Final Report

Kecheng Liang

12/12/2018

# Introduction

Basketball is one of the most popular sports in the world. National Basketball Association(NBA) is the largest league for this sport. There are lots of interesting data in the game. In basketball, an assist is attributed to a player who passes the ball to a teammate in a way that leads to a score by field goal, meaning that he or she was “assisting” in the basket. Because an assist can be scored for the passer even if the player who receives the pass makes a basket after dribbling the ball. In some situtation it becomes hard to define whether it is a assist. We may think that player who play in the home game are more easily get the tenth assist when the player already have nine assists. Same thing may happen in rebound. I want to do the analysis whether it really happens. Also I will do other interesting graph to show the miracle NBA data.

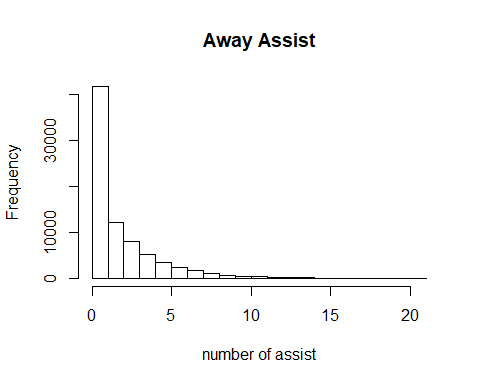
# Data

The data is downloaded from website and the data from 2012 to 2018. It is well organized with 51 variables and I removed some useless variables.

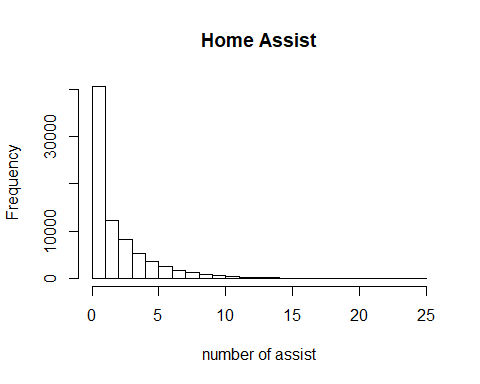
Initial <- read.csv("2012-18\_playerBoxScore.csv")  
Initial$playAST <- as.numeric(Initial$playAST)  
Initial$playTRB <- as.numeric(as.character(Initial$playTRB))  
Initial <- Initial[,c(-1,-2,-3,-4,-5,-8,-11,-12,-13,-14,-15,-16,-17,-24,-48,-49,-50,-51)]  
away\_data <- filter(Initial,Initial$teamLoc=="Away")  
home\_data <- filter(Initial,Initial$teamLoc=="Home")  
nba\_total <- rbind(away\_data,home\_data)

# EDA

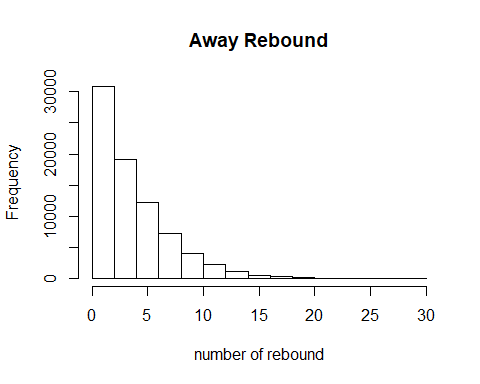
hist(x=away\_data$playAST,main = "Away Assist",xlab = "number of assist")



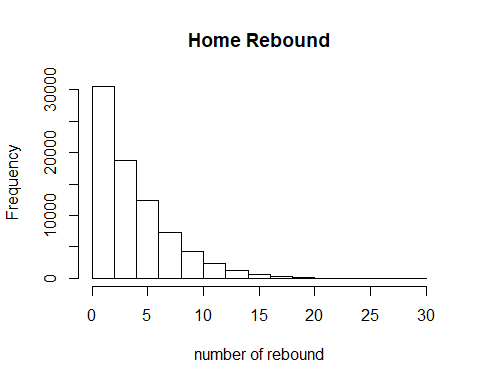
hist(x=home\_data$playAST,main = "Home Assist",xlab = "number of assist")



hist(x=away\_data$playTRB,main = "Away Rebound",xlab = "number of rebound")



hist(x=home\_data$playTRB,main = "Home Rebound",xlab = "number of rebound")



# Chi square test

##   
## Pearson's Chi-squared test  
##   
## data: total\_ast  
## X-squared = 95.549, df = 23, p-value = 8.177e-11

##   
## Pearson's Chi-squared test  
##   
## data: total\_trb  
## X-squared = 64.517, df = 29, p-value = 0.000164