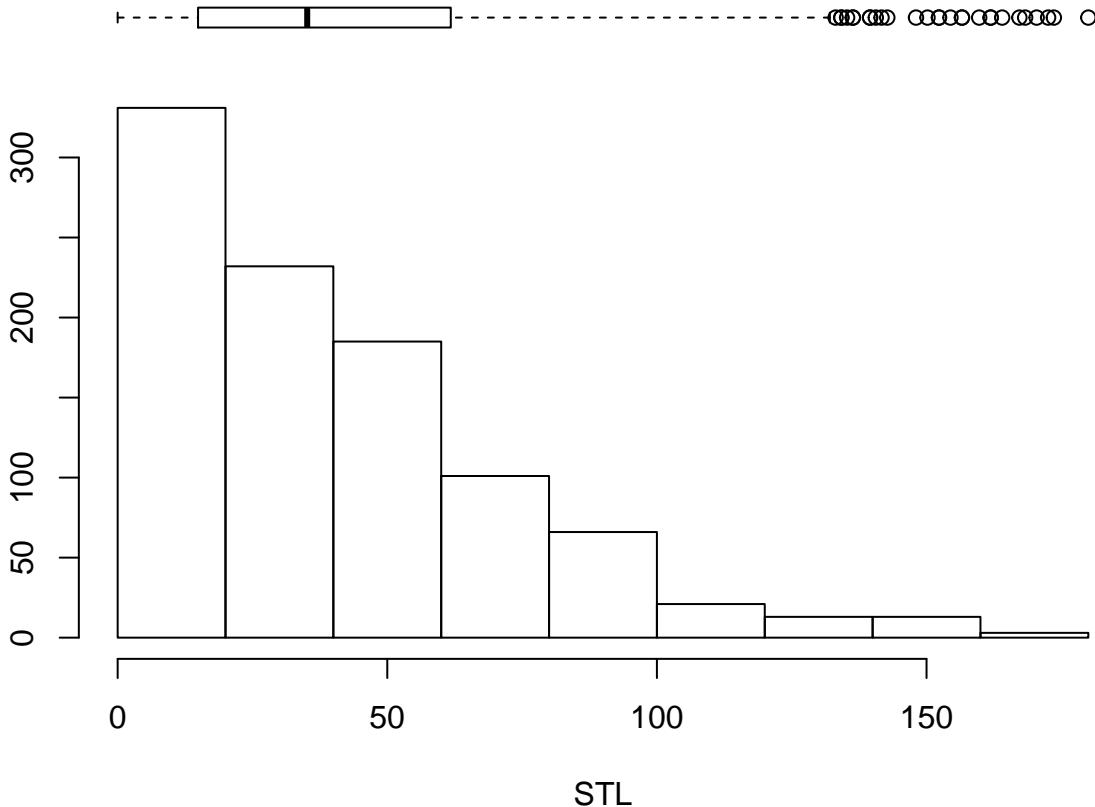
```

Histogram of AST



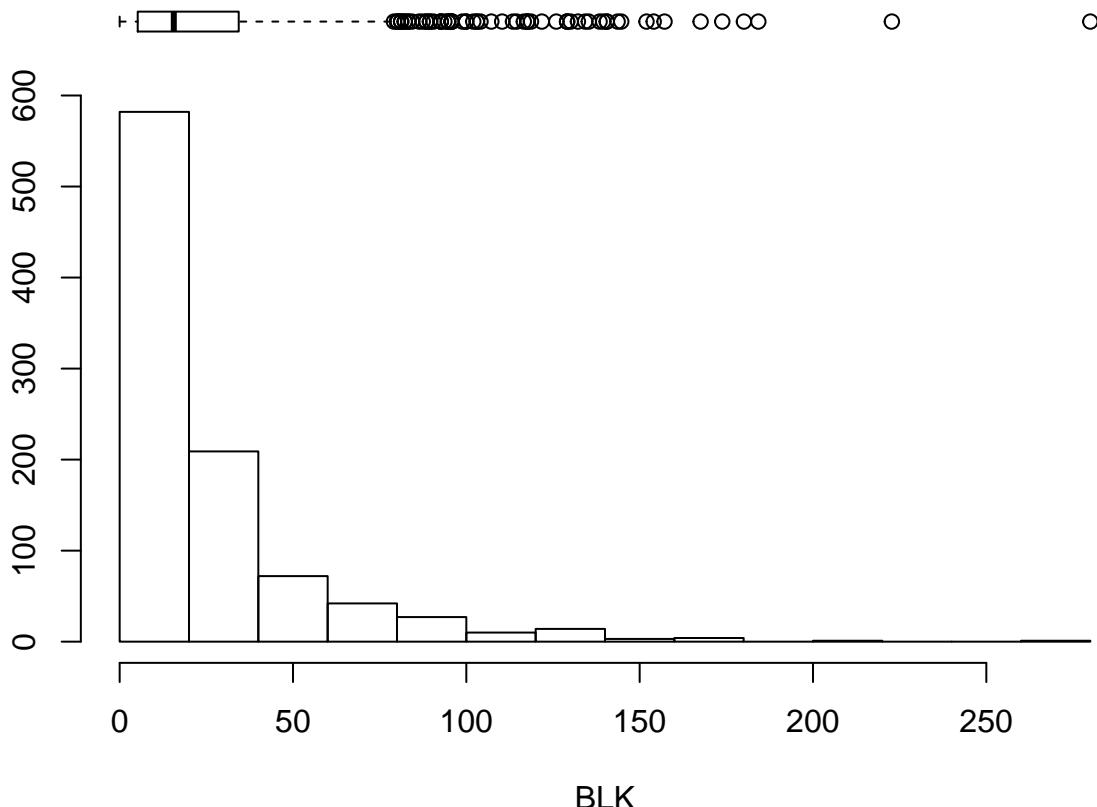
```
## Top 10 Players by AST
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 James~ 2.83e7 PG      27 HOU     81   81 2947 27.3 0.613 0.493
## 2 2017 Russe~ 2.85e7 PG      28 OKC     81   81 2802 30.6 0.554 0.3
## 3 2016 Rajon~ 1.40e7 PG      29 SAC     72   72 2537 16.9 0.506 0.217
## 4 2016 Russe~ 2.65e7 PG      27 OKC     80   80 2750 27.6 0.554 0.236
## 5 2017 John ~ 1.81e7 PG      26 WAS     78   78 2836 23.2 0.541 0.19
## 6 2016 John ~ 1.70e7 PG      25 WAS     77   77 2784 19.8 0.51  0.243
## 7 2016 Chris~ 2.29e7 PG      30 LAC     74   74 2420 26.2 0.575 0.295
## 8 2017 Ricky~ 1.43e7 PG      26 MIN     75   75 2469 16.8 0.539 0.302
## 9 2016 Ricky~ 1.36e7 PG      25 MIN     76   76 2323 17.6 0.529 0.324
## 10 2017 LeBro~ 3.33e7 SF     32 CLE    74   74 2794 27   0.619 0.254
## # ... 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>, O BPM <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 STL



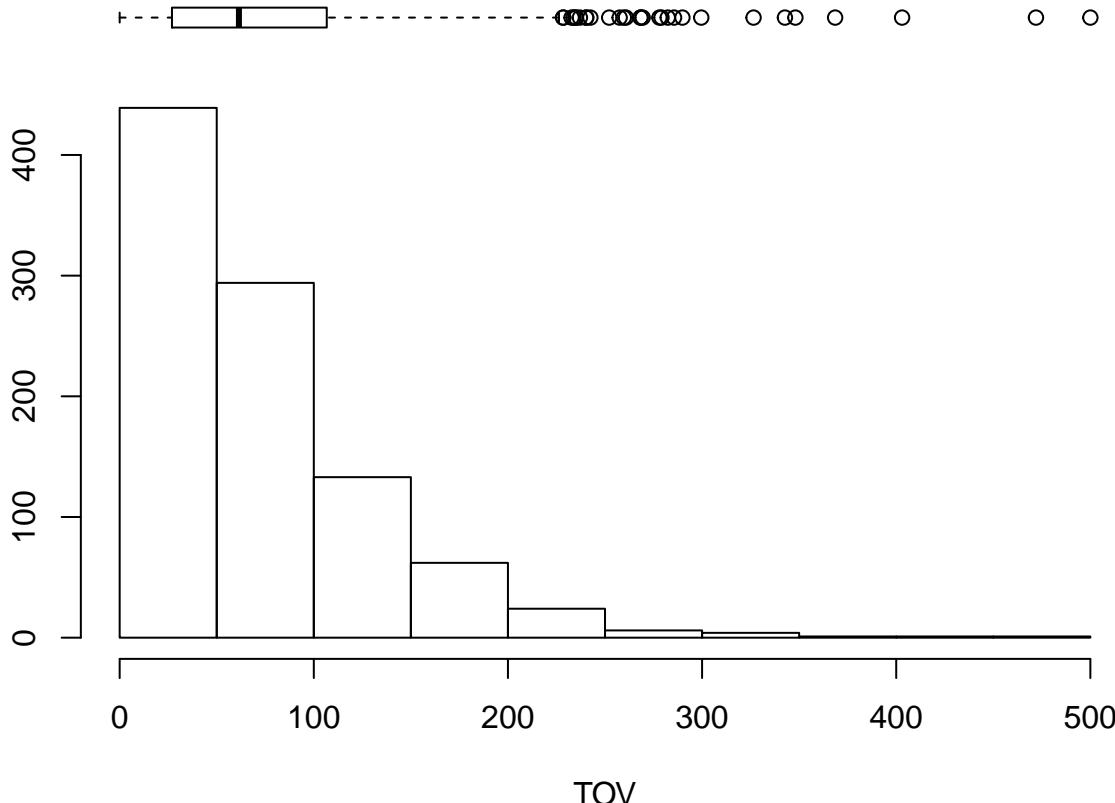
```
## Top 10 Players by STL
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016  Steph~ 1.21e7 PG     27 GSW    79    79  2700 31.5 0.669 0.554
## 2 2016  Russe~ 2.65e7 PG     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>, O BPM <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



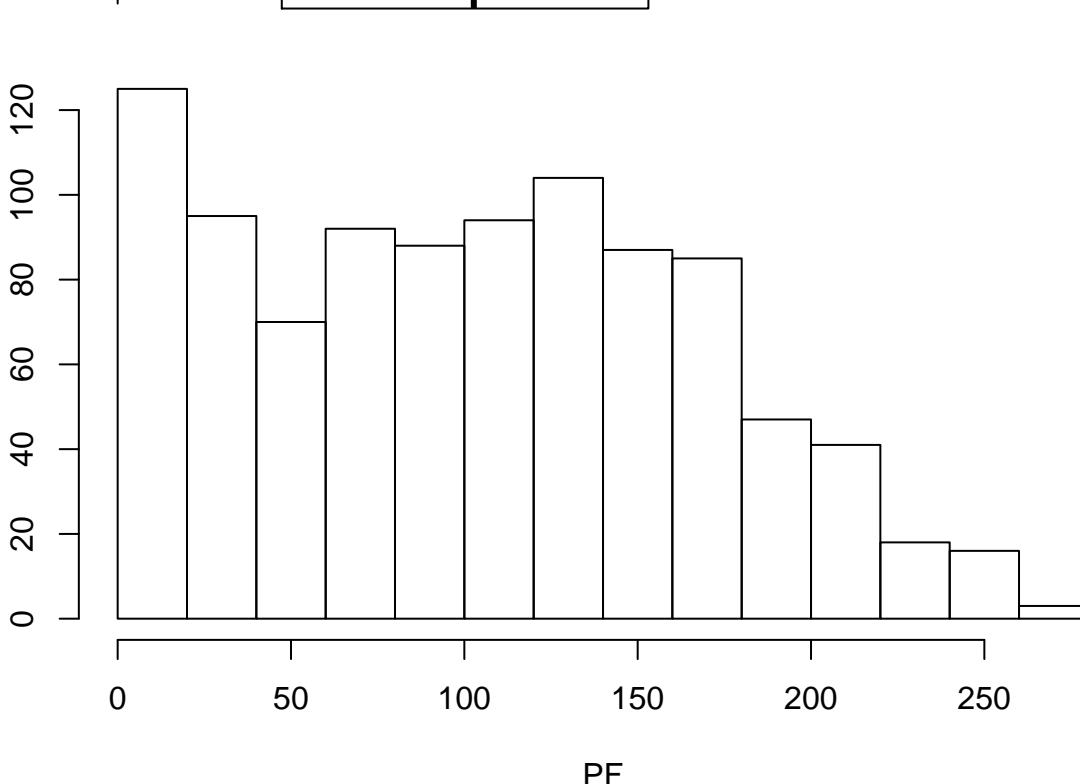
```
## Top 10 Players by BLK
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2016 Hassa~ 2.21e7 C       26 MIA    73   43  2125 25.7 0.629  0
## 2 2017 Rudy ~ 2.20e7 C       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>, O BPM <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



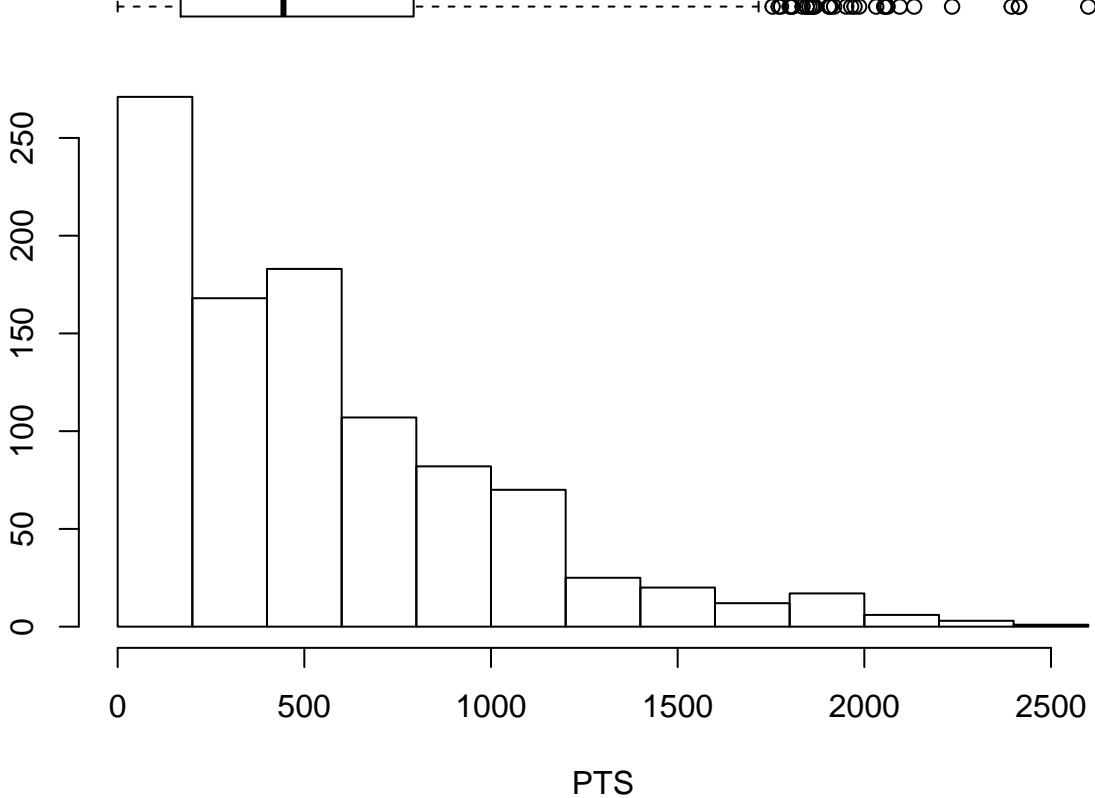
```
## Top 10 Players by TOV
## # A tibble: 10 x 51
##   year   name_p salary Pos    Age Tm      G   GS   MP   PER `TS%` `3PAr`
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2017 James~ 2.83e7 PG     27 HOU    81   81 2947 27.3 0.613 0.493
## 2 2017 Russe~ 2.85e7 PG     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 2016 Russe~ 2.65e7 PG     27 OKC    80   80 2750 27.6 0.554 0.236
## 5 2017 John ~ 1.81e7 PG     26 WAS    78   78 2836 23.2 0.541 0.19
## 6 2016 John ~ 1.70e7 PG     25 WAS    77   77 2784 19.8 0.51  0.243
## 7 2017 LeBro~ 3.33e7 SF     32 CLE    74   74 2794 27   0.619 0.254
## 8 2016 Rajon~ 1.40e7 PG     29 SAC    72   72 2537 16.9 0.506 0.217
## 9 2017 DeMar~ 1.81e7 C      26 TOT    72   72 2465 25.7 0.562 0.254
## 10 2016 Paul ~ 1.83e7 SF    25 IND   81   81 2819 20.9 0.557 0.391
## # ... 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>, O BPM <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>
## 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

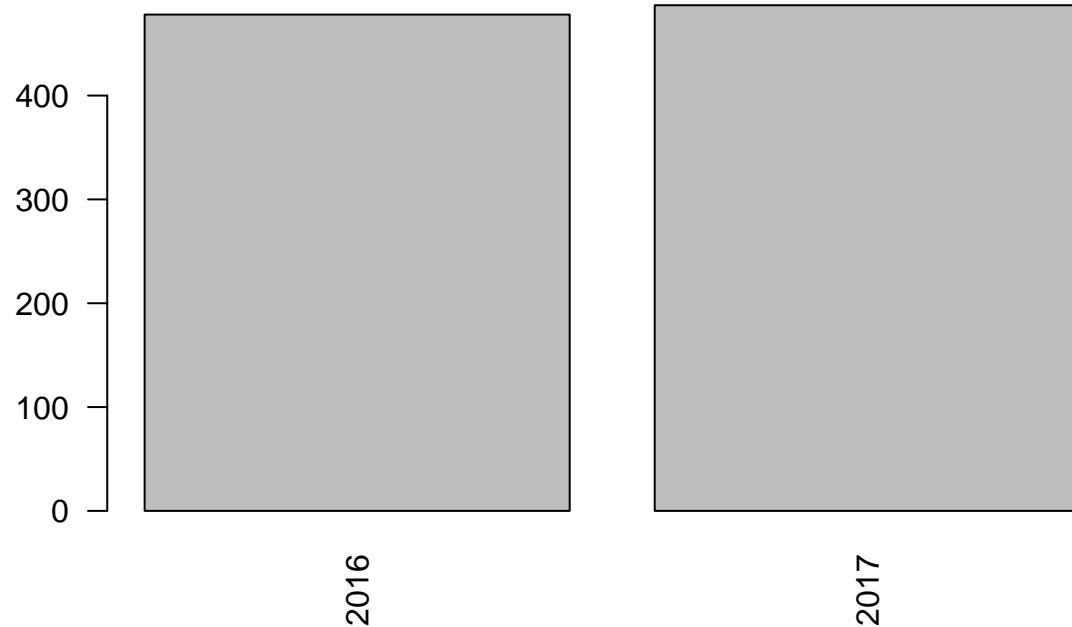


```
## Top 10 Players by PTS
## # A tibble: 10 x 51
##   year   name_p salary Pos     Age Tm      G   GS   MP    PER `TS%` `3PAr` 
##   <fct> <chr>   <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> 
## 1 2017  Russel~  2.85e7 PG      28  OKC    81   81  2802 30.6 0.554  0.3  
## 2 2016  James~   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  Isaias~  6.26e6 PG      27  BOS    76   76  2569 26.5 0.625  0.439 
## 6 2017  Anthro~  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>, O BPM <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

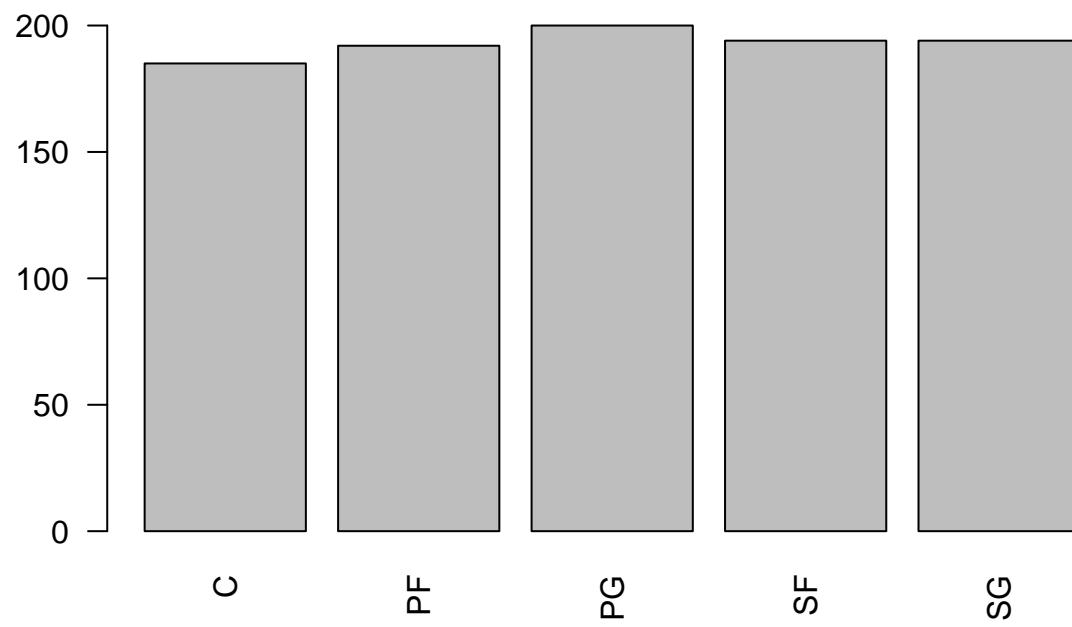
```
categorical_vars <- c('year', 'Pos', 'Tm')
for (col in categorical_vars){
  data <- df_primary[[col]]
  barplot(table(data), main=sprintf('Histogram of %s', col), las=2)
  print('\n')}
```

Histogram of year



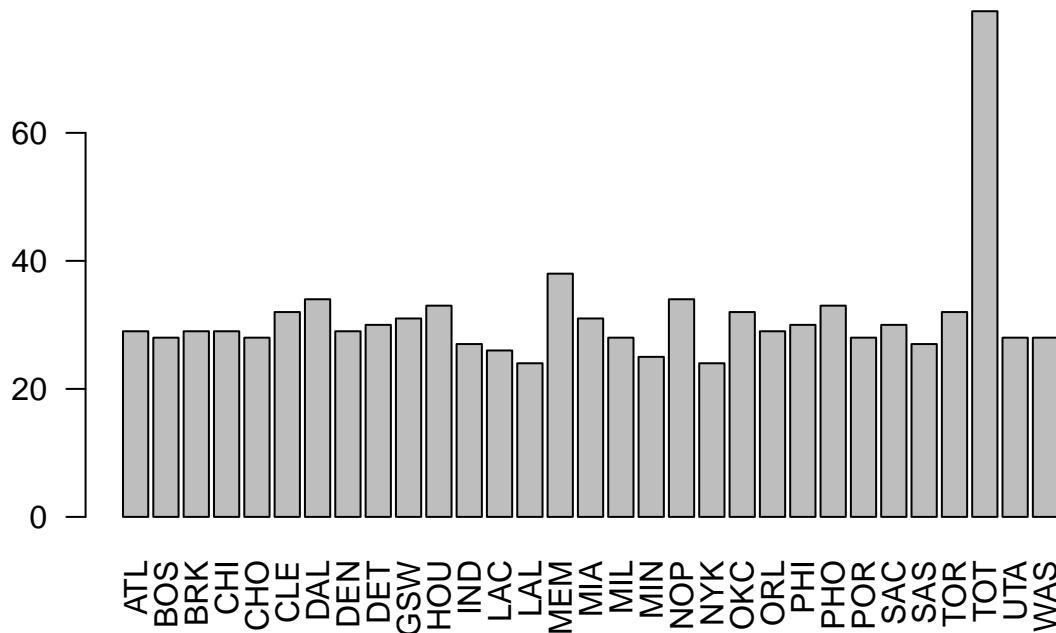
```
## [1] "\n"
```

Histogram of Pos



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

Histogram of Tm



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

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

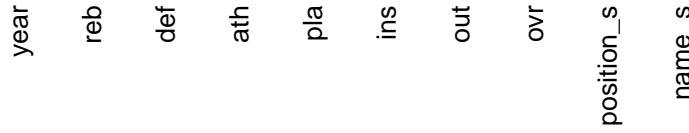
```
secondary_attributes <- c('name_s','position_s','ovr','out','ins','pla','ath','def','reb')
df_secondary <- vector('list',9)
names(df_secondary) <- secondary_attributes
path_f = '../data/raw/nba2k/nba2k_%d.csv'
for (year in c(16:20)){
  df_year <- read.csv(sprintf(path_f,year))
  headers <- names(df_year)
  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_attributes
  df_year[, 'year'] <- 2000+year
  df_secondary <- rbind(df_secondary,df_year)}
original_s <- df_secondary
# Checking for missing values in primary dataset
missCounts <- sapply(df_secondary,function(x) sum(is.na(x)))
missCounts

##      name_s position_s          ovr          out          ins          pla          ath
##            0            0            0            0            0            0            0
##      def      reb      year
##            0            0            0

library(Amelia)
missmap(df_secondary, main = "Missing values ")
```

Missing values

```
8283  
7938  
7593  
7248  
6903  
6558  
6213  
5868  
5523  
5178  
4833  
4488  
4143  
3798  
3453  
3108  
2763  
2418  
2073  
1728  
1383  
1038  
693  
348  
3
```



Missing (0%)
Observed (100%)

```
dev.copy(png, "../figures/Missing values/Missing_value_secondary.png")
```

```
## png  
## 3
```

```
dev.off()
```

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

```
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     G/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 ...
## $ 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 ...

```

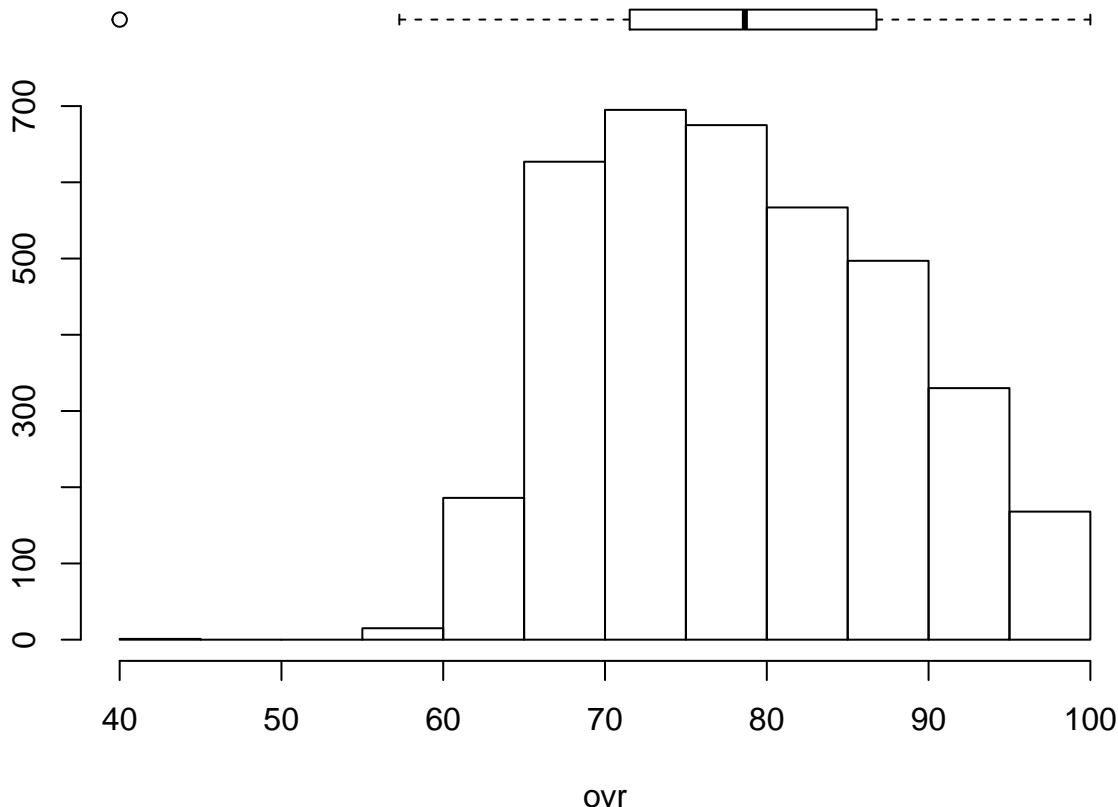
Histogram Barcharts for Numeric Variables

```

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,])
}

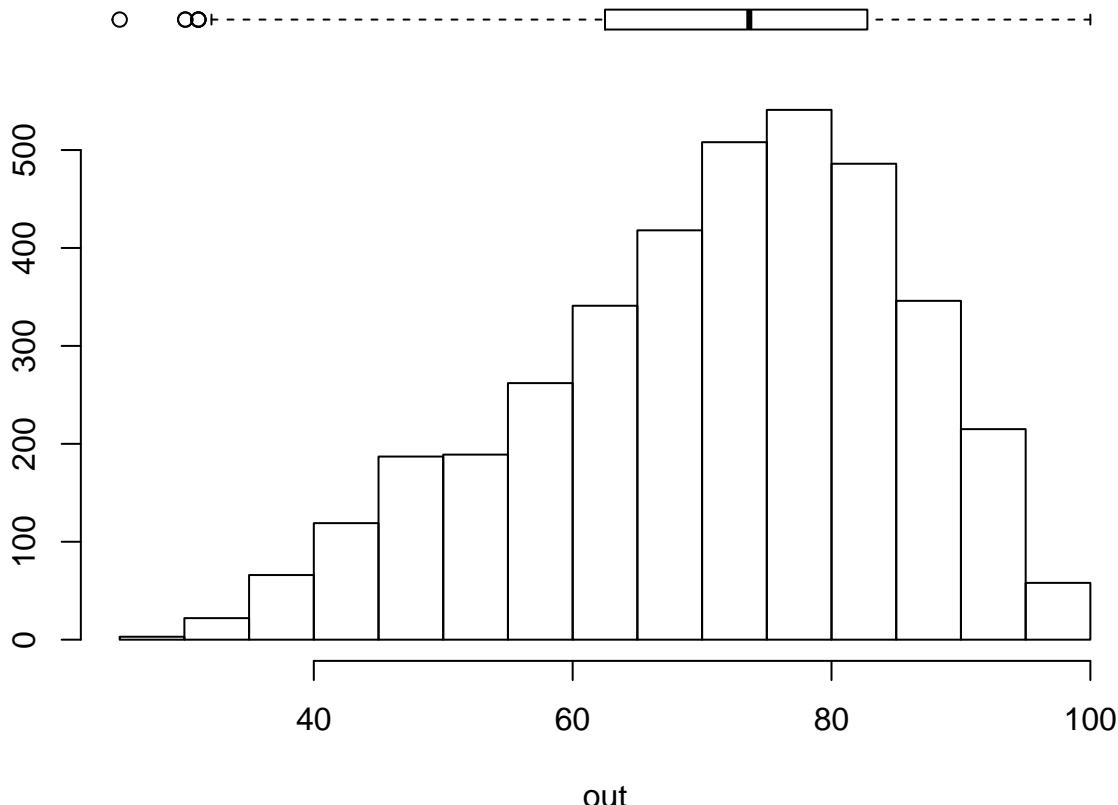
```

Histogram of ovr



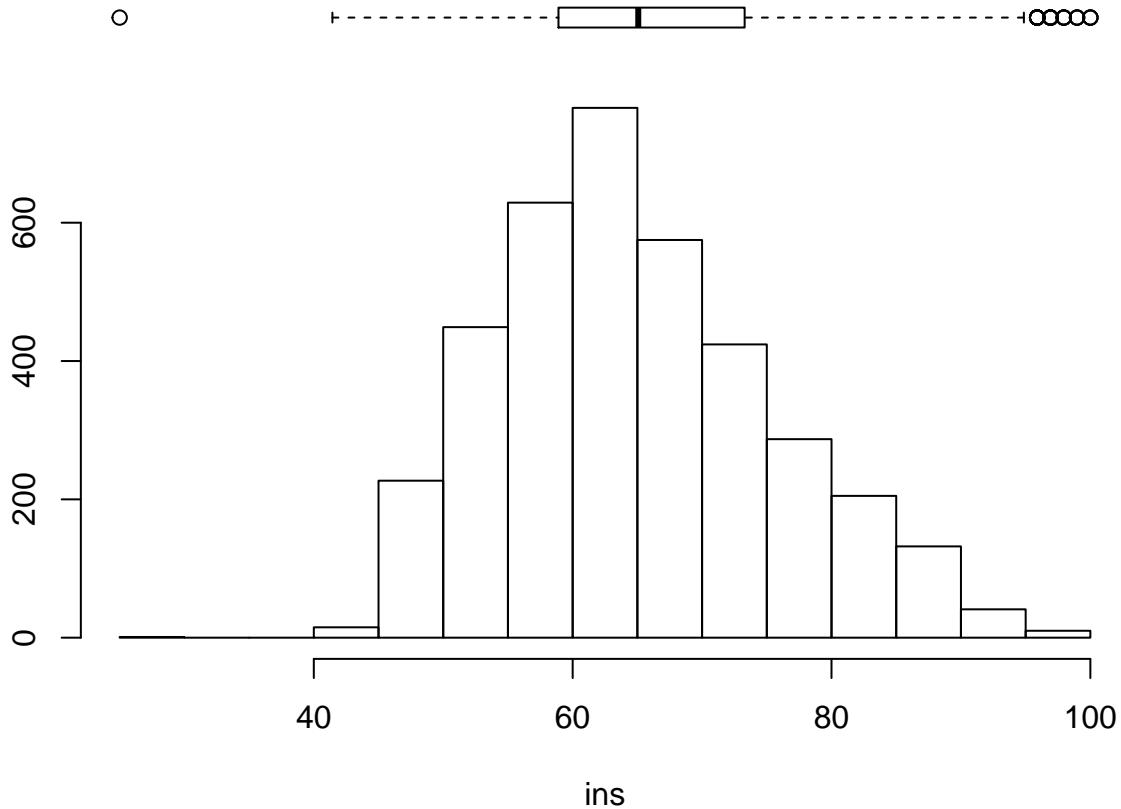
```
## Top 10 Players by ovr
##          name_s position_s   ovr   out   ins   pla   ath   def   reb year
## 1      '96 Michael Jordan     SG    99    95    88    91    93    92    75 2016
## 2      '15 Kobe Bryant      SG    99    97    79    95    84    88    65 2016
## 3      Stephen Curry       PG    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



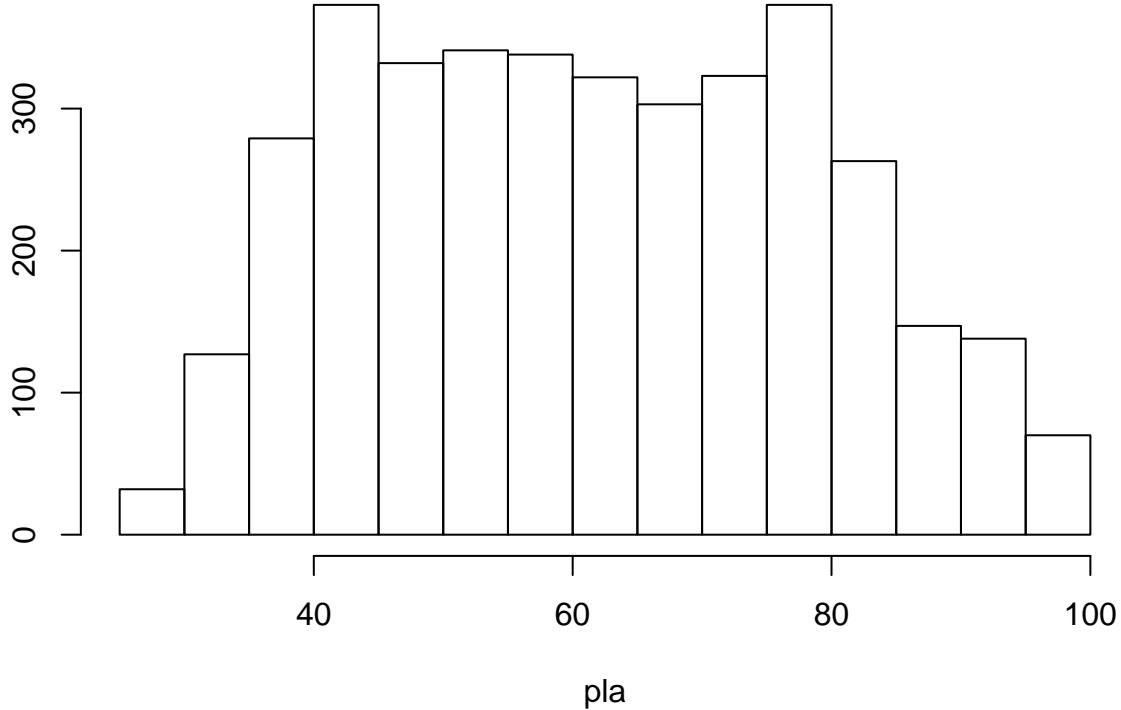
```
## Top 10 Players by out
##          name_s position_s ovr out ins pla ath def reb year
## 2098    Stephen Curry      PG  98  99  70  98  92  86  78 2017
## 3        Stephen Curry      PG  99  98  66  98  89  78  54 2016
## 6        Kyrie Irving       PG  98  98  70  95  91  74  49 2016
## 17       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



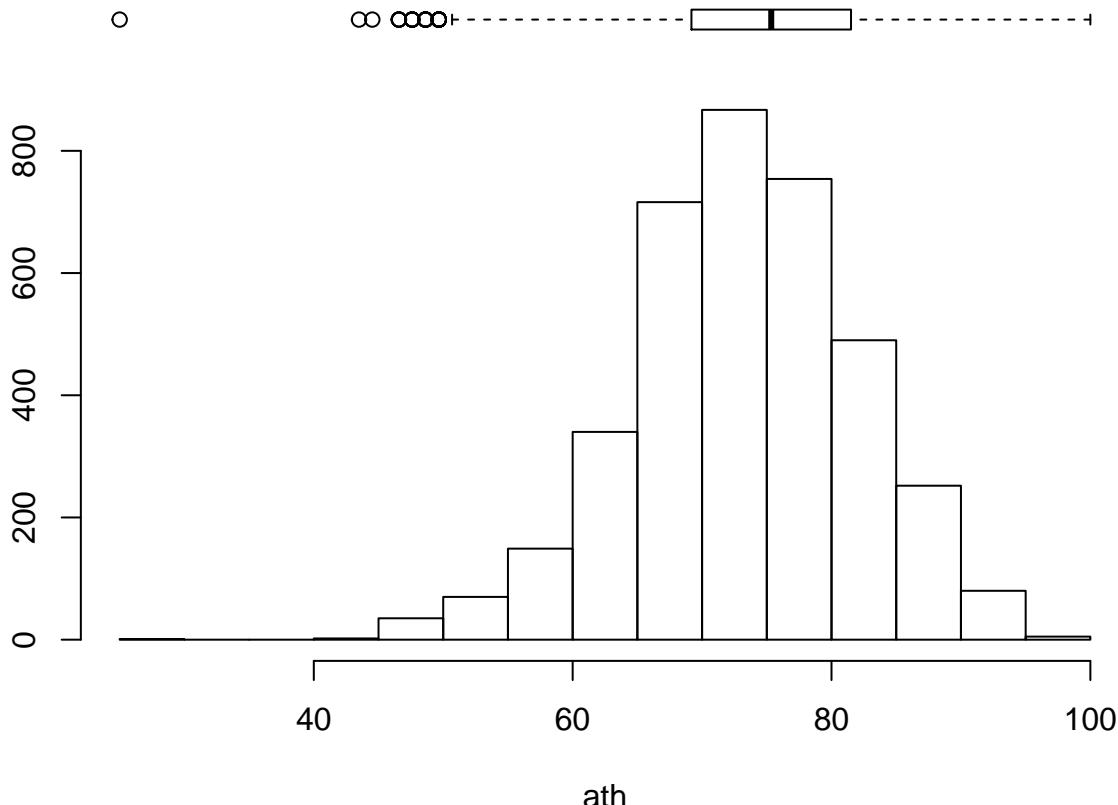
```
## Top 10 Players by ins
##          name_s position_s ovr out ins pla ath def reb year
## 2091      Charles Barkley      PF  99  95  98  89  95  97  98 2017
## 2106    Kareem Abdul-Jabbar      C   98  83  98  85  90  95  95  98 2017
## 2107      Karl Malone       PF  98  88  98  80  96  96  94  98 2017
## 2110    Anthony Davis       PF  98  87  97  65  91  94  94  97 2017
## 2135     Kevin Garnett       PF  97  86  97  80  93  93  94  98 2017
## 2227 Amar'e Stoudemire      PF  95  82  97  63  88  82  82  94 2017
## 2096     Michael Jordan      SG  99  97  96  95  96  96  95  80 2017
## 2103     Bill Russell        C   98  57  96  79  92  97  97  99 2017
## 2205     Wes Unseld         C   95  78  96  89  90  94  94  98 2017
## 2213     Shawn Kemp        PF  95  83  96  63  92  85  95  95 2017
```

Histogram of pla



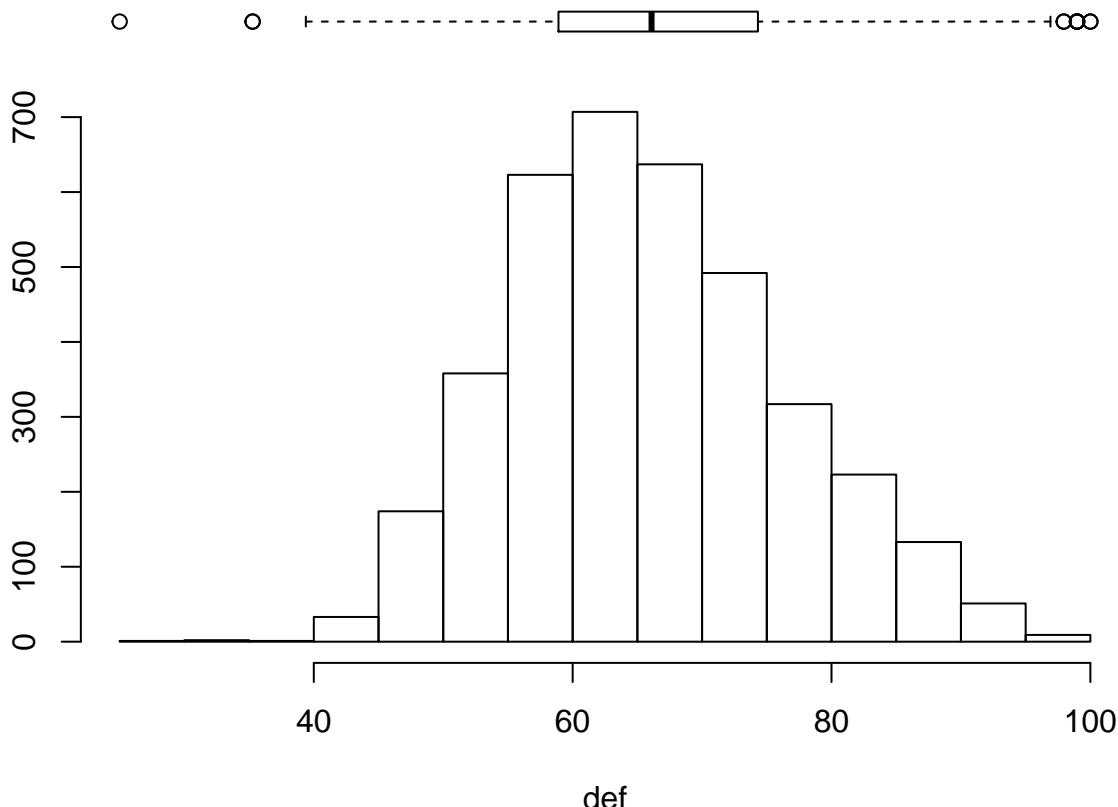
```
## Top 10 Players by pla
##          name_s position_s ovr out ins pla ath def reb year
## 2101      Magic Johnson      PG 98  95  90  99  96  95  89 2017
## 2115      John Stockton      PG 98  97  70  99  92  90  51 2017
## 3        Stephen Curry      PG 99  98  66  98  89  78  54 2016
## 7     '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



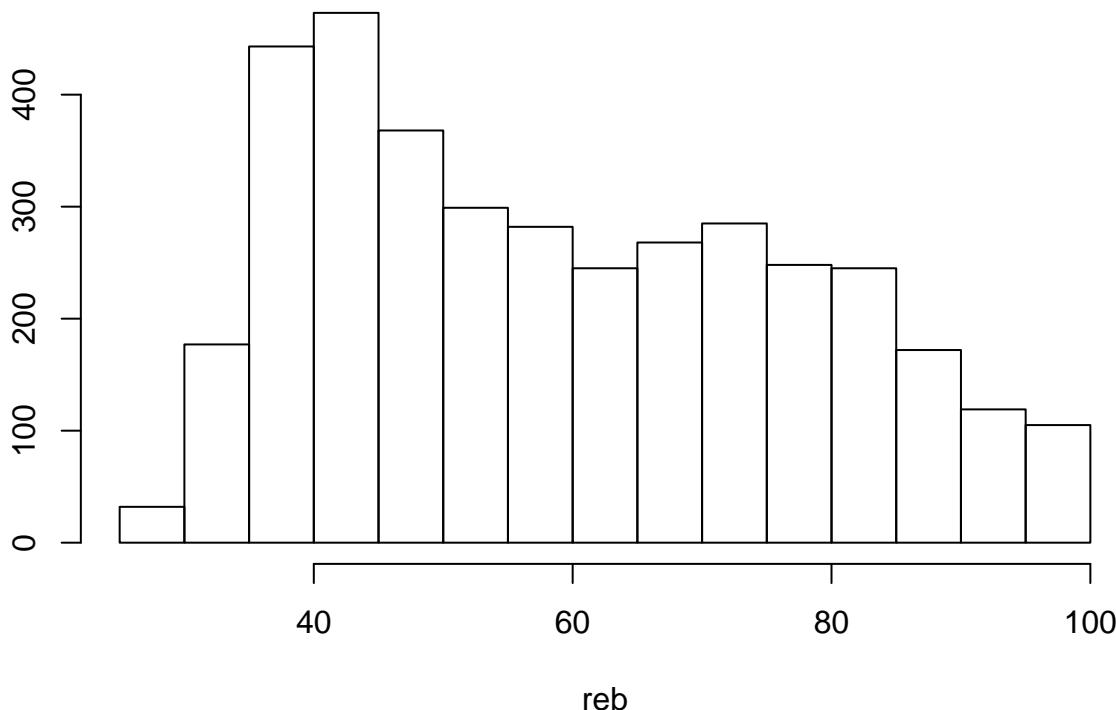
```
## Top 10 Players by ath
##          name_s position_s ovr out ins pla ath def reb year
## 2097 Russell Westbrook      PG  99  97  83  98  98  91  97 2017
## 2099   Lebron James        SF  98  94  94  95  95  97  92  74 2017
## 2096 Michael Jordan       SG  99  97  96  95  95  96  95  80 2017
##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

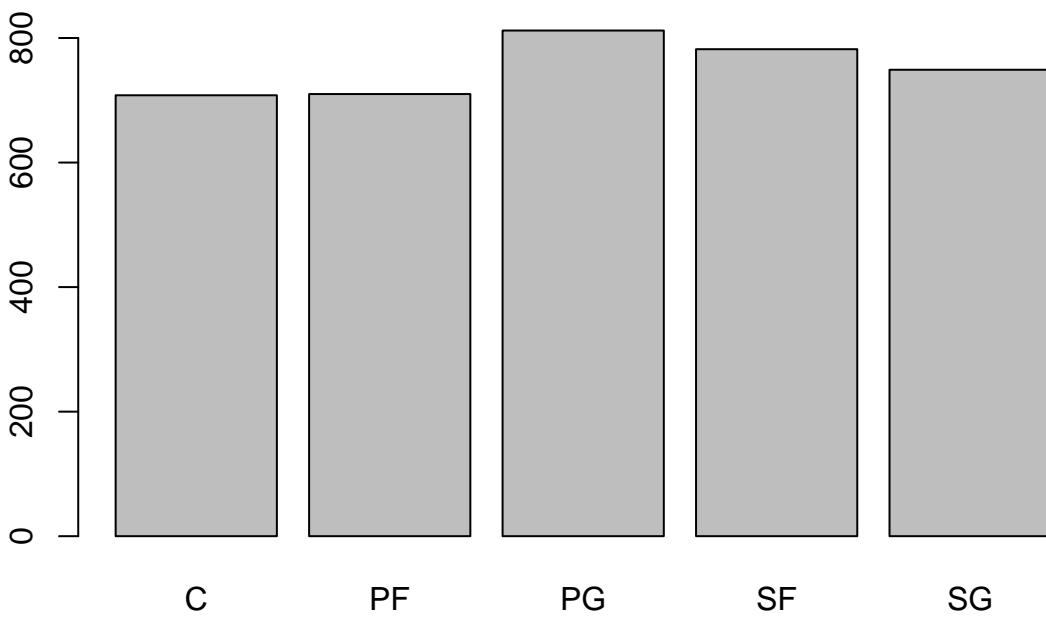
```
categorical_vars <- c('year', 'position_s')
for (col in categorical_vars){
  data <- df_secondary[[col]]
  barplot(table(data), main=sprintf('Histogram of %s', col))
  print('\n')}
```

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 ascii
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(
  `isiah thomas` = 'isaiah thomas',
  `jonathan simmons` = 'jonathon simmons',
  `lance stepheson` = 'lance stephenson',
  `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
## 162412	zoran dragic	goran dragic	1
## 299490	luke babbitt	luke babbittt	1
## 609192	goran dragic	zoran dragic	1
## 1009318	luke babbittt	luke babbitt	1
## 7438	ryan anderson	alan anderson	2
## 8954	alvin williams	alan williams	2
## 90435	damon jones	damian jones	2
## 104165	david wear	david west	2
## 133476	dryamond green	draymond green	2
## 161685	flynn robinson	glenn robinson	2
## 179137	josh smith	ish smith	2
## 217008	brian grant	jerian grant	2
## 247409	ish smith	josh smith	2
## 295807	mo williams	lou williams	2
## 311210	darius morris	marcus morris	2
## 319499	alvin williams	marvin williams	2
## 337975	lou williams	mo williams	2
## 375741	paul pressey	phil pressey	2
## 404018	alan anderson	ryan anderson	2
## 478902	willie green	willie reed	2
## 487381	joe bryant	kobe bryant	2
## 524082	willie reed	willie green	2
## 586816	darius miles	darius miller	2
## 608993	drew gooden	drew gordon	2
## 652349	daivid west	david wear	2
## 667630	marcus morris	darius morris	2
## 686256	charles oakley	charles barkley	2
## 749680	ervin johnson	kevin johnson	2
## 805879	shareef abdur rahim	shareef adburr ahim	2
## 819208	draymond green	dryamond green	2
## 849908	charles barkley	charles oakley	2
## 878167	shareef abdur rahim	shareef abdur rahim	2
## 879749	phil pressey	paul pressey	2

```

## 888581      glenn robinson      flynn robinson      2
## 905084      darius miller      darius miles      2
## 917574      alan williams      alvin williams      2
## 917883      marvin williams      alvin williams      2
## 936876      jerian grant      brian grant      2
## 959516      kevin johnson      ervin johnson      2
## 965285      kobe bryant      joe bryant      2
## 972441      drew gordon      drew gooden      2
## 994035      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] 733

```

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:733 2016:369  Min.   :11534  C :157  Min.   :19.00
## Class :character 2017:364  1st Qu.:2116955  PF:147  1st Qu.:23.00
## Mode  :character                      Median :5200000  PG:138  Median :26.00
##                               Mean   :7837816  SF:142  Mean   :26.55
##                               3rd Qu.:12016854  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.: 846  1st Qu.:10.90  

##  TOR : 27  Median  :68.00  Median  :20.00  Median  :1508  Median :13.70  

##  PHO : 25  Mean    :61.17  Mean    :31.68  Mean    :1475  Mean   :14.17  

##  UTA : 25  3rd Qu.:77.00  3rd Qu.:63.00  3rd Qu.:2119  3rd Qu.:16.90  

##  DET : 24  Max.    :82.00  Max.    :82.00  Max.    :3125  Max.   :32.00  

## (Other):535  

##      TS%          3PAr          FTr          ORB%  

##  Min.   :0.1250  Min.   :0.000  Min.   :0.0000  Min.   : 0.000  

##  1st Qu.:0.5090  1st Qu.:0.106  1st Qu.:0.1790  1st Qu.: 2.000  

##  Median :0.5410  Median  :0.305  Median  :0.2480  Median : 3.600  

##  Mean   :0.5389  Mean    :0.293  Mean    :0.2724  Mean   : 5.073  

##  3rd Qu.:0.5720  3rd Qu.:0.442  3rd Qu.:0.3400  3rd Qu.: 7.500  

##  Max.   :1.0000  Max.   :0.900  Max.   :1.2190  Max.   :21.800  

##  

##      DRB%          TRB%          AST%          STL%          BLK%  

##  Min.   : 0.00  Min.   : 0.0  Min.   : 0.0  Min.   : 0.000  Min.   : 0.000  

##  1st Qu.:10.50  1st Qu.: 6.3  1st Qu.: 7.2  1st Qu.: 1.100  1st Qu.:0.600  

##  Median :14.60  Median  : 9.3  Median  :10.3  Median  : 1.500  Median :1.200  

##  Mean   :15.53  Mean    :10.3  Mean    :13.4  Mean    : 1.586  Mean   :1.737  

##  3rd Qu.:19.60  3rd Qu.:13.3  3rd Qu.:17.7  3rd Qu.: 1.900  3rd Qu.:2.500  

##  Max.   :36.30  Max.   :25.6  Max.   :57.3  Max.   :11.100  Max.   :9.700  

##  

##      TOV%          USG%          OWS          DWS  

##  Min.   : 0.00  Min.   : 0.00  Min.   :-3.300  Min.   : 0.000  

##  1st Qu.:10.00  1st Qu.:15.40  1st Qu.: 0.200  1st Qu.: 0.700  

##  Median :12.50  Median  :18.50  Median  : 1.100  Median  :1.300  

##  Mean   :12.77  Mean    :19.18  Mean    : 1.722  Mean   : 1.522  

##  3rd Qu.:15.10  3rd Qu.:22.20  3rd Qu.: 2.500  3rd Qu.: 2.200  

##  Max.   :43.60  Max.   :41.70  Max.   :13.800  Max.   : 6.000  

##  

##      WS          WS/48          OBPM          DBPM  

##  Min.   :-2.100  Min.   :-0.28300  Min.   :-17.3000  Min.   :-8.20000  

##  1st Qu.: 1.100  1st Qu.: 0.05600  1st Qu.: -2.1000  1st Qu.:-1.30000  

##  Median : 2.500  Median  : 0.09100  Median  : -0.7000  Median :-0.10000  

##  Mean   : 3.243  Mean    : 0.09283  Mean    : -0.6689  Mean   :-0.08868  

##  3rd Qu.: 4.500  3rd Qu.: 0.12700  3rd Qu.:  0.5000  3rd Qu.: 1.10000  

##  Max.   :17.900  Max.   : 0.34300  Max.   : 12.4000  Max.   :12.00000  

##  

##      BPM          VORP          FG          FGA  

##  Min.   :-24.1000  Min.   :-1.4000  Min.   : 0.0  Min.   : 0.0  

##  1st Qu.: -2.7000  1st Qu.:-0.1000  1st Qu.:116.0  1st Qu.: 256.0  

##  Median : -0.7000  Median  : 0.4000  Median  :208.0  Median : 462.0  

##  Mean   : -0.7569  Mean    : 0.8359  Mean    :240.3  Mean   : 526.3  

##  3rd Qu.:  1.0000  3rd Qu.: 1.3000  3rd Qu.:338.0  3rd Qu.: 731.0  

##  Max.   : 15.6000  Max.   :12.4000  Max.   :824.0  Max.   :1941.0  

##  

##      FG%          3P          3PA          3P%  

##  Min.   :0.0830  Min.   : 0.00  Min.   : 0.0  Min.   :0.0000  

##  1st Qu.:0.4110  1st Qu.: 4.00  1st Qu.: 16.0  1st Qu.:0.2711  

##  Median :0.4450  Median  :42.00  Median  :120.0  Median :0.3330  

##  Mean   :0.4533  Mean    :56.32  Mean    :157.1  Mean   :0.3034  

##  3rd Qu.:0.4880  3rd Qu.:91.00  3rd Qu.:256.0  3rd Qu.:0.3730  

##  Max.   :1.0000  Max.   :402.00  Max.   :886.0  Max.   :1.0000  

##  

##      2P          2PA          2P%          eFG%  

##  Min.   : 0  Min.   : 0.0  Min.   :0.0000  Min.   :0.1000  

##  1st Qu.: 75  1st Qu.:157.0  1st Qu.:0.4530  1st Qu.:0.4730  

##  Median :153  Median  :308.0  Median :0.4850  Median :0.5060  

##  Mean   :184  Mean    :369.2  Mean    :0.4889  Mean   :0.5043  

##  3rd Qu.:258  3rd Qu.:513.0  3rd Qu.:0.5310  3rd Qu.:0.5370  

##  Max.   :730  Max.   :1421.0  Max.   :1.0000  Max.   :1.0000

```

```

##  

##      FT          FTA          FT%          ORB  

## Min.   : 0.0   Min.   : 0.0   Min.   :0.0000   Min.   : 0.00  

## 1st Qu.: 38.0  1st Qu.: 51.0  1st Qu.:0.7000  1st Qu.: 21.00  

## Median : 80.0  Median :110.0  Median :0.7690  Median : 44.00  

## Mean   :111.6  Mean   :145.3  Mean   :0.7518  Mean   : 63.34  

## 3rd Qu.:145.0  3rd Qu.:194.0  3rd Qu.:0.8310  3rd Qu.: 86.00  

## Max.   :746.0   Max.   :881.0   Max.   :1.0000  Max.   :395.00  

##  

##      DRB          TRB          AST          STL  

## Min.   : 0.0   Min.   : 0.0   Min.   : 0.0   Min.   : 0.00  

## 1st Qu.:103.0  1st Qu.: 128.0  1st Qu.: 47.0  1st Qu.: 22.00  

## Median :180.0  Median : 229.0  Median : 97.0  Median : 42.00  

## Mean   :206.8  Mean   : 270.1  Mean   :137.2  Mean   : 47.39  

## 3rd Qu.:279.0  3rd Qu.: 365.0  3rd Qu.:176.0  3rd Qu.: 66.00  

## Max.   :817.0   Max.   :1198.0  Max.   :906.0  Max.   :169.00  

##  

##      BLK          TOV          PF          PTS  

## Min.   : 0.00   Min.   : 0.00   Min.   : 0.0   Min.   : 0.0  

## 1st Qu.: 9.00  1st Qu.: 39.00  1st Qu.: 79.0  1st Qu.: 307.0  

## Median :20.00  Median : 69.00  Median :125.0  Median : 544.0  

## Mean   :30.21  Mean   : 83.28  Mean   :121.6  Mean   : 648.6  

## 3rd Qu.:39.00  3rd Qu.:114.00  3rd Qu.:165.0  3rd Qu.: 898.0  

## Max.   :269.00  Max.   :464.00  Max.   :278.0  Max.   :2558.0  

##  

##      out          ovr          ins          pla          ath  

## Min.   :30.00   Min.   :61.0   Min.   :44.00   Min.   :28.00   Min.   :49.00  

## 1st Qu.:61.00  1st Qu.:71.0   1st Qu.:58.00  1st Qu.:47.00  1st Qu.:68.00  

## Median :72.00  Median : 76.0   Median :64.00  Median :59.00  Median :73.00  

## Mean   :71.19  Mean   : 78.5   Mean   :65.35  Mean   :61.27  Mean   :73.49  

## 3rd Qu.:83.00  3rd Qu.:85.0   3rd Qu.:71.00  3rd Qu.:76.00  3rd Qu.:79.00  

## Max.   :99.00  Max.   :99.0   Max.   :97.00  Max.   :98.00  Max.   :98.00  

##  

##      def          reb  

## Min.   :43.00   Min.   :27.00  

## 1st Qu.:58.00  1st Qu.:44.00  

## Median :64.00  Median : 59.00  

## Mean   :65.49  Mean   : 60.86  

## 3rd Qu.:72.00  3rd Qu.:74.00  

## Max.   :98.00  Max.   :98.00  

##
```

```

missCounts <- sapply(df_final,function(x) sum(is.na(x)))
missCounts

```

	name	year	salary	Pos	Age	Tm	G	GS	MP	PER	TS%
##	0	0	0	0	0	0	0	0	0	0	0
##	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	OWS
##	0	0	0	0	0	0	0	0	0	0	0
##	DWS	WS	WS/48	OBPM	DBPM	BPM	VORP	FG	FGA	FG%	3P
##	0	0	0	0	0	0	0	0	0	0	0
##	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB
##	0	0	0	0	0	0	0	0	0	0	0
##	TRB	AST	STL	BLK	TOV	PF	PTS	out	ovr	ins	pla
##	0	0	0	0	0	0	0	0	0	0	0
##	ath	def	reb								
##	0	0	0								

```

library(Amelia)

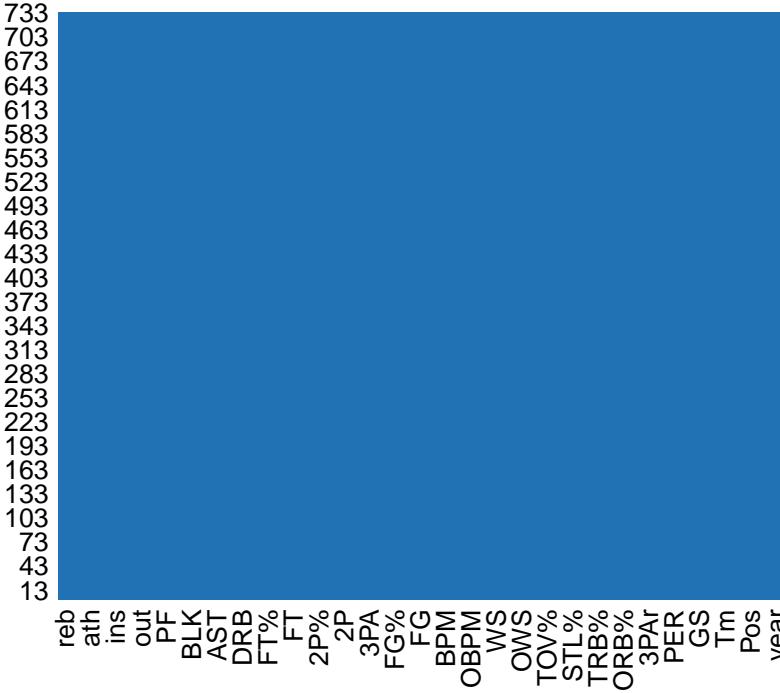
```

```

missmap(df_final, main = "Missing values ")

```

Missing values



■ Missing (0%)
■ Observed (100%)

```
# 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 DBPM 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':    733 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 ...
## $ 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':    733 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)

##
## Attaching package: 'purrr'

## The following object is masked from 'package:rvest':
##
##     pluck

library(tidyr)
library(ggplot2)
df_final %>%
  keep(is.numeric) %>%
  gather() %>%
  ggplot(aes(value)) +

```

```

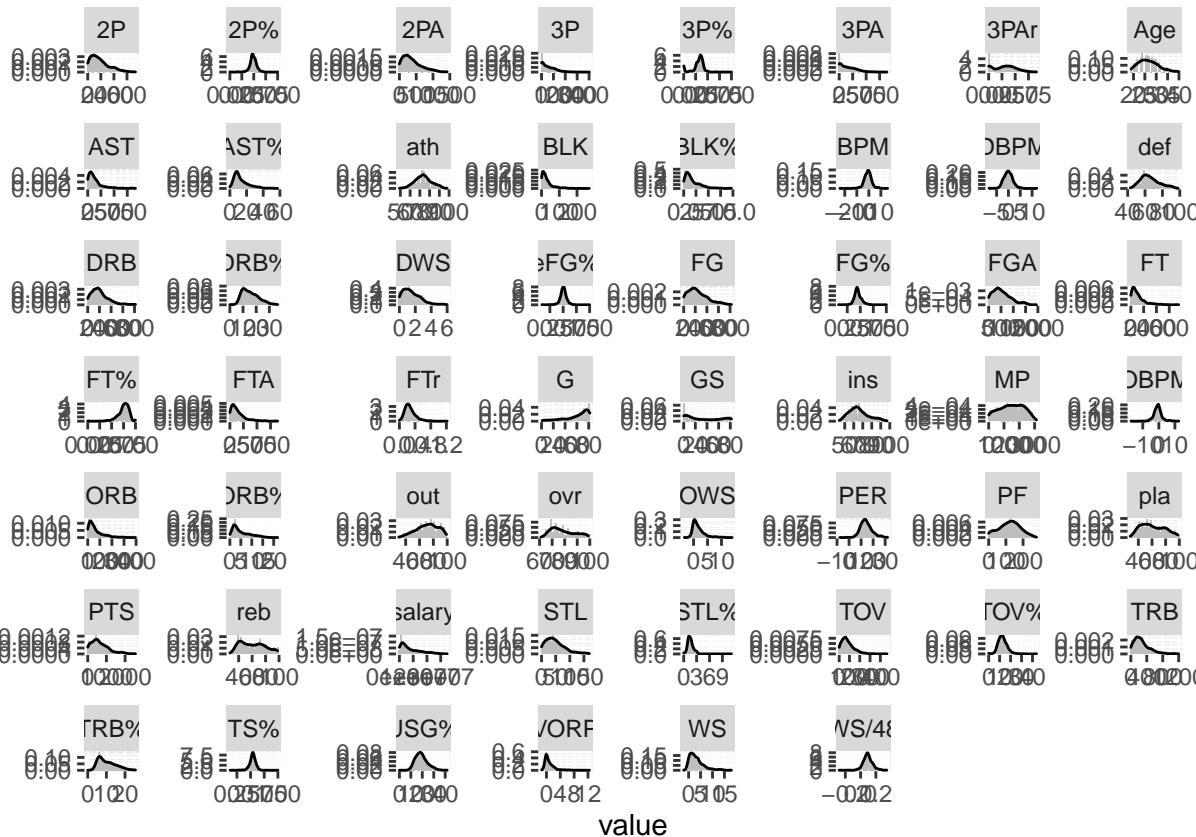
facet_wrap(~ key, scales = "free") +
geom_histogram(aes(y=..density..), fill = "grey") +
geom_density()

```

```

## `stat_bin()` using `bins = 30` . Pick better value with `binwidth` .

```



```

ggsave("../figures/hist_complete_vars.png", width=15, height=13)

```

```

## `stat_bin()` using `bins = 30` . Pick better value with `binwidth` .

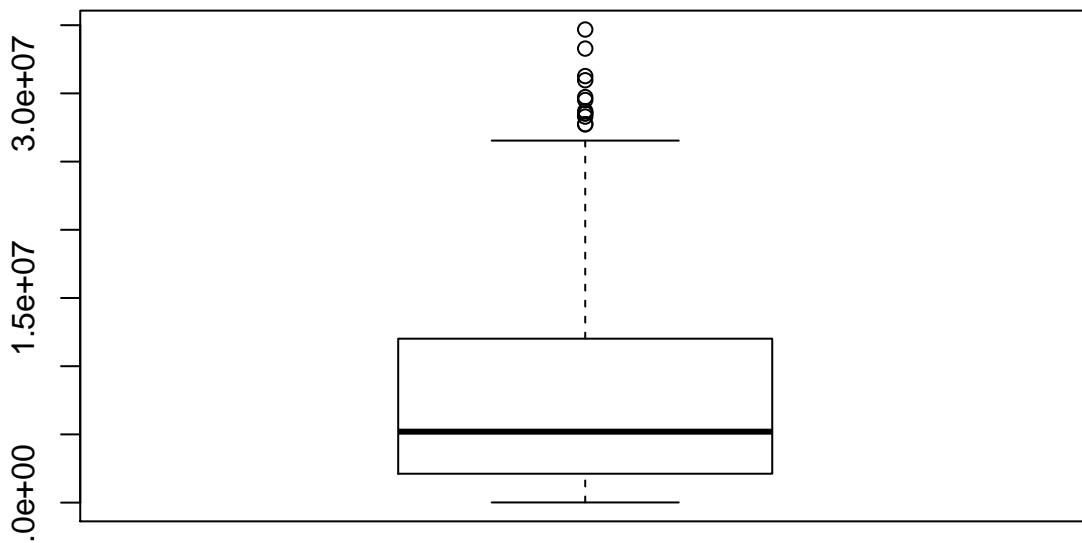
```

Salary variable

```

library(ggplot2)
boxplot(df_final$salary)

```



```

ggplot(df_final, aes(x = salary)) + geom_histogram(fill = "grey")

```

```

## `stat_bin()` using `bins = 30` . Pick better value with `binwidth` .

```

count

100
50
0

0e+00 1e+07 2e+07 3e+07

salary

```
ggplot(df_final, aes(x = salary)) + geom_density()
```

density

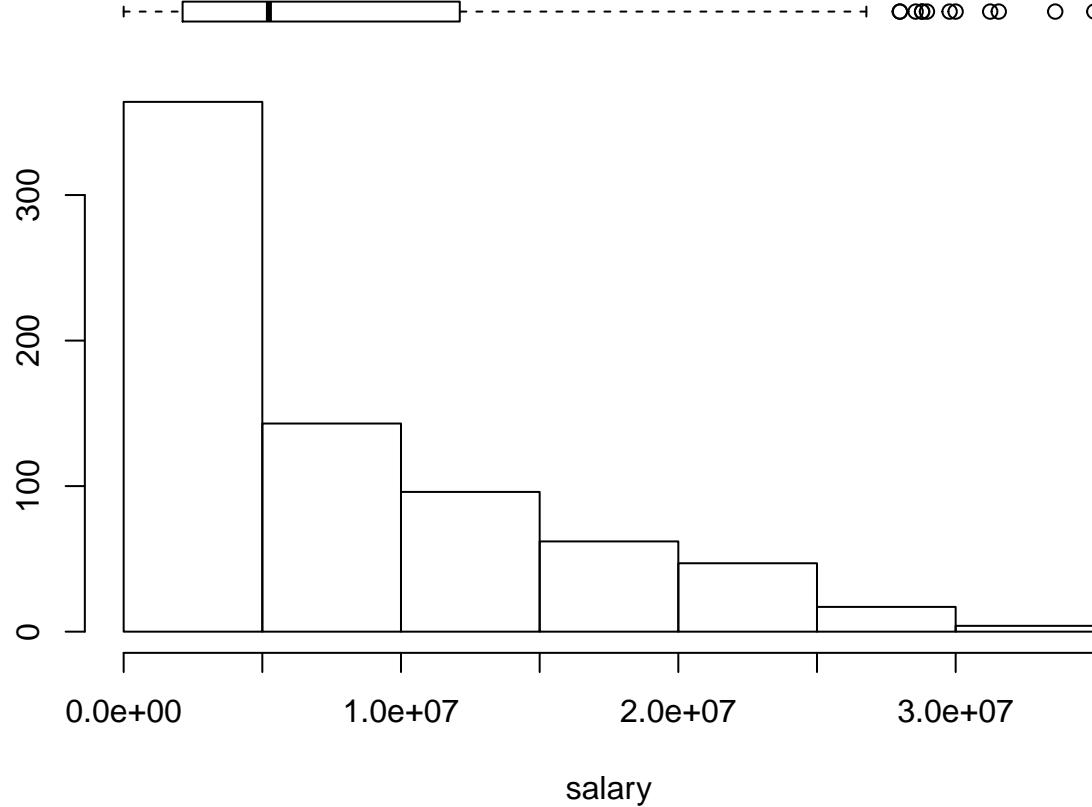
7.5e-08
5.0e-08
2.5e-08
0.0e+00

0e+00 1e+07 2e+07 3e+07

salary

```
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(df_final$salary , horizontal=TRUE , xaxt="n", frame=F, main=sprintf('Histogram of salary'))
par(mar=c(4, 3.1, 1.1, 2.1))
hist(df_final$salary,main=' ', xlab = "salary", ylab = "count")
```

Histogram of salary



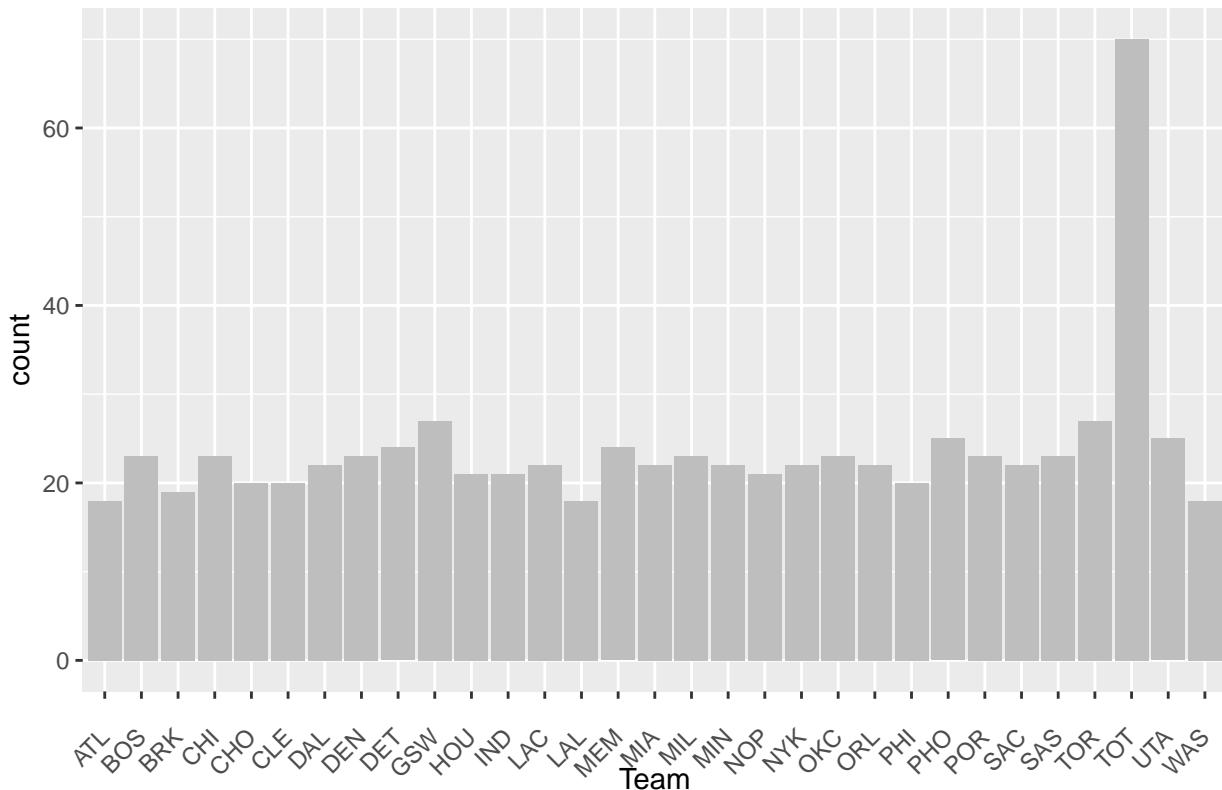
salary stats

```
mean(df_p_final$salary)
## [1] 7837816
median(df_p_final$salary)
## [1] 5200000
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
result <- getmode(df_p_final$salary)
print(result)
## [1] 1312611
```

Bar Chart of Player by Team from Complete Dataset

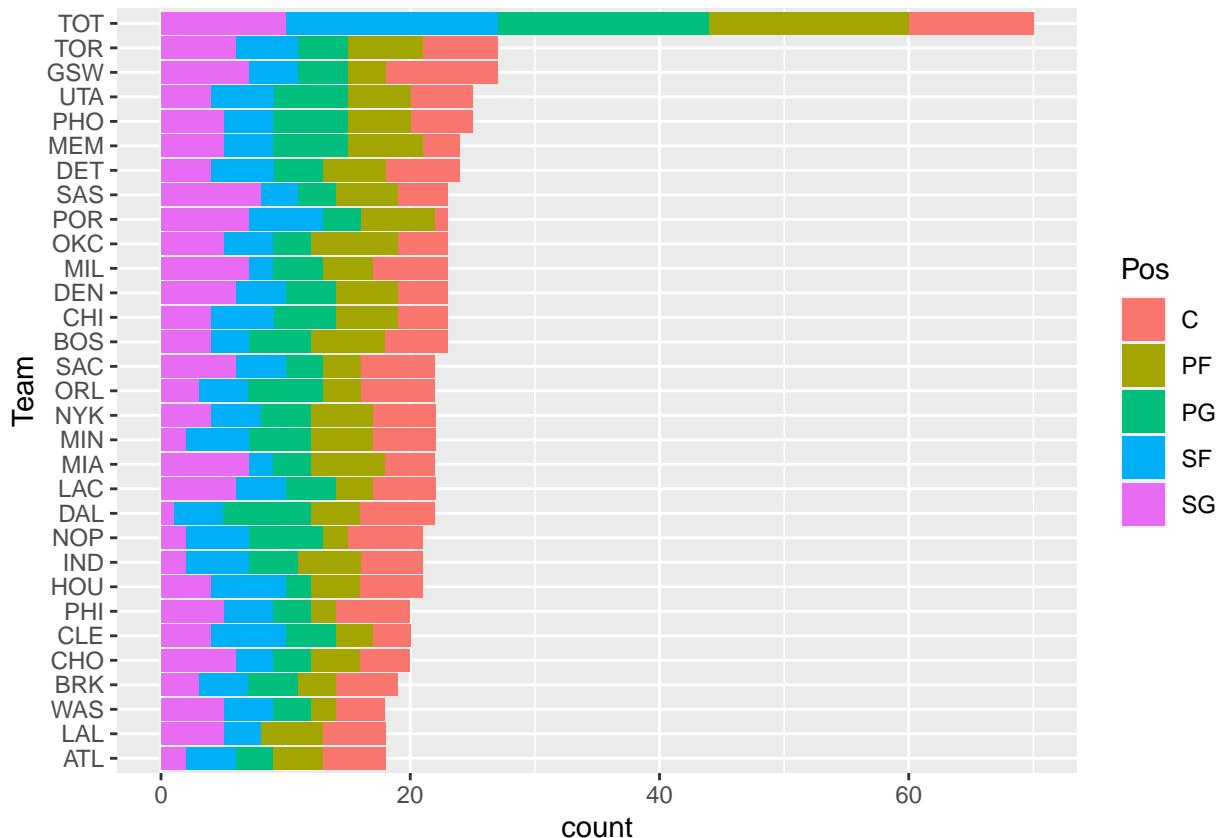
```
library(ggplot2)
ggplot(df_final, aes(x = Tm)) +
  geom_bar(fill = "grey") +
  labs(x = "Team", title = "Players per team") +
  theme(axis.text.x=element_text(angle=45,hjust=1,vjust=0.5))
```

Players per team



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

```
ggplot(df_final, aes(x=reorder(Tm,Tm,FUN = length), fill = Pos)) + geom_bar() + xlab("Team") + coord_flip()
```



Sum of Salaries per Team for Complete Dataset

```
library(ggplot2)
ggplot(df_final,aes(x = Tm, y = salary, fill = Tm)) + stat_summary(fun.y = "sum", geom = "bar") +
```

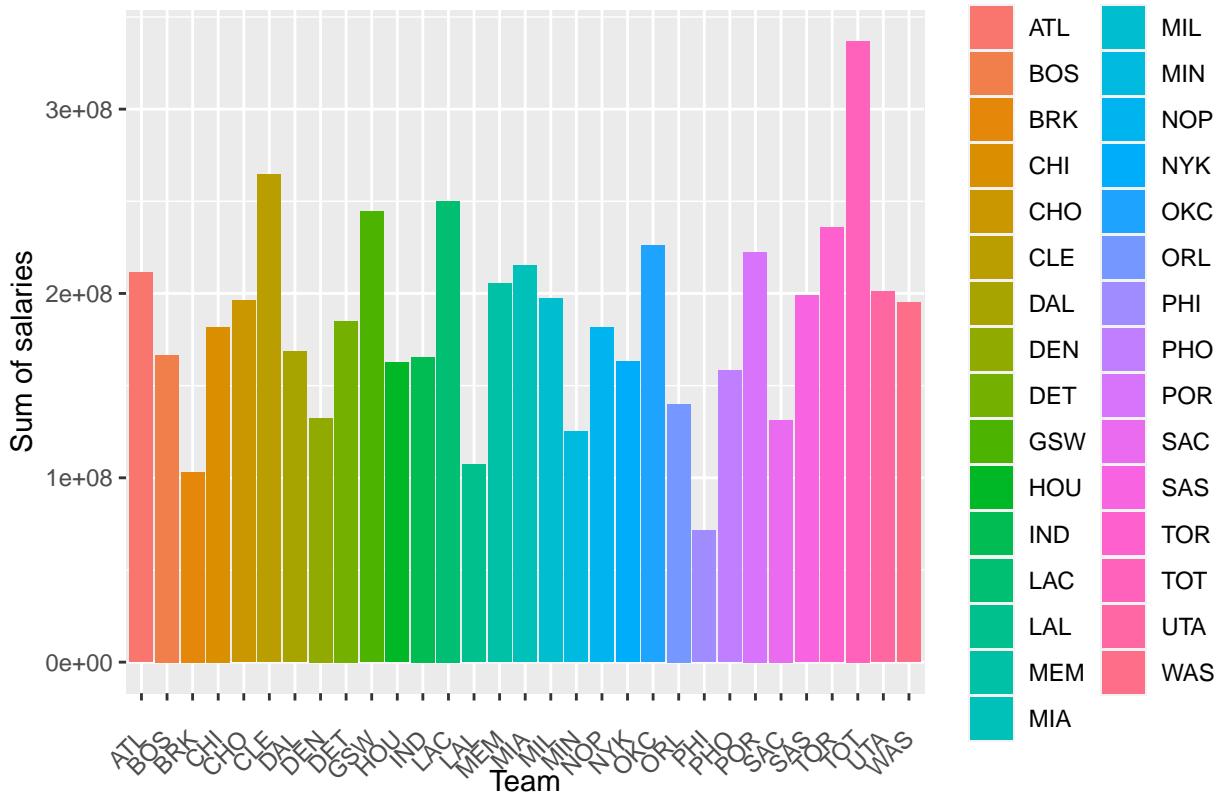
```

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))

```

Warning: `fun.y` is deprecated. Use `fun` instead.

Sum of salaries per team



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

Mean Salaries per Team for Complete Dataset

```

library(ggplot2)
ggplot(df_final,aes(x = Tm, y = salary, fill = Tm)) + stat_summary(fun.y = "mean", geom = "bar") +
  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))

```

Warning: `fun.y` is deprecated. Use `fun` instead.

mean salary per team



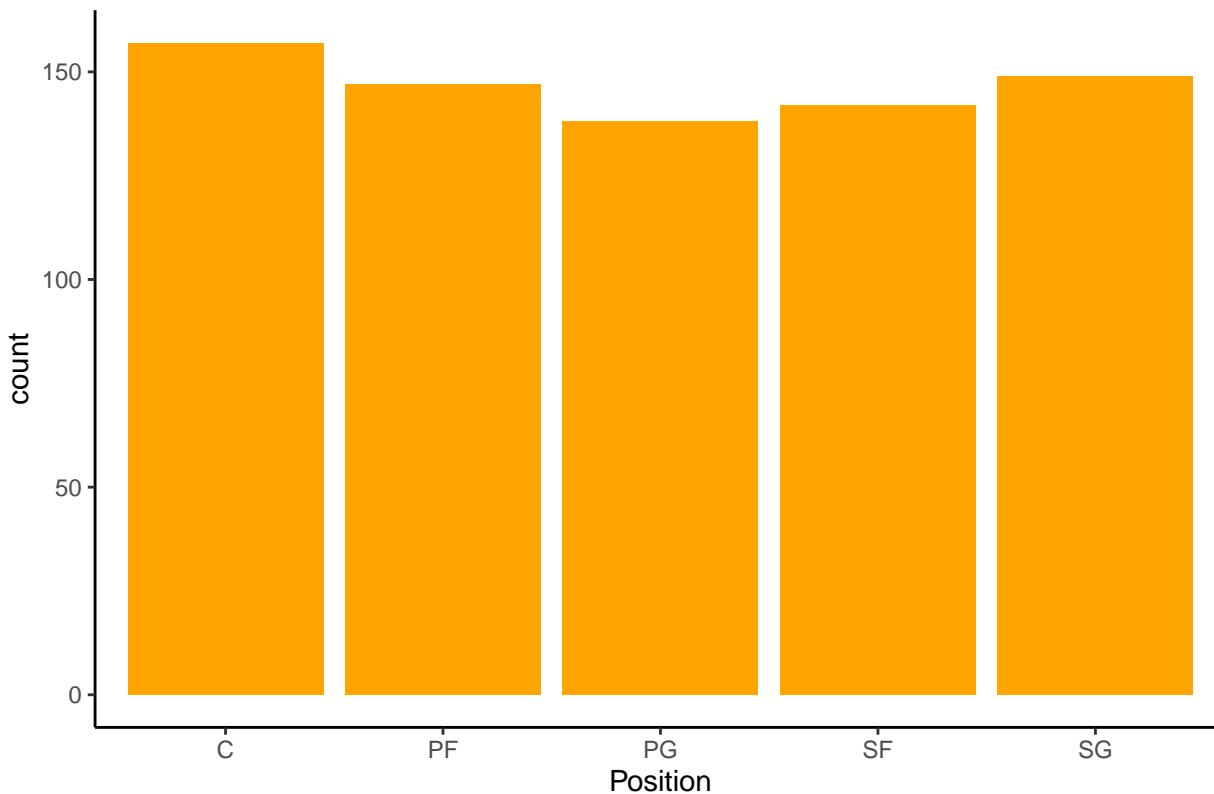
```
ggsave("../figures/bar_complete_mean_salaries_per_team.png", width=10, height=7)
```

Players in each position

```
nrow(df_p_final)
## [1] 733

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

No of Player in each position



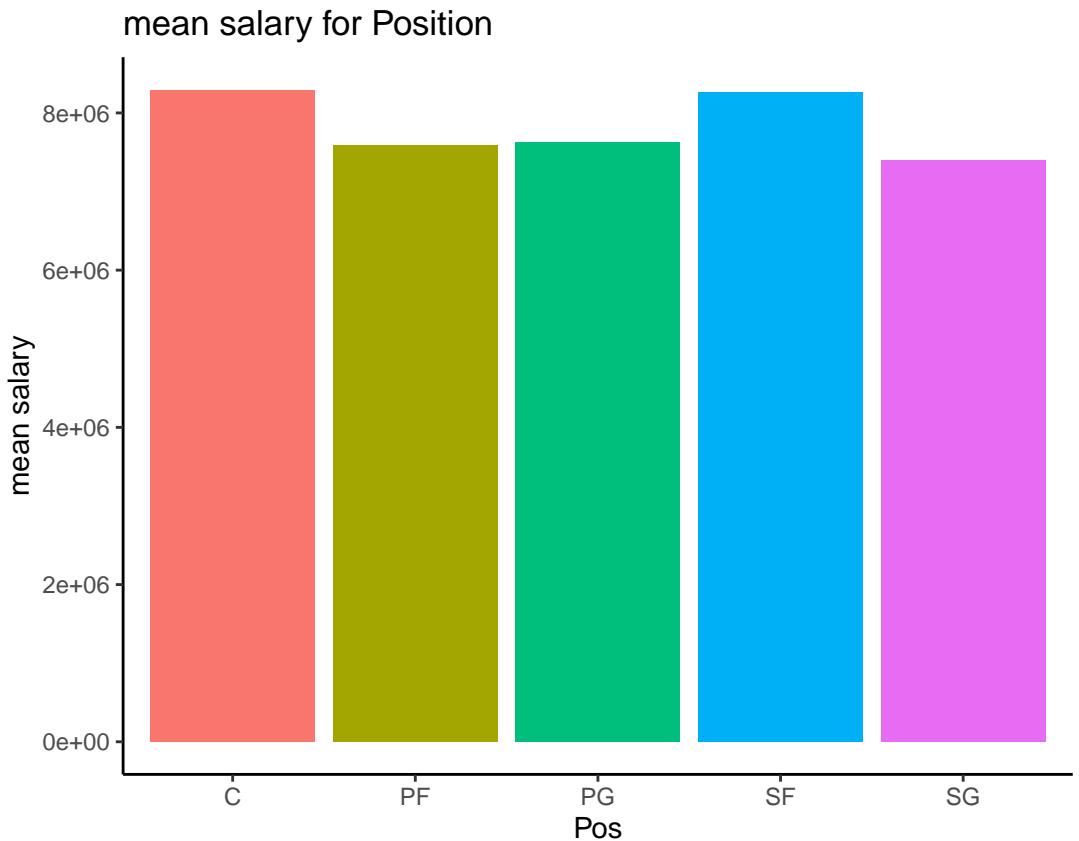
```
ggsave("../figures/bar_complete_player_Position.png", width=10, height=7)
```

Mean salaries for each position

```
library(ggplot2)
library(tidyr)
library(dplyr)

## 
## Attaching package: 'dplyr'
## The following object is masked from 'package:reshape':
## 
##     rename
## The following objects are masked from 'package:stats':
## 
##     filter, lag
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
ggplot(df_final,aes(x = Pos, y = salary, fill = Pos)) + stat_summary(fun.y = "mean", geom = "bar") +
  labs(
    x = "Pos",
    y = "mean salary",
    title = paste(
      "mean salary for Position")) +
  theme_classic()

## Warning: `fun.y` is deprecated. Use `fun` instead.
```



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

Correlation Matrix for complete dataset

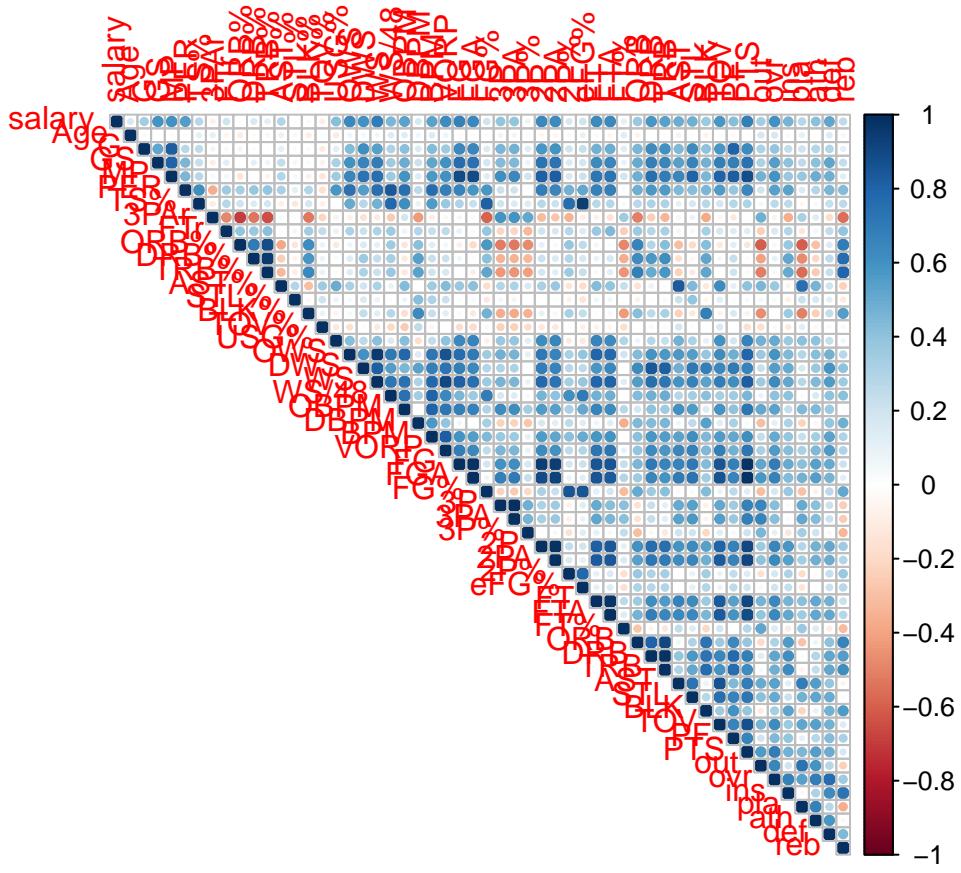
```
corr_matrix_c <- cor(Filter(is.numeric,df_final[2:ncol(df_final)]),method = "pearson")
correlation_salary_c <- sort(corr_matrix_c[, 'salary'],decreasing = TRUE)
correlation_salary_c

##      salary         WS        PTS        FG        FGA        2P
## 1.00000000 0.69654545 0.68147201 0.67871374 0.64964867 0.64945688
##      FTA         OWS        FT        2PA        VORP        MP
## 0.64822250 0.64762530 0.63630269 0.63534810 0.62400574 0.60387680
##      ovr         DWS        GS        TOV        DRB        PER
## 0.60386589 0.60093014 0.59461775 0.58439384 0.58130461 0.55068657
##      TRB         BPM        OBPM       def        AST        STL
## 0.54195071 0.53967876 0.53657917 0.52198890 0.49574308 0.48956302
##      ins        WS/48       USG%        PF        ath        3P
## 0.46613871 0.45295278 0.42790513 0.42275773 0.41069301 0.39214372
##      3PA         ORB        BLK         G        out        AST%
## 0.39058935 0.37213223 0.36870041 0.34675456 0.34594270 0.29653627
##      pla        TS%        reb       eFG%        FTr        FG%
## 0.28565474 0.27181444 0.25525211 0.20570361 0.20202229 0.19929701
##      DBPM        Age       DRB%       2P%        TRB%        FT%
## 0.17061810 0.16977681 0.16761161 0.16597870 0.12064951 0.11487058
##      3P%        BLK%       STL%       ORB%       3Par       TOV%
## 0.08619947 0.03554661 0.01677941 0.01510555 -0.08755677 -0.09488064
```

Correlation Plot for complete dataset

```
library(corrplot)

## corrplot 0.84 loaded
corrplot(corr_matrix_c,type = "upper")
```

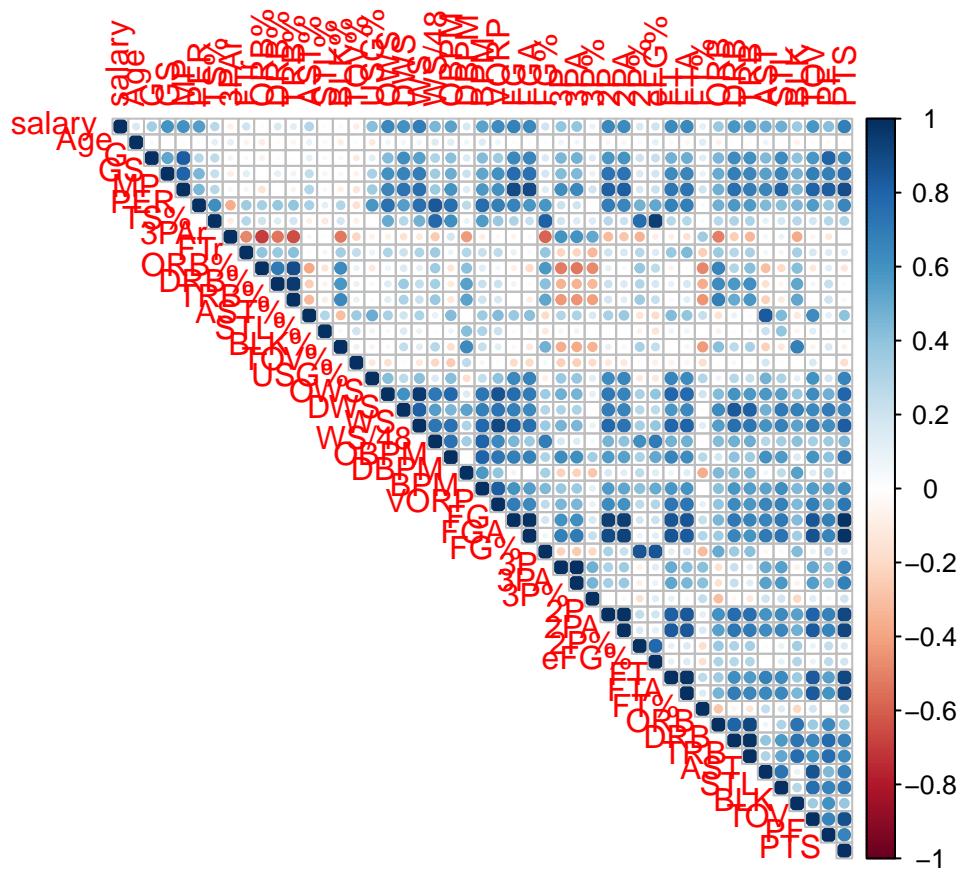


Correlation plot for Primary dataset

```
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 0.69654545 0.68147201 0.67871374 0.64964867 0.64945688
##      FTA        OWS       FT       2PA      VORP       MP
## 0.64822250 0.64762530 0.63630269 0.63534810 0.62400574 0.60387680
##      DWS        GS       TOV       DRB       PER       TRB
## 0.60093014 0.59461775 0.58439384 0.58130461 0.55068657 0.54195071
##      BPM        OBPM      AST       STL      WS/48      USG%
## 0.53967876 0.53657917 0.49574308 0.48956302 0.45295278 0.42790513
##      PF         3P       3PA      ORB       BLK        G
## 0.42275773 0.39214372 0.39058935 0.37213223 0.36870041 0.34675456
##      AST%       TS%     eFG%      FTr       FG%      DBPM
## 0.29653627 0.27181444 0.20570361 0.20202229 0.19929701 0.17061810
##      Age      DRB%      2P%      TRB%      FT%      3P%
## 0.16977681 0.16761161 0.16597870 0.12064951 0.11487058 0.08619947
##      BLK%      STL%      ORB%      3PAr      TOV%
## 0.03554661 0.01677941 0.01510555 -0.08755677 -0.09488064
```

```
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
```

Salary vs all other variable scatterplots

```
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){
```

```

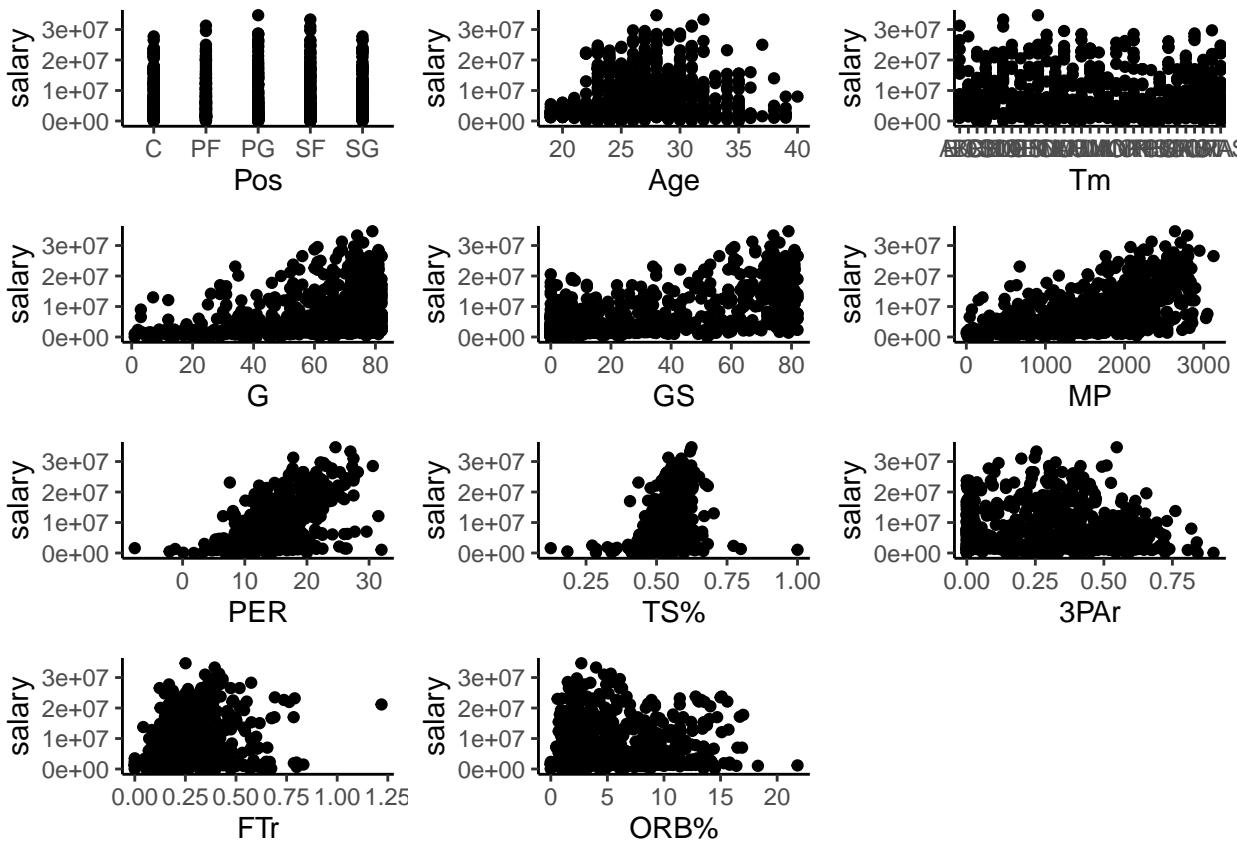
name = names(d[i+val])
p[[i]] = plot(as.character(name))]
val = i+val
do.call(grid.arrange,p)
p <- NULL}

```

```

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

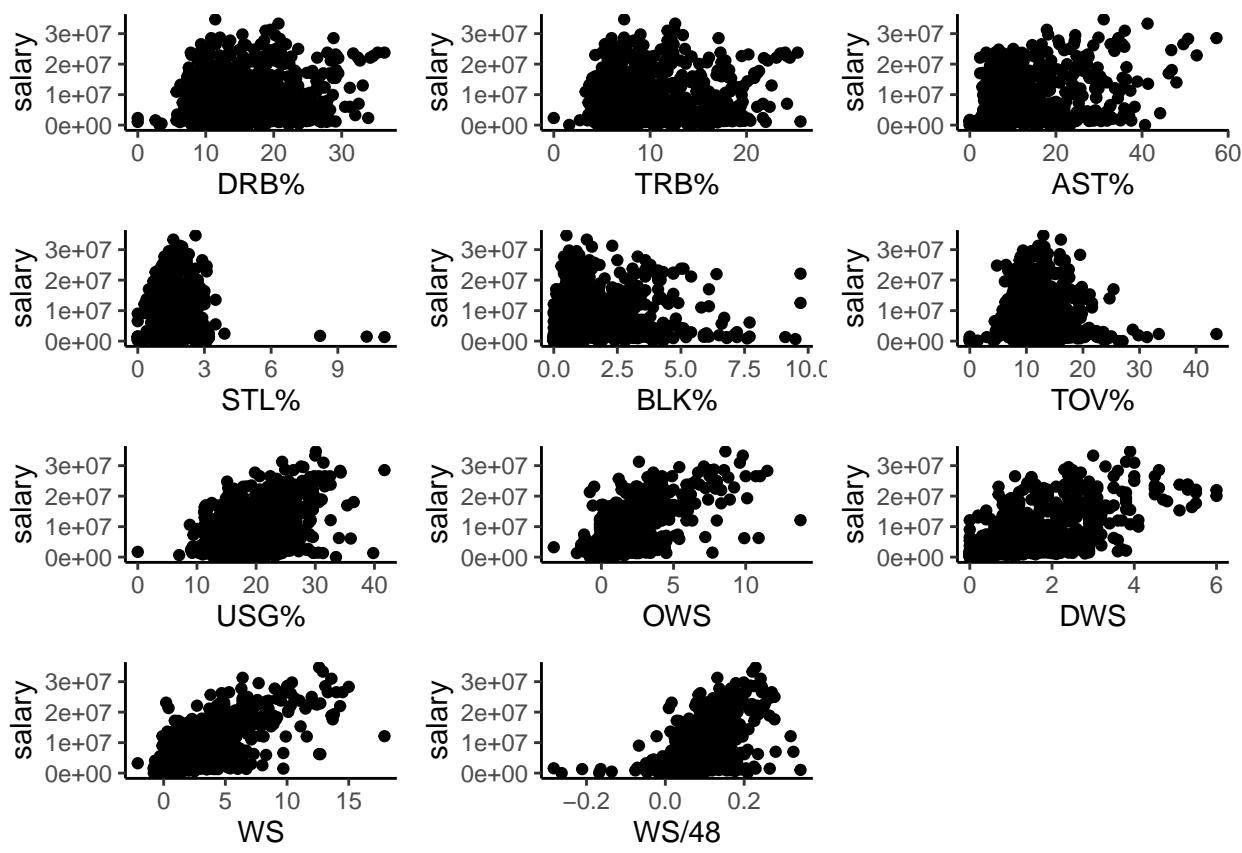
```



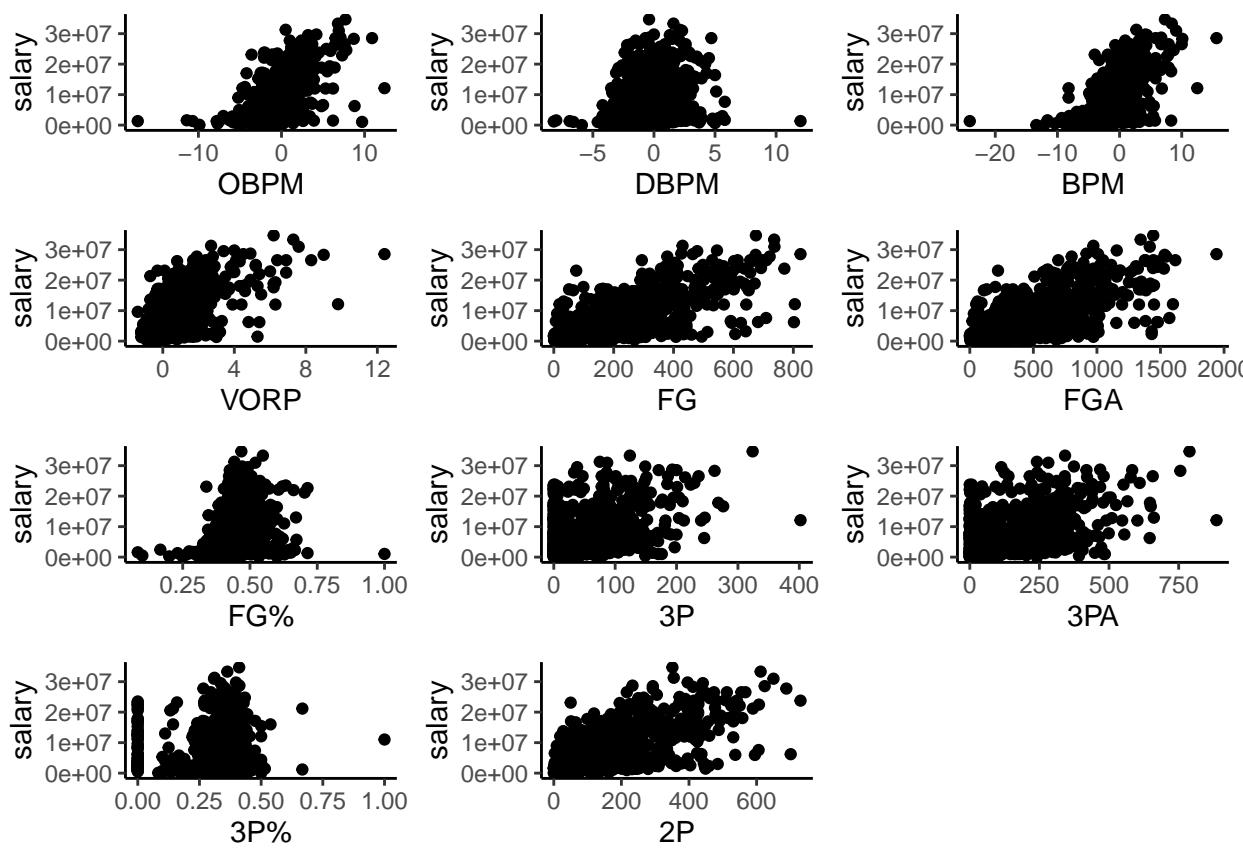
```

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

```



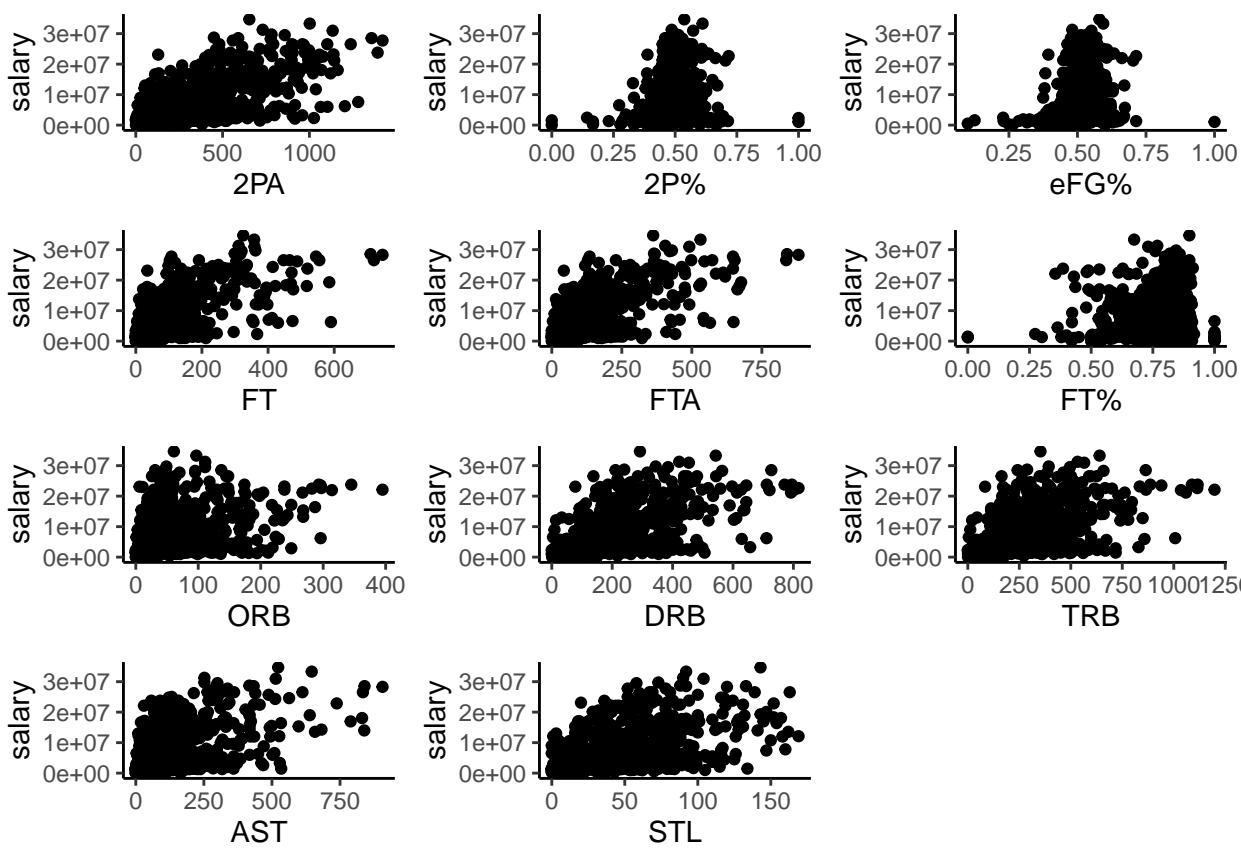
```
## [1] "OBPM"
## [1] "DBPM"
## [1] "BPM"
## [1] "VORP"
## [1] "FG"
## [1] "FGA"
## [1] "FG%"
## [1] "3P"
## [1] "3PA"
## [1] "3P%"
## [1] "2P"
```



```

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

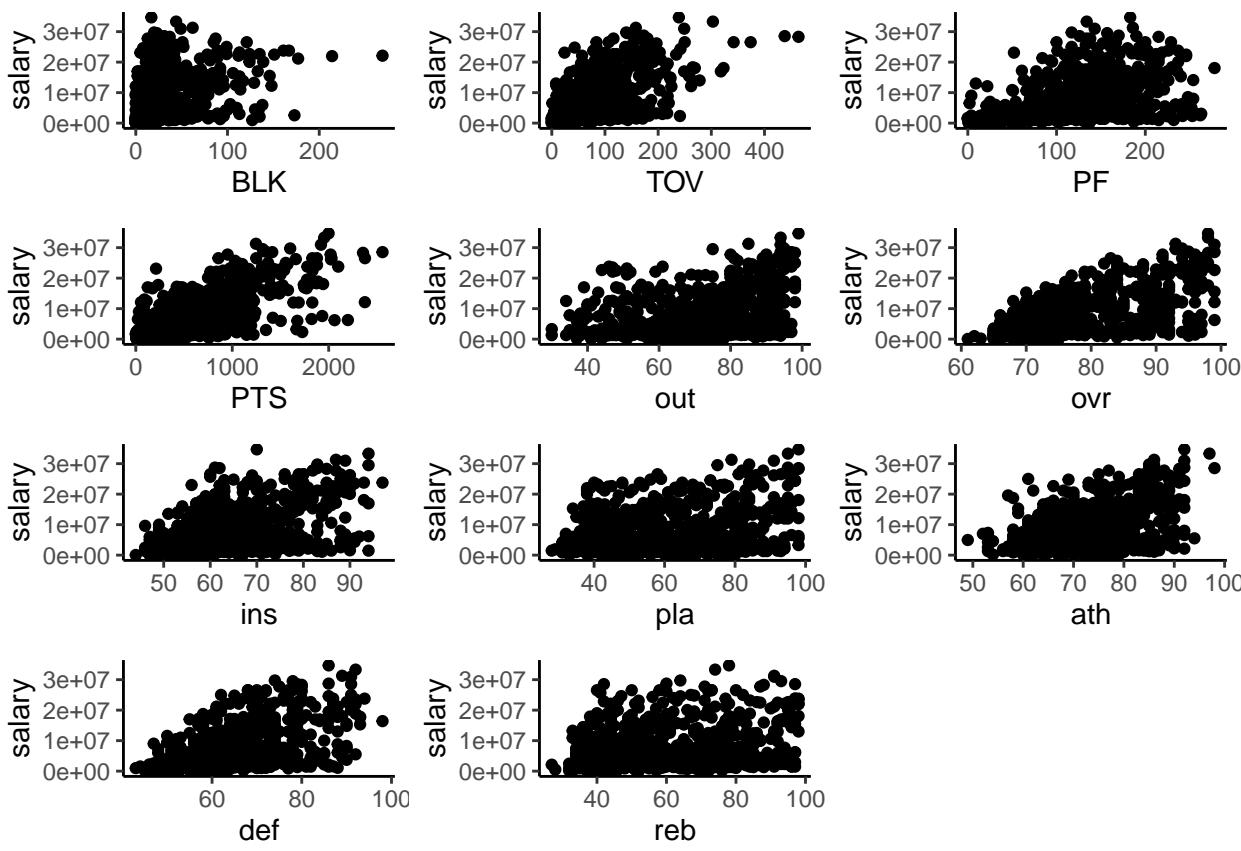
```



```

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

```



selecting relevant variables

```

rel_p <- c("name", "year", "salary", "Pos", "Age", "Tm", "GS", "MP", "PER", "TS%", "TRB%", "AST%", "STL%", "BLK%", "TOV%", "USG%")

rel_c <- c("name", "year", "salary", "Pos", "Age", "Tm", "GS", "MP", "PER", "TS%", "TRB%", "AST%", "STL%", "BLK%", "TOV%", "USG%")

df_p_final_relevant <- df_p_final[names(df_p_final) %in% rel_p]
head(df_p_final_relevant)

```

```

##          name year  salary Pos Age   Tm GS    MP   PER   TS% TRB% AST% STL% BLK%
## 1 aaron brooks 2016 2700000 PG 31 CHI 0 1108 11.8 0.494  4.8 26.0  1.4  0.7
## 2 aaron brooks 2017 2116955 PG 32 IND 0  894  9.5 0.507  4.3 20.7  1.4  0.9
## 3 aaron gordon 2016 4351320 PF 20 ORL 37 1863 17.0 0.541 15.1 10.3  1.6  2.4
## 4 aaron gordon 2017 5504420 SF 21 ORL 72 2298 14.4 0.530  9.6 10.5  1.4  1.4
## 5 adreian payne 2016 2022240 PF 24 MIN 2  486  5.6 0.422 13.3  8.9  1.7  1.8
## 6 aj hammons 2017 1312611 C  24 DAL 0  163  8.4 0.472 12.8  3.8  0.3  7.2
##      TOV% USG% WS BPM VORP FG   FG% 3P% 2P% 2P% eFG%  FT  FT% AST  PF
## 1 14.2 22.9 0.9 -3.3 -0.4 188 0.401 66 0.357 122 0.430 0.471 49 0.766 180 132
## 2 17.2 19.2 0.3 -4.6 -0.6 121 0.403 48 0.375  73 0.424 0.483 32 0.800 125  93
## 3  9.0 17.3 5.4  1.8  1.8 274 0.473 42 0.296 232 0.531 0.509 129 0.668 128 153
## 4  8.5 20.1 3.7 -0.7  0.8 393 0.454 77 0.288 316 0.528 0.499 156 0.719 150 172
## 5 18.7 17.7 -0.5 -6.1 -0.5  53 0.366  9 0.281  44 0.389 0.397 17 0.654  29  77
## 6 16.4 17.6  0.0 -5.6 -0.1  17 0.405  5 0.500 12 0.375 0.464  9 0.450   4  21
##      PTS
## 1 491
## 2 322
## 3 719
## 4 1019
## 5 132
## 6 48

```

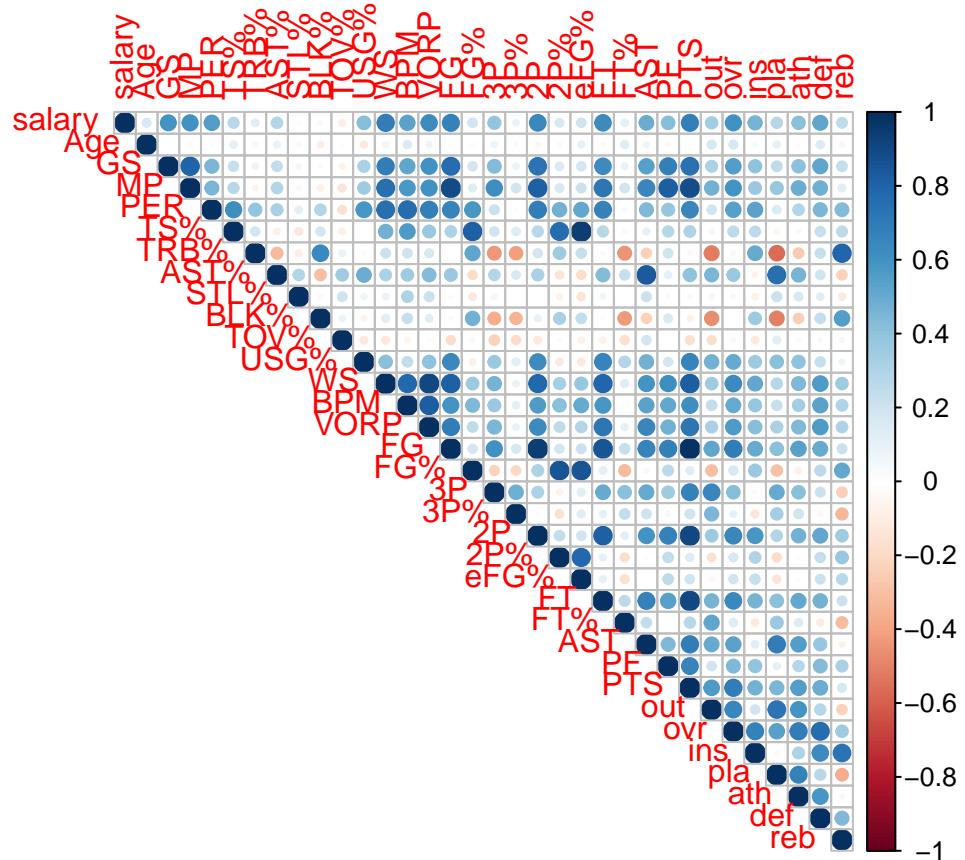
```
df_final_relevant <- df_final[names(df_final) %in% rel_c]
```

correlation plot of relevant variables

```
library(corrplot)
# primary dataset
corr_matrix_p_relevant <- cor(Filter(is.numeric,df_p_final_relevant[2:ncol(df_p_final_relevant)]),method = "pearson")
correlation_salary_p_relevant <- sort(df_p_final_relevant[, 'salary'],decreasing = TRUE)

# complete dataset
corr_matrix_c_relevant <- cor(Filter(is.numeric,df_final_relevant[2:ncol(df_final_relevant)]),method = "pearson")
correlation_salary_c_relevant <- sort(df_final_relevant[, 'salary'],decreasing = TRUE)

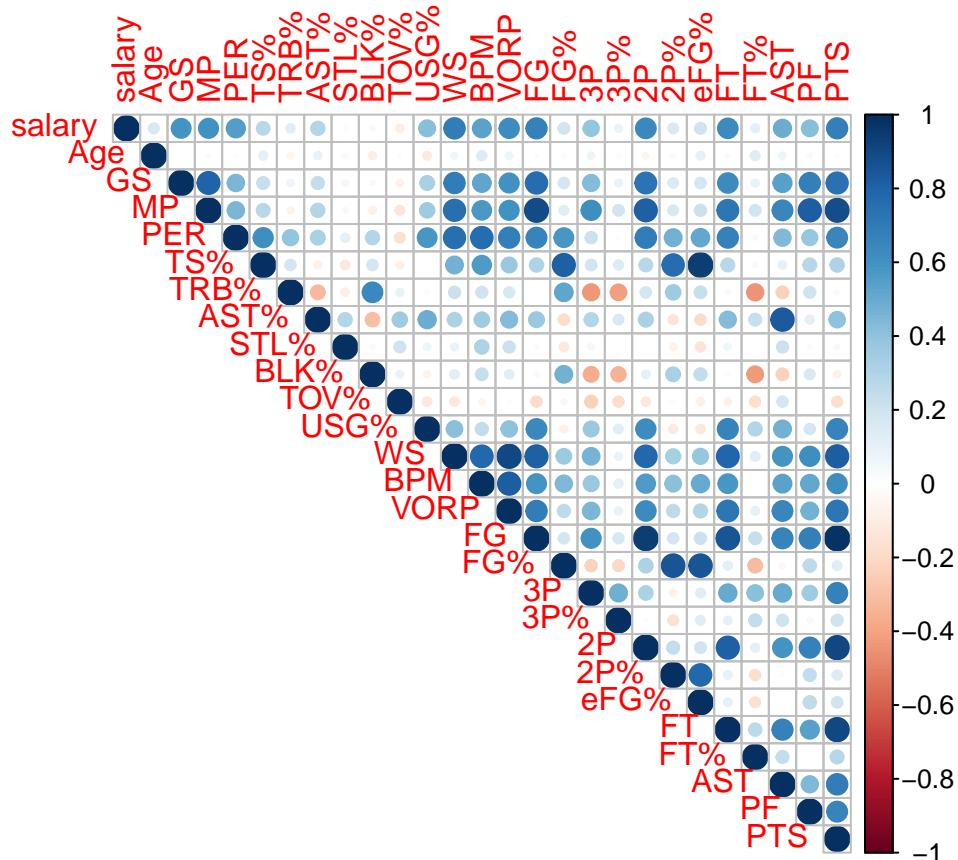
corrplot(corr_matrix_c_relevant,type = "upper")
```



```
dev.copy(png,"../figures/Correlation_plot_c_relevant.png")
```

```
## png
## 4
dev.off()

## pdf
## 2
# primary dataset
corrplot(corr_matrix_p_relevant,type = "upper")
```



```
dev.copy(png, "../figures/Correlation_plot_p_relevant.png")
```

```
## png
## 4
dev.off()

## pdf
## 2
```

Detecting outliers

```
df_cook <- df_final_relevant[which(sapply(df_final_relevant, is.numeric))]
mod <- lm(salary ~ ., data = df_cook )
cooksrd <- cooks.distance(mod)

# All outliers
influential <- as.numeric(names(cooksrd)[(cooksrd > 9*mean(cooksrd, na.rm=T))]) # influential row numbers
influential <- na.omit(influential)
df_final_relevant[influential, names(df_final_relevant) %in% c("name", "year", "salary", "Pos", "Age", "Tm", "GS", "MP", "M
##          name year   salary Pos Age Tm GS MP AST% WS FG%
## 260      isaiah thomas 2017 6261395 PG 27 BOS 76 2569 32.6 12.6 0.463
## 294      jarrett jack 2017 2328652 PG 33 NOP 0 33 20.3 0.0 0.667
## 384      karl anthony towns 2017 6216840 C 21 MIN 82 3030 13.2 12.7 0.542
## 530      nikola mirotic 2016 5782450 PF 24 CHI 38 1646 9.4 3.9 0.407
## 616      sam dekker 2017 1794600 SF 22 HOU 2 1419 7.7 3.1 0.473
## 641      stephen curry 2017 34682550 PG 28 GSW 79 2638 31.1 12.6 0.468
## 642      stephen zimmerman 2017 1312611 C 20 ORL 0 108 5.3 0.0 0.323
##          3P% 2P% FT% PF PTS
## 260 0.3790000 0.528 0.909 167 2199
## 294 0.0000000 1.000 1.000 4 6
## 384 0.3670000 0.582 0.832 241 2061
## 530 0.3900000 0.430 0.807 151 777
## 616 0.3210000 0.591 0.559 83 504
## 641 0.4110000 0.537 0.898 183 1999
```

```
## 642 0.2711419 0.323 0.600 17 23
```

Plot Cook's Distance

```
plot(cooksd, pch="*", cex=2, main="Outliers") # plot cook's distance
abline(h = 8*mean(cooksd, na.rm=T), col="red")
# add cutoff line. Those observations that have a cook's distance greater than 8
# times the mean may be classified as influential.
text(x=1:length(cooksd)+1, y=cooksd, labels=ifelse(cooksd>8*mean(cooksd, na.rm=T), names(cooksd), ""), col="red")
dev.copy(png, "../figures/Missing values/outliers.png")
dev.off()
```

Variable Selection

Split Train-Test

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

Primary Dataset

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

```
## [1] 588
```

```
nrow(p_test_df)
```

```
## [1] 145
```

```
names(p_train_df)
```

```
## [1] "name"    "year"     "salary"   "Pos"      "Age"      "Tm"       "G"        "GS"
## [9] "MP"       "PER"      "TS%"      "3PAr"     "FTr"      "ORB%"     "DRB%"    "TRB%"
## [17] "AST%"    "STL%"     "BLK%"     "TOV%"     "USG%"     "OWS"      "DWS"      "WS"
## [25] "WS/48"   "OBPM"    "DBPM"     "BPM"      "VORP"     "FG"       "FGA"     "FG%"
## [33] "3P"       "3PA"      "3P%"      "2P"       "2PA"      "2P%"     "eFG%"    "FT"
## [41] "FTA"      "FT%"      "ORB"      "DRB"      "TRB"      "AST"      "STL"      "BLK"
## [49] "TOV"      "PF"       "PTS"
```

```
head(p_train_df)
```

```
##          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

write.csv(p_train_df,'../data/train_test/primary/train.csv',row.names=F)
write.csv(p_test_df,'../data/train_test/primary/test.csv',row.names=F)

train_rows_relevant <- createDataPartition(y=df_p_final_relevant[, 'salary'], list=FALSE, p=.8)
p_train_relevant_df <- df_p_final_relevant[train_rows,]
p_test_relevant_df <- df_p_final_relevant[-train_rows,]
stopifnot(nrow(p_train_relevant_df) + nrow(p_test_relevant_df) == nrow(df_p_final_relevant))
head(p_train_relevant_df)

##           name year salary Pos Age Tm GS MP PER TS% TRB% AST% STL% BLK%
## 1 aaron brooks 2016 2700000 PG 31 CHI 0 1108 11.8 0.494 4.8 26.0 1.4 0.7
## 2 aaron brooks 2017 2116955 PG 32 IND 0 894 9.5 0.507 4.3 20.7 1.4 0.9
## 3 aaron gordon 2016 4351320 PF 20 ORL 37 1863 17.0 0.541 15.1 10.3 1.6 2.4
## 4 aaron gordon 2017 5504420 SF 21 ORL 72 2298 14.4 0.530 9.6 10.5 1.4 1.4
## 5 adreian payne 2016 2022240 PF 24 MIN 2 486 5.6 0.422 13.3 8.9 1.7 1.8
## 6 aj hammons 2017 1312611 C 24 DAL 0 163 8.4 0.472 12.8 3.8 0.3 7.2
##   TOV% USG% WS BPM VORP FG FG% 3P% 2P% 2P% eFG% FT FT% AST PF
## 1 14.2 22.9 0.9 -3.3 -0.4 188 0.401 66 0.357 122 0.430 0.471 49 0.766 180 132
## 2 17.2 19.2 0.3 -4.6 -0.6 121 0.403 48 0.375 73 0.424 0.483 32 0.800 125 93
## 3 9.0 17.3 5.4 1.8 1.8 274 0.473 42 0.296 232 0.531 0.509 129 0.668 128 153
## 4 8.5 20.1 3.7 -0.7 0.8 393 0.454 77 0.288 316 0.528 0.499 156 0.719 150 172
## 5 18.7 17.7 -0.5 -6.1 -0.5 53 0.366 9 0.281 44 0.389 0.397 17 0.654 29 77
## 6 16.4 17.6 0.0 -5.6 -0.1 17 0.405 5 0.500 12 0.375 0.464 9 0.450 4 21
##   PTS
## 1 491
## 2 322
## 3 719
## 4 1019
## 5 132
## 6 48

head(p_test_relevant_df)

##           name year salary Pos Age Tm GS MP PER TS% TRB% AST% STL%
## 7 al farouq aminu 2016 7680965 SF 25 POR 82 2341 12.7 0.533 11.5 8.8 1.5
## 13 alan anderson 2016 1315448 SG 33 WAS 0 192 9.2 0.495 7.9 10.3 1.0
## 24 alonzo gee 2016 1400000 SF 28 NOP 38 1632 8.4 0.572 8.3 6.2 1.9
## 29 andre drummond 2016 22116750 C 22 DET 81 2666 21.2 0.499 24.5 4.4 2.3
## 32 andre iguodala 2017 14814815 SF 33 GSW 0 1998 14.4 0.624 8.3 16.7 1.8
## 45 anthony davis 2016 22116750 C 22 NOP 61 2164 25.0 0.559 16.1 10.0 1.8
##   BLK% TOV% USG% WS BPM VORP FG FG% 3P% 3P% 2P% 2P% eFG% FT FT%
## 7 1.8 13.2 16.9 4.0 0.2 1.3 299 0.416 126 0.361 173 0.468 0.503 115 0.737
## 13 0.4 3.0 15.4 0.3 -2.8 0.0 21 0.356 12 0.324 9 0.409 0.458 11 0.733
## 24 0.7 16.2 9.2 1.9 -0.8 0.5 132 0.518 17 0.283 115 0.590 0.551 44 0.667
## 29 3.2 10.5 24.1 7.4 -0.5 1.0 552 0.521 2 0.333 550 0.522 0.522 208 0.355
## 32 1.5 11.2 11.2 6.9 3.0 2.5 219 0.528 64 0.362 155 0.651 0.605 72 0.706
## 45 4.7 8.4 29.6 7.2 2.2 2.3 560 0.493 35 0.324 525 0.511 0.508 326 0.758
##   AST PF PTS
## 7 138 171 839
## 13 14 25 65
## 24 72 169 325

```

```

## 29 67 245 1314
## 32 262 97 574
## 45 116 148 1481
write.csv(p_train_relevant_df,'../data/train_test/primary/train_relevant.csv',row.names=F)
write.csv(p_test_relevant_df,'../data/train_test/primary/test_relevant.csv',row.names=F)

```

Complete Dataset

```

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

## [1] 588
nrow(c_test_df)

## [1] 145
names(c_train_df)

## [1] "name"    "year"     "salary"   "Pos"      "Age"      "Tm"       "G"        "GS"
## [9] "MP"       "PER"      "TS%"      "3PAr"     "FTr"      "ORB%"     "DRB%"    "TRB%"
## [17] "AST%"    "STL%"    "BLK%"    "TOV%"    "USG%"    "OWS"      "DWS"      "WS"
## [25] "WS/48"   "OBPM"    "DBPM"    "BPM"     "VORP"     "FG"       "FGA"      "FG%"
## [33] "3P"       "3PA"      "3P%"      "2P"       "2PA"      "2P%"      "eFG%"    "FT"
## [41] "FTA"      "FT%"      "ORB"      "DRB"      "TRB"      "AST"      "STL"      "BLK"
## [49] "TOV"      "PF"       "PTS"      "out"      "ovr"      "ins"      "pla"      "ath"
## [57] "def"      "reb"

head(c_train_df)

##           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

```

```

write.csv(c_train_df,'../data/train_test/complete/train.csv',row.names=F)
write.csv(c_test_df,'../data/train_test/complete/test.csv',row.names=F)

train_rows_relevant <- createDataPartition(y=df_final_relevant[, 'salary'], list=FALSE, p=.8)
c_train_relevant_df <- df_final_relevant[train_rows,]
c_test_relevant_df <- df_final_relevant[-train_rows,]
stopifnot(nrow(c_train_relevant_df) + nrow(c_test_relevant_df) == nrow(df_final_relevant))
head(c_train_relevant_df)

##          name year salary Pos Age Tm GS MP PER TS% TRB% AST% STL% BLK%
## 1 aaron brooks 2016 2700000 PG 31 CHI 0 1108 11.8 0.494 4.8 26.0 1.4 0.7
## 2 aaron brooks 2017 2116955 PG 32 IND 0 894 9.5 0.507 4.3 20.7 1.4 0.9
## 3 aaron gordon 2016 4351320 PF 20 ORL 37 1863 17.0 0.541 15.1 10.3 1.6 2.4
## 4 aaron gordon 2017 5504420 SF 21 ORL 72 2298 14.4 0.530 9.6 10.5 1.4 1.4
## 5 adreian payne 2016 2022240 PF 24 MIN 2 486 5.6 0.422 13.3 8.9 1.7 1.8
## 6 aj hammons 2017 1312611 C 24 DAL 0 163 8.4 0.472 12.8 3.8 0.3 7.2
##   TOV% USG% WS BPM VORP FG FG% 3P 3P% 2P 2P% eFG% FT FT% AST PF
## 1 14.2 22.9 0.9 -3.3 -0.4 188 0.401 66 0.357 122 0.430 0.471 49 0.766 180 132
## 2 17.2 19.2 0.3 -4.6 -0.6 121 0.403 48 0.375 73 0.424 0.483 32 0.800 125 93
## 3 9.0 17.3 5.4 1.8 1.8 274 0.473 42 0.296 232 0.531 0.509 129 0.668 128 153
## 4 8.5 20.1 3.7 -0.7 0.8 393 0.454 77 0.288 316 0.528 0.499 156 0.719 150 172
## 5 18.7 17.7 -0.5 -6.1 -0.5 53 0.366 9 0.281 44 0.389 0.397 17 0.654 29 77
## 6 16.4 17.6 0.0 -5.6 -0.1 17 0.405 5 0.500 12 0.375 0.464 9 0.450 4 21
##   PTS out ovr ins pla ath def reb
## 1 491 79 75 52 74 77 52 36
## 2 322 87 85 51 81 82 57 37
## 3 719 87 90 91 69 86 69 87
## 4 1019 86 92 91 49 86 75 94
## 5 132 56 69 65 43 66 64 68
## 6 48 47 66 64 40 58 57 71

head(c_test_relevant_df)

##          name year salary Pos Age Tm GS MP PER TS% TRB% AST% STL% BLK%
## 7 al farouq aminu 2016 7680965 SF 25 POR 82 2341 12.7 0.533 11.5 8.8 1.5
## 13 alan anderson 2016 1315448 SG 33 WAS 0 192 9.2 0.495 7.9 10.3 1.0
## 24 alonzo gee 2016 1400000 SF 28 NOP 38 1632 8.4 0.572 8.3 6.2 1.9
## 29 andre drummond 2016 22116750 C 22 DET 81 2666 21.2 0.499 24.5 4.4 2.3
## 32 andre iguodala 2017 14814815 SF 33 GSW 0 1998 14.4 0.624 8.3 16.7 1.8
## 45 anthony davis 2016 22116750 C 22 NOP 61 2164 25.0 0.559 16.1 10.0 1.8
##   BLK% TOV% USG% WS BPM VORP FG FG% 3P 3P% 2P 2P% eFG% FT FT%
## 7 1.8 13.2 16.9 4.0 0.2 1.3 299 0.416 126 0.361 173 0.468 0.503 115 0.737
## 13 0.4 3.0 15.4 0.3 -2.8 0.0 21 0.356 12 0.324 9 0.409 0.458 11 0.733
## 24 0.7 16.2 9.2 1.9 -0.8 0.5 132 0.518 17 0.283 115 0.590 0.551 44 0.667
## 29 3.2 10.5 24.1 7.4 -0.5 1.0 552 0.521 2 0.333 550 0.522 0.522 208 0.355
## 32 1.5 11.2 11.2 6.9 3.0 2.5 219 0.528 64 0.362 155 0.651 0.605 72 0.706
## 45 4.7 8.4 29.6 7.2 2.2 2.3 560 0.493 35 0.324 525 0.511 0.508 326 0.758
##   AST PF PTS out ovr ins pla ath def reb
## 7 138 171 839 90 91 77 60 81 76 94
## 13 14 25 65 73 69 57 49 71 60 42
## 24 72 169 325 65 70 61 50 76 63 46
## 29 67 245 1314 47 86 81 42 77 73 97
## 32 262 97 574 81 91 79 89 87 90 45
## 45 116 148 1481 91 97 89 64 87 87 98

write.csv(c_train_relevant_df,'../data/train_test/complete/train_relevant.csv',row.names=F)
write.csv(c_test_relevant_df,'../data/train_test/complete/test_relevant.csv',row.names=F)

```

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)
p_x_vars <- names(p_train_relevant_df)[!(names(p_train_relevant_df))%in%c('salary','name','2P','PTS')]
# 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 + GS + MP + PER + `TS%` + `TRB%` +
## `AST%` + `STL%` + `BLK%` + `TOV%` + `USG%` + WS + BPM + VORP +
## FG + `FG%` + `3P` + `3P%` + `2P%` + `eFG%` + FT + `FT%` +
## AST + PF
## <environment: 0x564994bedea8>

p_selection_model <- ols(p_formula, data = p_train_relevant_df)
p_selection_model

## Linear Regression Model
##
## ols(formula = p_formula, data = p_train_relevant_df)
##
##          Model Likelihood      Discrimination
##                  Ratio Test           Indexes
## Obs            588    LR chi2   611.37    R2       0.646
## sigma4596847.8837 d.f.          59    R2 adj   0.607
## d.f.            528    Pr(> chi2) 0.0000    g  6402520.583
##
## Residuals
##
##        Min       1Q     Median      3Q      Max
## -13700680 -2839583 -118251  2432935 16150653
##
##
##          Coef        S.E.         t     Pr(>|t|)
## Intercept 1553909.3793 5487540.3109  0.28 0.7772
## year=2017 1396393.5240 410654.1299  3.40 0.0007
## Pos=PF   -202092.5519 747694.9011 -0.27 0.7870
## Pos=PG   -3292889.8860 1274714.2015 -2.58 0.0101
## Pos=SF   -787449.7161 996472.5655 -0.79 0.4297
## Pos=SG   -1667019.3745 1109742.2133 -1.50 0.1337
## Age        228455.3982 52049.9531  4.39 <0.0001
## Tm=BOS  -2785931.6466 1728002.0114 -1.61 0.1075
## Tm=BRK  -3486926.0974 1861094.6892 -1.87 0.0615
## Tm=CHI  -3432180.3405 1701529.4924 -2.02 0.0442
## Tm=CHO  -2296047.1673 1755368.1972 -1.31 0.1914
## Tm=CLE  -269514.0025 1773216.8239 -0.15 0.8793
## Tm=DAL  -3544609.1832 1823155.7160 -1.94 0.0524
## Tm=DEN  -5724530.8187 1803188.8342 -3.17 0.0016
## Tm=DET  -3034455.8277 1759690.2742 -1.72 0.0852
## Tm=GSW  -3038617.9359 1723530.0875 -1.76 0.0785
## Tm=HOU  -4310794.4377 1787829.5830 -2.41 0.0162
## Tm=IND  -2861199.3161 1773768.9052 -1.61 0.1073
## Tm=LAC  -1112210.6476 1714078.2922 -0.65 0.5167
## Tm=LAL  -2619114.2524 1901871.7607 -1.38 0.1691
## Tm=MEM  -1198674.8873 1713918.4356 -0.70 0.4846
## Tm=MIA  -2265976.3922 1824684.7594 -1.24 0.2148
## Tm=MIL  -1506577.8752 1830047.4931 -0.82 0.4107
## Tm=MIN  -4413367.8770 1758662.1429 -2.51 0.0124
## Tm=NOP  -543968.6628 1809940.5978 -0.30 0.7639
## Tm=NYK  -3869428.0379 1739634.9390 -2.22 0.0266
## Tm=OKC  -1536298.1613 1773700.4062 -0.87 0.3868
## Tm=ORL  -2344680.9015 1764761.8299 -1.33 0.1846
## Tm=PHI  -4180411.0786 1834550.6960 -2.28 0.0231
## Tm=PHO  -1962454.0170 1729161.0671 -1.13 0.2569
```

```

## Tm=POR      -901135.7685 1786072.3010 -0.50 0.6141
## Tm=SAC     -2519878.2277 1820260.8287 -1.38 0.1668
## Tm=SAS     -4502772.1861 1765011.7078 -2.55 0.0110
## Tm=TOR     -1706921.1054 1664347.5002 -1.03 0.3056
## Tm=TOT     -4009263.3457 1517228.0922 -2.64 0.0085
## Tm=UTA     -2913684.6345 1710244.0288 -1.70 0.0890
## Tm=WAS     -824536.3771 1833293.5143 -0.45 0.6531
## GS          39308.3338   11806.9938  3.33 0.0009
## MP          2064.9760    1229.7283  1.68 0.0937
## PER         126764.2182   244045.7934  0.52 0.6037
## TS%        -24140528.8730 19763737.5697 -1.22 0.2225
## TRB%       -29773.9229   127930.1738 -0.23 0.8161
## AST%        68316.3597   96508.0542  0.71 0.4793
## STL%       -360976.0005   376028.1501 -0.96 0.3375
## BLK%       -265298.4560   224013.2607 -1.18 0.2368
## TOV%        143716.7979   95378.5587  1.51 0.1325
## USG%       159027.7466   127640.1549  1.25 0.2134
## WS          1302637.7917   335318.9879  3.88 0.0001
## BPM         199060.6721   230806.7374  0.86 0.3888
## VORP       -805540.9322   546414.6389 -1.47 0.1410
## FG          -4424.3274    6111.8509 -0.72 0.4695
## FG%        18670714.2339  12580853.5108  1.48 0.1384
## 3P          15955.3209    8166.2767  1.95 0.0513
## 3P%        1279080.8965  2175294.4764  0.59 0.5568
## 2P%       -4892750.5433  5316650.5216 -0.92 0.3579
## eFG%       -474083.8835  17197941.6970 -0.03 0.9780
## FT          6098.7192    5459.5719  1.12 0.2645
## FT%        -823725.5012  2683898.3634 -0.31 0.7590
## AST         -1619.3377    5027.6268 -0.32 0.7475
## PF          -27671.9031   7966.2727 -3.47 0.0006
##

```

```

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
## eFG%    0.00    1  0.9780  0.00      1  0.9780 -2.00 0.646
## TRB%    0.05    1  0.8156  0.06      2  0.9728 -3.94 0.646
## FT%     0.10    1  0.7513  0.16      3  0.9844 -5.84 0.646
## AST     0.10    1  0.7495  0.26      4  0.9924 -7.74 0.646
## PER     0.20    1  0.6539  0.46      5  0.9936 -9.54 0.646
## 3P%     0.41    1  0.5241  0.86      6  0.9902 -11.14 0.646
## FG      0.52    1  0.4716  1.38      7  0.9861 -12.62 0.646
## STL%    0.89    1  0.3466  2.27      8  0.9717 -13.73 0.645
## BLK%    0.72    1  0.3967  2.99      9  0.9648 -15.01 0.644
## BPM     1.31    1  0.2532  4.29     10  0.9332 -15.71 0.644
## TOV%    1.27    1  0.2590  5.57     11  0.9007 -16.43 0.643
## 2P%     1.59    1  0.2073  7.16     12  0.8471 -16.84 0.642
## MP      2.10    1  0.1469  9.26     13  0.7530 -16.74 0.640
## USG%    1.89    1  0.1697 11.15     14  0.6746 -16.85 0.639
##

```

```

## Approximate Estimates after Deleting Factors
##

```

```

##           Coef      S.E.      Wald Z      P
## Intercept 456240 2758634 0.165386 8.686e-01
## year=2017 1148156 388402  2.956106 3.115e-03
## Pos=PF    -11982 655951 -0.018267 9.854e-01
## Pos=PG   -3909901 1005868 -3.887090 1.015e-04
## Pos=SF   -822935 748241 -1.099826 2.714e-01
## Pos=SG   -1676955 835061 -2.008182 4.462e-02
## Age      223020 49448  4.510194 6.477e-06
## Tm=BOS  -2740006 1714345 -1.598281 1.100e-01
## Tm=BRK  -3114659 1832586 -1.699598 8.921e-02

```

```

## Tm=CHI      -2969940 1676308 -1.771715 7.644e-02
## Tm=CHO      -2055523 1730507 -1.187816 2.349e-01
## Tm=CLE      10815 1752056  0.006173 9.951e-01
## Tm=DAL      -3271557 1805859 -1.811635 7.004e-02
## Tm=DEN      -5341317 1777449 -3.005047 2.655e-03
## Tm=DET      -2467550 1711062 -1.442116 1.493e-01
## Tm=GSW      -2802620 1675123 -1.673083 9.431e-02
## Tm=HOU      -4610332 1756599 -2.624579 8.676e-03
## Tm=IND      -2574200 1749950 -1.471014 1.413e-01
## Tm=LAC      -908521 1698406 -0.534925 5.927e-01
## Tm=LAL      -2580728 1845602 -1.398312 1.620e-01
## Tm=MEM      -904715 1684780 -0.536993 5.913e-01
## Tm=MIA      -1866583 1799064 -1.037530 2.995e-01
## Tm=MIL      -1400747 1806075 -0.775575 4.380e-01
## Tm=MIN      -4619876 1719838 -2.686228 7.226e-03
## Tm=NOP      -539496 1775019 -0.303938 7.612e-01
## Tm=NYK      -3342984 1707081 -1.958305 5.019e-02
## Tm=OKC      -1310920 1721128 -0.761664 4.463e-01
## Tm=ORL      -2180258 1736059 -1.255867 2.092e-01
## Tm=PHI      -3635593 1799022 -2.020872 4.329e-02
## Tm=PHO      -1899465 1688962 -1.124635 2.607e-01
## Tm=POR      -629952 1749315 -0.360114 7.188e-01
## Tm=SAC      -2618708 1804384 -1.451302 1.467e-01
## Tm=SAS      -3666757 1711710 -2.142160 3.218e-02
## Tm=TOR      -1246616 1637512 -0.761286 4.465e-01
## Tm=TOT      -3752179 1493017 -2.513153 1.197e-02
## Tm=UTA      -2085085 1669685 -1.248789 2.117e-01
## Tm=WAS      -706936 1816507 -0.389173 6.971e-01
## GS          47059   10642  4.422134 9.773e-06
## TS%        -20082937 8093218 -2.481453 1.308e-02
## AST%       176331   39929  4.416133 1.005e-05
## WS          1422260  243711  5.835842 5.352e-09
## VORP       -1079988  374926 -2.880538 3.970e-03
## FG%        17131465 7566993  2.263973 2.358e-02
## 3P          22732    6296  3.610792 3.053e-04
## FT          11611    3549  3.271312 1.070e-03
## PF          -19874    4847 -4.100289 4.126e-05
##
## Factors in Final Model
##
## [1] year Pos  Age  Tm   GS   TS%  AST%  WS   VORP  FG%  3P   FT   PF

```

Checking for Multicollinearity Among Optimal Subset of Primary Variables.

```

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

## salary ~ year + Pos + Age + Tm + GS + `TS%` + `AST%` + WS + VORP +
##     `FG%` + `3P` + FT + PF
## <environment: 0x56499468f150>

p_subset_lm <- lm(p_subset_formula , data=p_train_relevant_df)
summary(p_subset_lm)

##
## Call:
## lm(formula = p_subset_formula, data = p_train_relevant_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -14236239 -2846907 -194149  2554876  15549671 
##
## Coefficients:

```

```

##               Estimate Std. Error t value Pr(>|t|) 
## (Intercept) 456240    2751361   0.166 0.868358
## year2017    1148156    387378   2.964 0.003171 ** 
## PosPF       -11982     654222  -0.018 0.985394
## PosPG      -3909901    1003216  -3.897 0.000109 *** 
## PosSF      -822935     746268  -1.103 0.270632
## PosSG      -1676955    832860  -2.013 0.044557 *  
## Age          223020     49318   4.522 7.53e-06 *** 
## TmBOS      -2740006    1709825  -1.603 0.109626
## TmBRK      -3114659    1827754  -1.704 0.088937 .  
## TmCHI      -2969940    1671888  -1.776 0.076228 .  
## TmCHO      -2055523    1725944  -1.191 0.234192
## TmCLE        10815     1747437   0.006 0.995064
## TmDAL      -3271557    1801097  -1.816 0.069858 .
## TmDEN      -5341317    1772762  -3.013 0.002708 ** 
## TmDET      -2467550    1706550  -1.446 0.148775
## TmGSW      -2802620    1670706  -1.678 0.094020 .
## TmHOU      -4610332    1751967  -2.632 0.008743 ** 
## TmIND      -2574200    1745336  -1.475 0.140819
## TmLAC        908520    1693928  -0.536 0.591944
## TmLAL      -2580728    1840736  -1.402 0.161485
## TmMEM      -904715     1680338  -0.538 0.590513
## TmMIA      -1866583    1794321  -1.040 0.298677
## TmMIL      -1400746    1801313  -0.778 0.437129
## TmMIN      -4619876    1715303  -2.693 0.007293 ** 
## TmNOP      -539496     1770339  -0.305 0.760680
## TmNYK      -3342984    1702580  -1.963 0.050101 .
## TmOKC      -1310920    1716590  -0.764 0.445392
## TmORL      -2180258    1731481  -1.259 0.208505
## TmPHI      -3635593    1794279  -2.026 0.043232 *  
## TmPHO      -1899465    1684509  -1.128 0.259985
## TmPOR      -629952     1744703  -0.361 0.718191
## TmSAC      -2618708    1799627  -1.455 0.146210
## TmSAS      -3666757    1707197  -2.148 0.032170 *  
## TmTOR      -1246616    1633195  -0.763 0.445617
## TmTOT      -3752179    1489080  -2.520 0.012029 *  
## TmUTA      -2085085    1665283  -1.252 0.211077
## TmWAS      -706936     1811718  -0.390 0.696541
## GS          47059      10614   4.434 1.12e-05 *** 
## `TS%`     -20082937    8071879  -2.488 0.013145 *
## `AST%`      176331     39824   4.428 1.15e-05 *** 
## WS          1422260    243069   5.851 8.45e-09 *** 
## VORP      -1079988    373937  -2.888 0.004030 ** 
## `FG%`      17131465    7547042  2.270 0.023602 *  
## `3P`         22732      6279   3.620 0.000322 *** 
## FT          11612      3540   3.280 0.001105 ** 
## PF          -19874     4834  -4.111 4.55e-05 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 4585000 on 542 degrees of freedom
## Multiple R-squared:  0.639, Adjusted R-squared:  0.609
## F-statistic: 21.32 on 45 and 542 DF, p-value: < 2.2e-16

```

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

	WS	`FG%`	VORP	`TS%`	TmTOT	PosPG	FT	`AST%`
## 14.914773	9.299794	8.728394	7.144946	5.086995	4.190961	4.167082	3.650485	
## `3P`	PosSG	TmTOR	GS	TmPHO	TmGSW	TmUTA	TmMEM	
## 3.456991	3.092757	3.037526	2.863495	2.858784	2.812127	2.793899	2.720141	
## TmCHI	TmSAS	TmLAC	TmOKC	TmMIN	TmBOS	TmDET	TmNYK	
## 2.692851	2.678805	2.637327	2.577475	2.573613	2.557200	2.547414	2.535575	
## TmPOR	PF	PosSF	TmORL	TmCHO	TmHOU	TmCLE	TmIND	
## 2.526887	2.499607	2.498992	2.488734	2.472843	2.410641	2.398189	2.392428	

```

##      TmPHI      TmDEN      TmNOP      TmMIL      TmDAL      TmSAC      TmMIA      TmLAL
## 2.383919  2.327086  2.320728  2.256420  2.255879  2.252197  2.238935  2.203021
##      TmBRK      TmWAS      PosPF       Age   year2017
## 2.172056  2.134109  1.908324  1.280955  1.049395

```

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

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

```
library(rms)
```

```
c_x_vars <- names(c_train_relevant_df) [!(names(c_train_relevant_df) %in% c('salary', 'name', '2P', 'PTS'))]
# 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 + GS + MP + PER + `TS%` + `TRB%` +
##        `AST%` + `STL%` + `BLK%` + `TOV%` + `USG%` + WS + BPM + VORP +
##        FG + `FG%` + `3P` + `3P%` + `2P%` + `eFG%` + FT + `FT%` +
##        AST + PF + out + ovr + ins + pla + ath + def + reb
## <environment: 0x5649982dfd88>
```

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

```
## Linear Regression Model
##
## ols(formula = c_formula, data = c_train_relevant_df)
##
##                               Model Likelihood      Discrimination
##                               Ratio Test          Indexes
## Obs            588    LR chi2     628.30      R2      0.656
## sigma          4561484.3734 d.f.         66      R2 adj     0.613
## d.f.           521    Pr(> chi2) 0.0000      g  6461689.853
##
## Residuals
```

```
##             Min       1Q     Median       3Q      Max
## -14615864 -2715524    -58045    2320543  15483115
```

```
##
##               Coef        S.E.        t    Pr(>|t|)
## Intercept -4450716.0365 5998793.1509 -0.74 0.4585
## year=2017    760705.4915 482461.2858  1.58 0.1155
## Pos=PF      74539.8571 803155.8480  0.09 0.9261
## Pos=PG     -3836878.9447 1555010.3953 -2.47 0.0139
## Pos=SF     -1258327.9963 1109153.8535 -1.13 0.2571
## Pos=SG     -2229161.2986 1315989.1726 -1.69 0.0909
## Age        183833.9201 56612.2582  3.25 0.0012
## Tm=BOS     -2628991.6776 1720351.8481 -1.53 0.1271
## Tm=BRK     -2736554.4384 1862720.9863 -1.47 0.1424
## Tm=CHI     -2995521.2555 1697962.6680 -1.76 0.0783
## Tm=CHO     -1761601.5566 1751683.2846 -1.01 0.3150
## Tm=CLE     -301085.0774 1768353.6568 -0.17 0.8649
## Tm=DAL     -2936783.5564 1820380.4139 -1.61 0.1073
## Tm=DEN     -5271247.2217 1796587.4649 -2.93 0.0035
## Tm=DET     -2311494.8222 1766632.9505 -1.31 0.1913
## Tm=GSW     -2929811.2390 1717307.2321 -1.71 0.0886
## Tm=HOU     -4103929.8613 1780682.0268 -2.30 0.0216
## Tm=IND     -2489194.9946 1770797.7073 -1.41 0.1604
## Tm=LAC     -1202650.9418 1709070.2142 -0.70 0.4819
## Tm=LAL     -2177721.0384 1905133.9114 -1.14 0.2535
## Tm=MEM     -1037106.6453 1720323.5834 -0.60 0.5469
## Tm=MIA     -2126510.1423 1820100.3235 -1.17 0.2432
```

```

## Tm=MIL      -1214985.2485 1828326.4484 -0.66 0.5066
## Tm=MIN      -4001352.1516 1758474.2533 -2.28 0.0233
## Tm=NOP       21140.4371 1811950.0683  0.01 0.9907
## Tm=NYK      -3336784.6326 1736758.5335 -1.92 0.0552
## Tm=OKC      -1352270.0393 1767269.8076 -0.77 0.4445
## Tm=ORL      -1799749.2956 1760933.0656 -1.02 0.3072
## Tm=PHI      -3793899.7880 1826979.5079 -2.08 0.0383
## Tm=PHO      -1577555.2151 1722415.0409 -0.92 0.3601
## Tm=POR       -640419.9138 1786764.8593 -0.36 0.7202
## Tm=SAC      -2169962.3157 1814299.0247 -1.20 0.2322
## Tm=SAS      -4275397.2442 1772555.1440 -2.41 0.0162
## Tm=TOR      -1719009.1870 1654753.7609 -1.04 0.2994
## Tm=TOT      -3792944.0219 1512534.5163 -2.51 0.0125
## Tm=UTA      -2664225.0325 1712794.5186 -1.56 0.1204
## Tm=WAS      -374397.8758 1833690.3460 -0.20 0.8383
## GS          31816.7256 12027.0596  2.65 0.0084
## MP          1629.4118 1229.3833  1.33 0.1856
## PER         71614.5033 245458.5622  0.29 0.7706
## TS%        -17583139.1749 19803751.3763 -0.89 0.3750
## TRB%       -26736.1539 135755.2647 -0.20 0.8439
## AST%        48530.1029 98264.9450  0.49 0.6216
## STL%       -431443.8808 374067.2963 -1.15 0.2493
## BLK%       -352613.5122 227975.4926 -1.55 0.1225
## TOV%        128179.6970 95344.5002  1.34 0.1794
## USG%       182409.1965 130260.4636  1.40 0.1620
## WS          1321244.6707 340113.5798  3.88 0.0001
## BPM         260554.8234 232173.0172  1.12 0.2623
## VORP       -868072.4082 561452.2193 -1.55 0.1227
## FG          -3206.8276 6231.9083 -0.51 0.6071
## FG%        8165411.8125 13699749.6412  0.60 0.5514
## 3P          13659.9759 8398.4451  1.63 0.1045
## 3P%        2152269.0839 2204548.1926  0.98 0.3294
## 2P%       -2560480.5425 5425189.0482 -0.47 0.6372
## eFG%       926144.8007 17351129.5191  0.05 0.9575
## FT          3901.0355 5466.8639  0.71 0.4758
## FT%        1017445.9093 2729863.9001  0.37 0.7095
## AST         -522.4391 5047.8340 -0.10 0.9176
## PF          -26497.9175 7919.0612 -3.35 0.0009
## out         -83202.4401 39290.7595 -2.12 0.0347
## ovr         154156.9747 88178.6903  1.75 0.0810
## ins         -6198.2512 47693.1011 -0.13 0.8966
## pla         -784.7456 31215.2598 -0.03 0.9800
## ath         12953.2418 50208.8708  0.26 0.7965
## def         41452.8120 38423.5914  1.08 0.2812
## reb        -40896.7907 26705.1724 -1.53 0.1263
##

```

```

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  0.9799  0.00      1  0.9799 -2.00 0.656
## eFG%     0.00   1  0.9573  0.00      2  0.9983 -4.00 0.656
## AST      0.01   1  0.9182  0.01      3  0.9996 -5.99 0.656
## ins      0.02   1  0.9000  0.03      4  0.9999 -7.97 0.656
## TRB%     0.04   1  0.8493  0.07      5  0.9999 -9.93 0.656
## PER      0.05   1  0.8302  0.11      6  1.0000 -11.89 0.656
## ath      0.05   1  0.8170  0.17      7  1.0000 -13.83 0.656
## FT%      0.16   1  0.6873  0.33      8  1.0000 -15.67 0.656
## FG       0.17   1  0.6808  0.50      9  1.0000 -17.50 0.656
## 2P%     0.20   1  0.6520  0.70     10  1.0000 -19.30 0.656
## FG%      0.07   1  0.7869  0.77     11  1.0000 -21.23 0.656
## FT       0.54   1  0.4619  1.31     12  0.9999 -22.69 0.656

```

```

##  AST%    0.90    1    0.3429   2.21    13    0.9996 -23.79  0.655
##  3P%     1.11    1    0.2920   3.32    14    0.9983 -24.68  0.654
##  def     1.23    1    0.2676   4.55    15    0.9953 -25.45  0.653
##  year    2.25    1    0.1334   6.81    16    0.9768 -25.19  0.652
##  VORP    2.06    1    0.1513   8.86    17    0.9444 -25.14  0.651
##  STL%    2.43    1    0.1194  11.29    18    0.8816 -24.71  0.649
##  TS%     1.82    1    0.1767  13.11    19    0.8326 -24.89  0.648
##  BPM     1.21    1    0.2715  14.32    20    0.8137 -25.68  0.647
##  BLK%    1.51    1    0.2195  15.83    21    0.7790 -26.17  0.646
##  TOV%    3.20    1    0.0737  19.03    22    0.6435 -24.97  0.644
##  Pos     7.05    4    0.1331  26.08    26    0.4585 -25.92  0.639
##  reb     0.17    1    0.6817  26.25    27    0.5047 -27.75  0.639
##  Tm      40.84   30   0.0895  67.09    57    0.1694 -46.91  0.612
##  3P      3.79    1    0.0516  70.88    58    0.1193 -45.12  0.610
##
## Approximate Estimates after Deleting Factors
##
##          Coef      S.E.  Wald Z      P
## Intercept -15775543 2146247.3 -7.350 1.977e-13
## Age        238362  46224.8  5.157 2.515e-07
## GS         29122   11158.4  2.610 9.057e-03
## MP         2578    668.5   3.857 1.150e-04
## USG%       206060  45994.2  4.480 7.460e-06
## WS         985550  109877.2  8.970 0.000e+00
## PF        -27716   6030.8  -4.596 4.313e-06
## out        -78216   19627.0 -3.985 6.745e-05
## ovr        184278   33719.4  5.465 4.628e-08
##
## Factors in Final Model
##
## [1] Age  GS  MP  USG% WS  PF  out  ovr

```

Checking for Multicollinearity Among Optimal Subset of Complete Variables.

```

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

## salary ~ Age + GS + MP + `USG%` + WS + PF + out + ovr
## <environment: 0x5649930af430>

c_subset_lm <- lm(c_subset_formula , data=c_train_relevant_df)
summary(c_subset_lm)

##
## Call:
## lm(formula = c_subset_formula, data = c_train_relevant_df)
##
## Residuals:
##      Min       1Q       Median      3Q      Max 
## -15368141 -2923935 -261742   2216792  17647516 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -1.578e+07  2.170e+06 -7.270 1.17e-12 ***
## Age          2.384e+05  4.674e+04  5.100 4.61e-07 ***
## GS           2.912e+04  1.128e+04  2.581 0.010087 *  
## MP           2.578e+03  6.759e+02  3.814 0.000151 *** 
## `USG%`       2.061e+05  4.650e+04  4.431 1.12e-05 *** 
## WS           9.856e+05  1.111e+05  8.871 < 2e-16 *** 
## PF          -2.772e+04  6.098e+03 -4.545 6.68e-06 *** 
## out          -7.822e+04  1.984e+04 -3.941 9.09e-05 *** 
## ovr          1.843e+05  3.409e+04  5.405 9.47e-08 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4612000 on 579 degrees of freedom
## Multiple R-squared:  0.6098, Adjusted R-squared:  0.6044
## F-statistic: 113.1 on 8 and 579 DF,  p-value: < 2.2e-16
sort(vif(c_subset_lm),decreasing=T) # All variables have low VIF values. So no multicollinearity.

##          MP         PF         GS         WS         ovr         out      `USG%`        Age
## 7.984500 3.929993 3.197309 3.078852 2.832498 2.563631 1.613312 1.136827
c_selected[['names.kept']] <- c_selected[['names.kept']][!(c_selected[['names.kept']] %in% c("FGA"))]
c_subset_formula <- get_salary_formula(c_selected[['names.kept']])
c_subset_formula

## salary ~ Age + GS + MP + `USG%` + WS + PF + out + ovr
## <environment: 0x564990cc50e0>
c_subset_lm <- lm(c_subset_formula , data=c_train_relevant_df)
summary(c_subset_lm)

##
## Call:
## lm(formula = c_subset_formula, data = c_train_relevant_df)
##
## Residuals:
##       Min     1Q     Median     3Q     Max 
## -15368141 -2923935 -261742   2216792  17647516 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -1.578e+07  2.170e+06 -7.270 1.17e-12 ***
## Age          2.384e+05  4.674e+04  5.100 4.61e-07 ***
## GS           2.912e+04  1.128e+04  2.581 0.010087 *  
## MP           2.578e+03  6.759e+02  3.814 0.000151 *** 
## `USG%`       2.061e+05  4.650e+04  4.431 1.12e-05 *** 
## WS           9.856e+05  1.111e+05  8.871 < 2e-16 *** 
## PF           -2.772e+04 6.098e+03 -4.545 6.68e-06 *** 
## out          -7.822e+04 1.984e+04 -3.941 9.09e-05 *** 
## ovr          1.843e+05  3.409e+04  5.405 9.47e-08 *** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4612000 on 579 degrees of freedom
## Multiple R-squared:  0.6098, Adjusted R-squared:  0.6044
## F-statistic: 113.1 on 8 and 579 DF,  p-value: < 2.2e-16
sort(vif(c_subset_lm),decreasing=T)

##          MP         PF         GS         WS         ovr         out      `USG%`        Age
## 7.984500 3.929993 3.197309 3.078852 2.832498 2.563631 1.613312 1.136827
c_vars_final <- c_selected[['names.kept']]

pred <- predict(c_subset_lm, c_test_relevant_df)
SST8 <- sum((c_test_relevant_df$salary - mean(c_test_relevant_df$salary))^2)
SSR8 <- sum((pred - mean(c_test_relevant_df$salary))^2)
Rsqm8<-SSR8/SST8
Rsqm8

## [1] 0.6111508

```

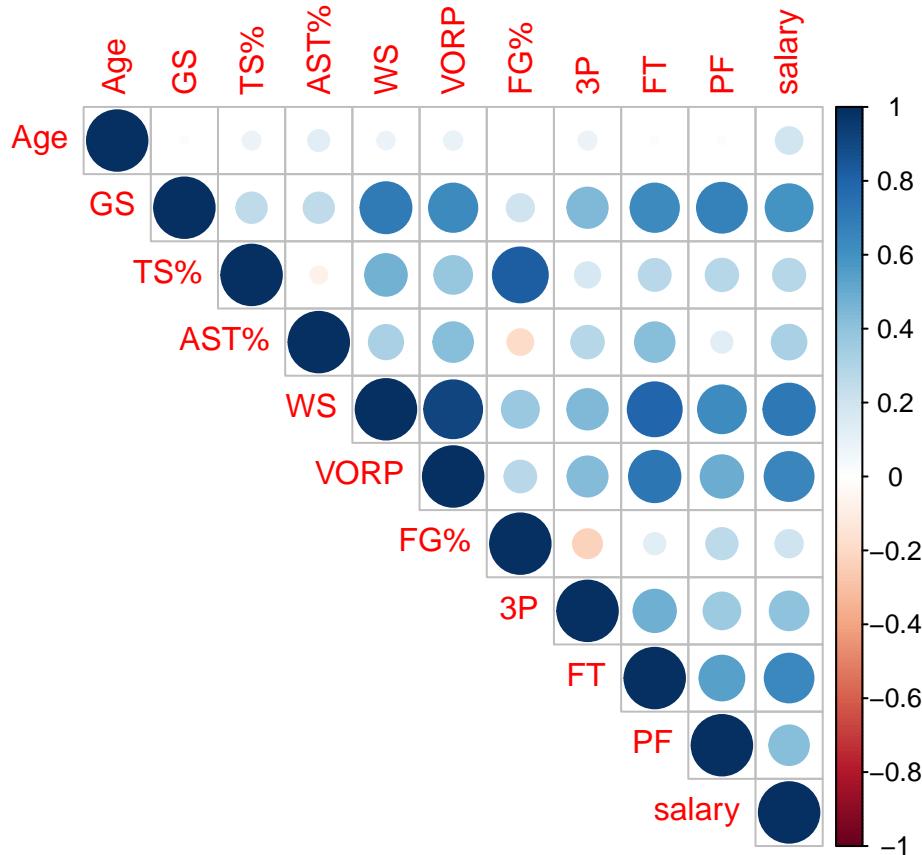
Correlation plots for relevant variables for complete and primary dataset

```

v <- c(p_vars_final,"salary")
corr_matrix_c <- cor(Filter(is.numeric,p_train_df[v]),method = "pearson")

```

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



```
dev.copy(png,"../figures/Correlation_selected_variables_primary.png")
```

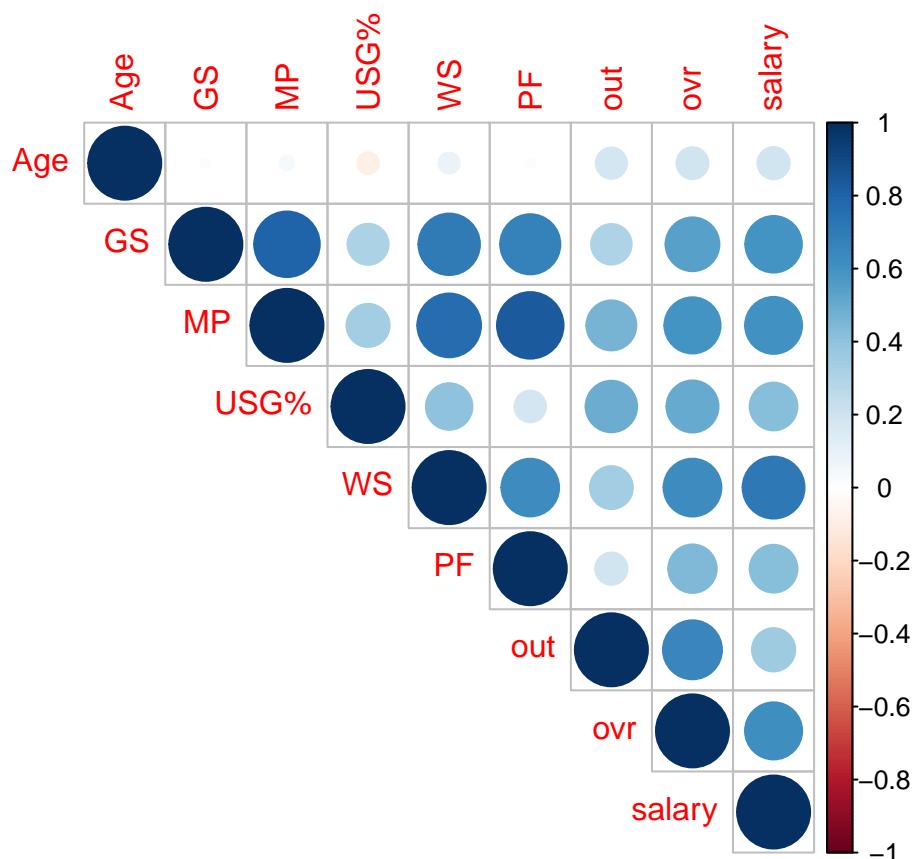
```
## png
## 4
dev.off()

## pdf
## 2

v <- c(c_vars_final,"salary")
corr_matrix_c <- cor(Filter(is.numeric,c_train_df[v]),method = "pearson")
correlation_salary_c <- sort(corr_matrix_c[, 'salary'],decreasing = TRUE)
correlation_salary_c

##      salary         WS        ovr        MP        GS        PF       USG%        out
## 1 1.0000000 0.7191658 0.6118097 0.6094748 0.5994794 0.4286004 0.4278609 0.3512832
##      Age
## 0.1951513

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



```
dev.copy(png, "../figures/Correlation_selected_variables_complete.png")
```

```
## png
## 4
dev.off()

## pdf
## 2
```

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]

train_rows_relevant <- createDataPartition(y=df_p_subset_final[, 'salary'], list=FALSE, p=.8)
p_train_selected_df <- df_p_subset_final[train_rows,]
p_test_selected_df <- df_p_subset_final[-train_rows,]
stopifnot(nrow(p_train_selected_df) + nrow(p_test_selected_df) == nrow(df_p_subset_final))
head(p_train_selected_df)
```

```
##          name salary year Pos Age Tm GS   TS% AST%    WS VORP   FG% 3P  FT
## 1 aaron brooks 2700000 2016 PG  31 CHI  0 0.494 26.0  0.9 -0.4 0.401 66  49
## 2 aaron brooks 2116955 2017 PG  32 IND  0 0.507 20.7  0.3 -0.6 0.403 48  32
## 3 aaron gordon 4351320 2016 PF  20 ORL 37 0.541 10.3  5.4  1.8 0.473 42 129
## 4 aaron gordon 5504420 2017 SF  21 ORL 72 0.530 10.5  3.7  0.8 0.454 77 156
## 5 adreian payne 2022240 2016 PF  24 MIN  2 0.422  8.9 -0.5 -0.5 0.366  9  17
## 6 aj hammons 1312611 2017 C   24 DAL  0 0.472  3.8  0.0 -0.1 0.405  5   9
## PF
## 1 132
## 2 93
## 3 153
## 4 172
```

```

## 5 77
## 6 21
head(p_test_selected_df)

##           name salary year Pos Age Tm GS   TS% AST% WS VORP   FG% 3P
## 7 al farouq aminu 7680965 2016 SF 25 POR 82 0.533 8.8 4.0 1.3 0.416 126
## 13 alan anderson 1315448 2016 SG 33 WAS 0 0.495 10.3 0.3 0.0 0.356 12
## 24 alonzo gee 1400000 2016 SF 28 NOP 38 0.572 6.2 1.9 0.5 0.518 17
## 29 andre drummond 22116750 2016 C 22 DET 81 0.499 4.4 7.4 1.0 0.521 2
## 32 andre iguodala 14814815 2017 SF 33 GSW 0 0.624 16.7 6.9 2.5 0.528 64
## 45 anthony davis 22116750 2016 C 22 NOP 61 0.559 10.0 7.2 2.3 0.493 35
##       FT PF
## 7 115 171
## 13 11 25
## 24 44 169
## 29 208 245
## 32 72 97
## 45 326 148

write.csv(p_train_selected_df,'../data/train_test/primary/train_selected.csv',row.names=F)
write.csv(p_test_selected_df,'../data/train_test/primary/test_selected.csv',row.names=F)

train_rows_relevant <- createDataPartition(y=df_c_subset_final[, 'salary'], list=FALSE, p=.8)
c_train_selected_df <- df_c_subset_final[train_rows,]
c_test_selected_df <- df_c_subset_final[-train_rows,]
stopifnot(nrow(c_train_selected_df) + nrow(c_test_selected_df) == nrow(df_c_subset_final))
head(c_train_selected_df)

##           name salary Age GS   MP USG%   WS  PF out ovr
## 1 aaron brooks 2700000 31 0 1108 22.9 0.9 132 79 75
## 2 aaron brooks 2116955 32 0 894 19.2 0.3 93 87 85
## 3 aaron gordon 4351320 20 37 1863 17.3 5.4 153 87 90
## 4 aaron gordon 5504420 21 72 2298 20.1 3.7 172 86 92
## 5 adreian payne 2022240 24 2 486 17.7 -0.5 77 56 69
## 6 aj hammons 1312611 24 0 163 17.6 0.0 21 47 66

head(c_test_selected_df)

##           name salary Age GS   MP USG%   WS  PF out ovr
## 7 al farouq aminu 7680965 25 82 2341 16.9 4.0 171 90 91
## 13 alan anderson 1315448 33 0 192 15.4 0.3 25 73 69
## 24 alonzo gee 1400000 28 38 1632 9.2 1.9 169 65 70
## 29 andre drummond 22116750 22 81 2666 24.1 7.4 245 47 86
## 32 andre iguodala 14814815 33 0 1998 11.2 6.9 97 81 91
## 45 anthony davis 22116750 22 61 2164 29.6 7.2 148 91 97

write.csv(c_train_selected_df,'../data/train_test/complete/train_selected.csv',row.names=F)
write.csv(c_test_selected_df,'../data/train_test/complete/test_selected.csv',row.names=F)

```