

Midrange

Kevin Chen

December 25, 2018

Webscraping Data

```
library(rvest)
```

```
## Loading required package: xml2
```

```
library(dplyr)
```

```
##
```

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

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(stringr)
```

```
datalist = list()
```

```
#webscraping for years 2007:2019
```

```
for (i in 2007:2019){
```

```
#widgets.sports-reference.com/wg.fcgi?css=1&site=bbr&url=%2Fleagues%2FNBA_2019.html&div=div_team_shooting
```

```
main <- "https://widgets.sports-reference.com/wg.fcgi?css=1&site=bbr&url=%2Fleagues%2FNBA_"
```

```
tail <- ".html&div=div_team_shooting"
```

```
url <- paste0(main, i, tail) #combines correct url according to year
```

```
first <- url %>%
```

```
read_html() %>%
```

```
html_table()
```

```
first <- first[[1]]
```

```
first[2,] <- paste(first[1,], first[2,], sep=" ")
```

```
colnames(first) <- first[2,]
```

```

first <- first[-c(1:2),]
first$i <- i
datalist[[i]] <- first

#names(first) <- as.character(unlist(first[1,]))
#first <- first[-1,]

}
teams_shooting <- bind_rows(datalist) # this is for fga% etc

#scrape offensive rating
datalist2 = list()
for(i in 2007:2019){
  main <- "https://widgets.sports-reference.com/wg.fcgi?css=1&site=bbr&url=%2Fleagues%2FNBA_"
  tail2 <- ".html&div=div_misc_stats"
  url <- paste0(main, i, tail2)

  first <- url %>%
    read_html() %>%
    html_table()

  first <- first[[1]]
  colnames(first) <- first[1,]
  first <- first[-1,]
  first$i <- i
  datalist2[[i]] <- first
}

#Combining data
teams_stats <- bind_rows(datalist2)
names(teams_shooting) <- trimws(names(teams_shooting))

```

Rename teams

```

#bobcats
teams_stats[teams_stats[, "Team"] == "Charlotte Bobcats", "Team"] = "Charlotte Hornets"
teams_shooting[teams_shooting[, "Team"] == "Charlotte Bobcats", "Team"] = "Charlotte Hornets"

#seattle
teams_stats[teams_stats[, "Team"] == "Seattle Supersonics", "Team"] = "Oklahoma City Thunder"
teams_shooting[teams_shooting[, "Team"] == "Seattle Supersonics", "Team"] = "Oklahoma City Thunder"

#new orleans
teams_stats[teams_stats[, "Team"] == "New Orleans Hornets", "Team"] = "New Orleans Pelicans"
teams_shooting[teams_shooting[, "Team"] == "New Orleans Hornets", "Team"] = "New Orleans Pelicans"

```

Select relevant columns for mid range data

```
#selecting relevant columns
teams_midrange<-teams_shooting[,c(1,2,5,9,10,11,12,15,16, 17, 18)]

#create year column
teams_midrange$Year<-teams_shooting%>%select(., i) %>% unlist()

#make columns numeric
teams_midrange[,3:ncol(teams_midrange)]<-sapply(teams_midrange[,3:ncol(teams_midrange)],as.numeric)

#create %FGA by midrange column
teams_midrange$`% of FGA by Midrange` <- rowSums(teams_midrange[,c(4,5,6)])
```

Remove unwanted characters from team stats data

```
teams_stats$Team <- str_remove_all(teams_stats$Team, "\\.*")
```

Arranging team stats each season in alphabetical order

```
end = 30
start = 1
c=0
#1-30 for 2007, the again for 2008 onward
while(c<13){
  teams_stats[start:end,]<-teams_stats[start:end,]%>% arrange(.,Team)
  start = end+2
  end = start+29
  c=c+1
}
```

Add Ortg to Midrange data

```
teams_midrange$ORTg <- as.numeric(teams_stats$ORTg) #adding ORtg from team_stats to team_shooting
teams_midrange<-teams_midrange[,c(1:7,13,8:11, ncol(teams_midrange),12)] #rearrange columns
```

sort according to ortg in desc order

```
end = 30
start = 1
c=0
while(c<13){
```

```

teams_midrange[start:end,]<-teams_midrange[start:end,]%>% arrange(.,-ORtg)
start = end+2
end = start+29
c=c+1
}

```

Lm predicting Ortg from FGA Midrange. Residuals looks good

```

offrtg<-lm(teams_midrange$ORtg~teams_midrange$`% of FGA by Midrange`)
summary(offrtg)

```

```

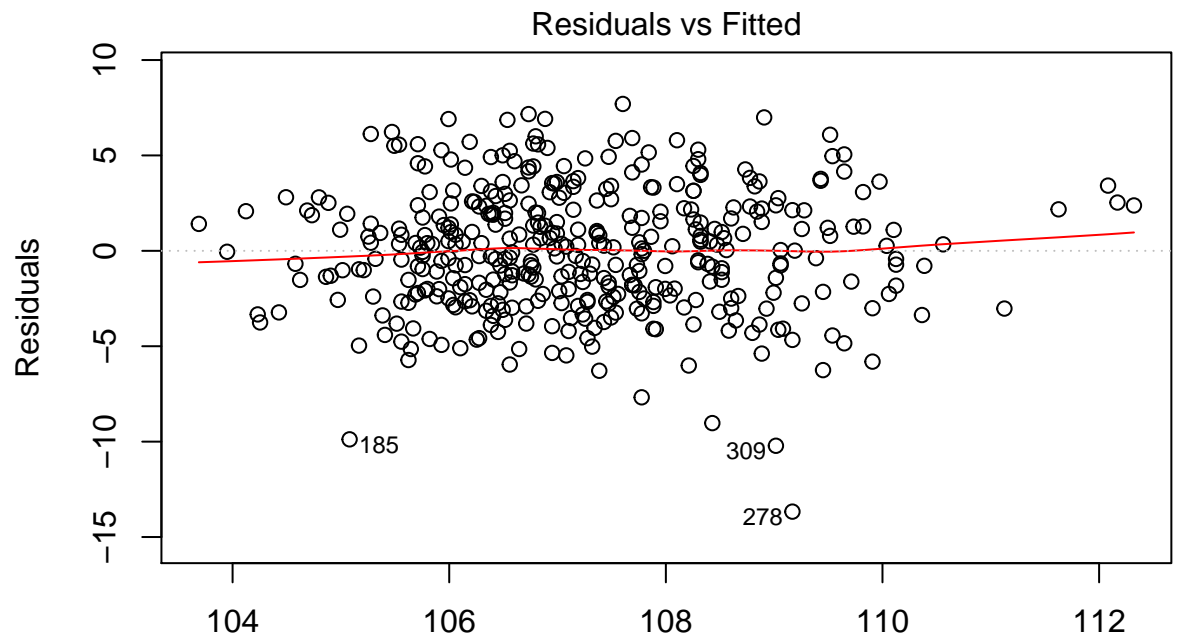
##
## Call:
## lm(formula = teams_midrange$ORtg ~ teams_midrange$`% of FGA by Midrange`)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.6684  -2.2763   0.0403   2.1705   7.6972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)          116.866      1.080 108.242  <2e-16
## teams_midrange$`% of FGA by Midrange`  -21.745      2.409  -9.026  <2e-16
##
## (Intercept)                ***
## teams_midrange$`% of FGA by Midrange` ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.203 on 401 degrees of freedom
## Multiple R-squared:  0.1689, Adjusted R-squared:  0.1668
## F-statistic: 81.48 on 1 and 401 DF,  p-value: < 2.2e-16

```

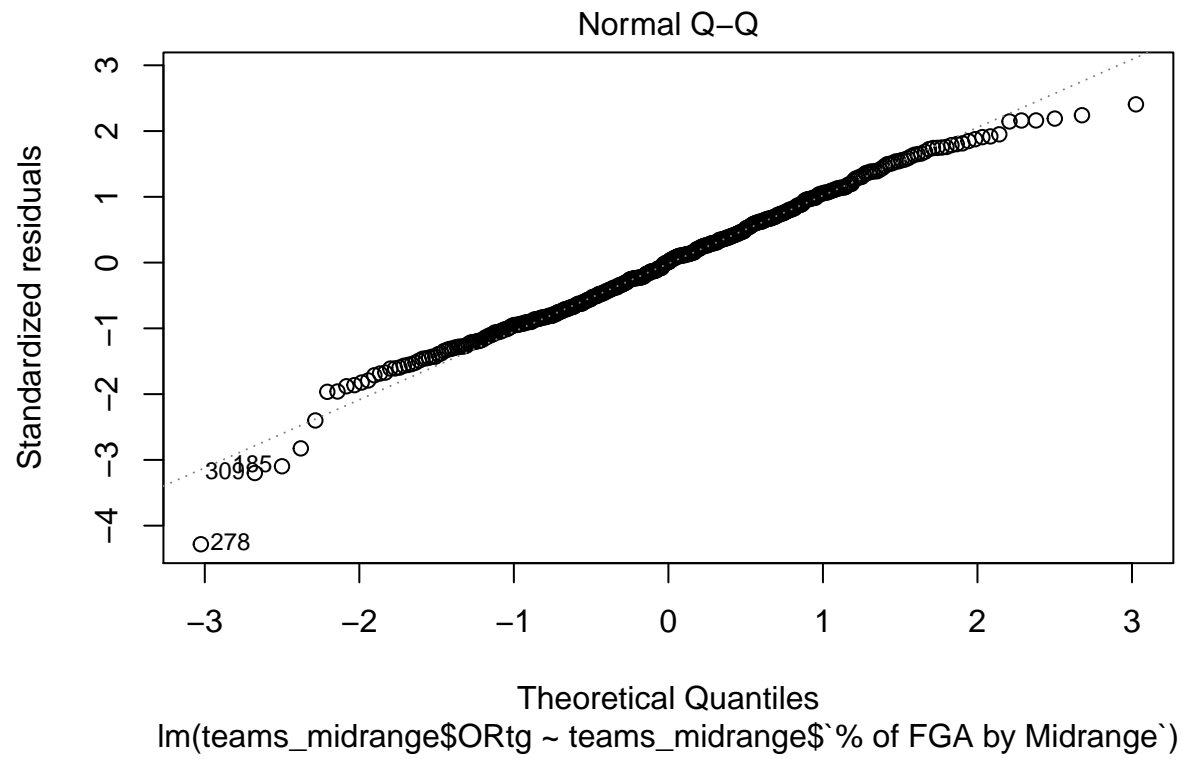
```

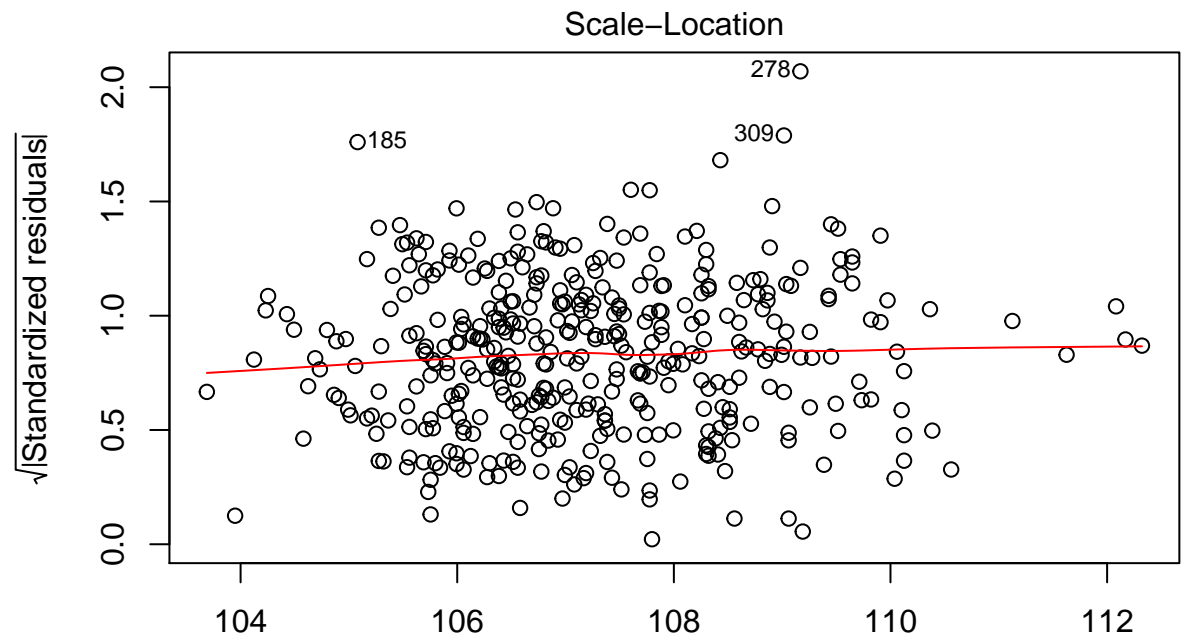
plot(offrtg)

```

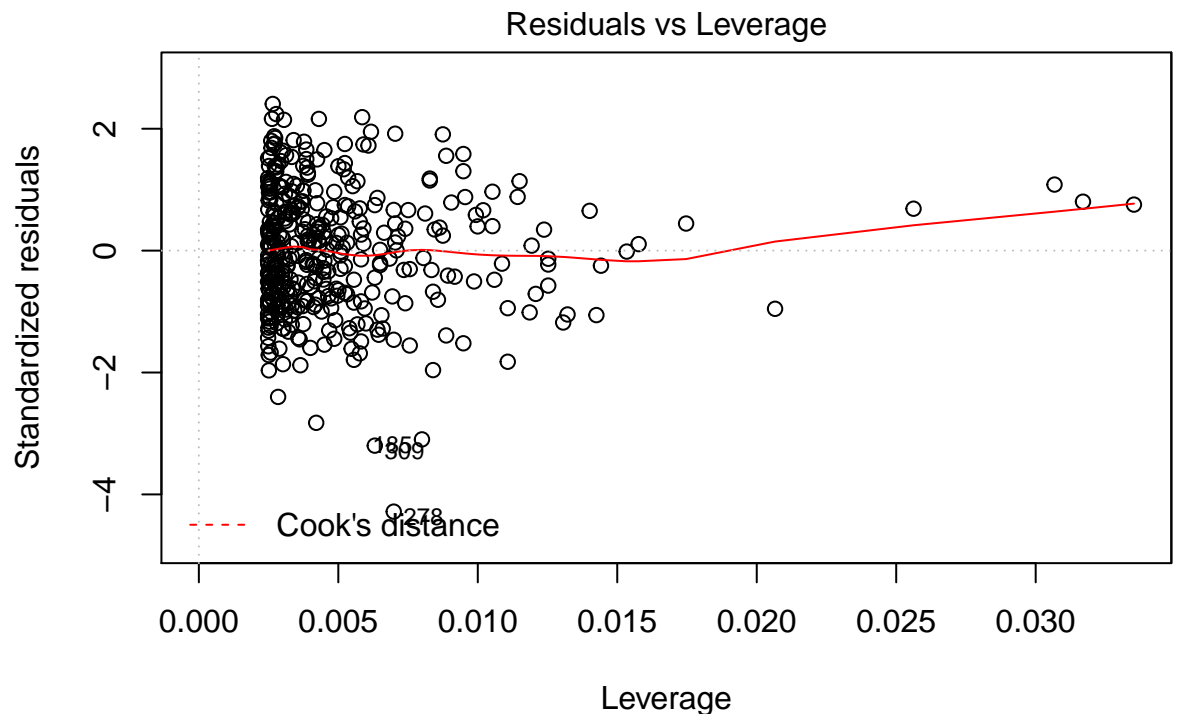


Fitted values
 $\text{lm}(\text{teams_midrange}\$ORTg \sim \text{teams_midrange}\$\text{'\% of FGA by Midrange'})$





Fitted values
`lm(teams_midrange$ORTg ~ teams_midrange$`% of FGA by Midrange`)`

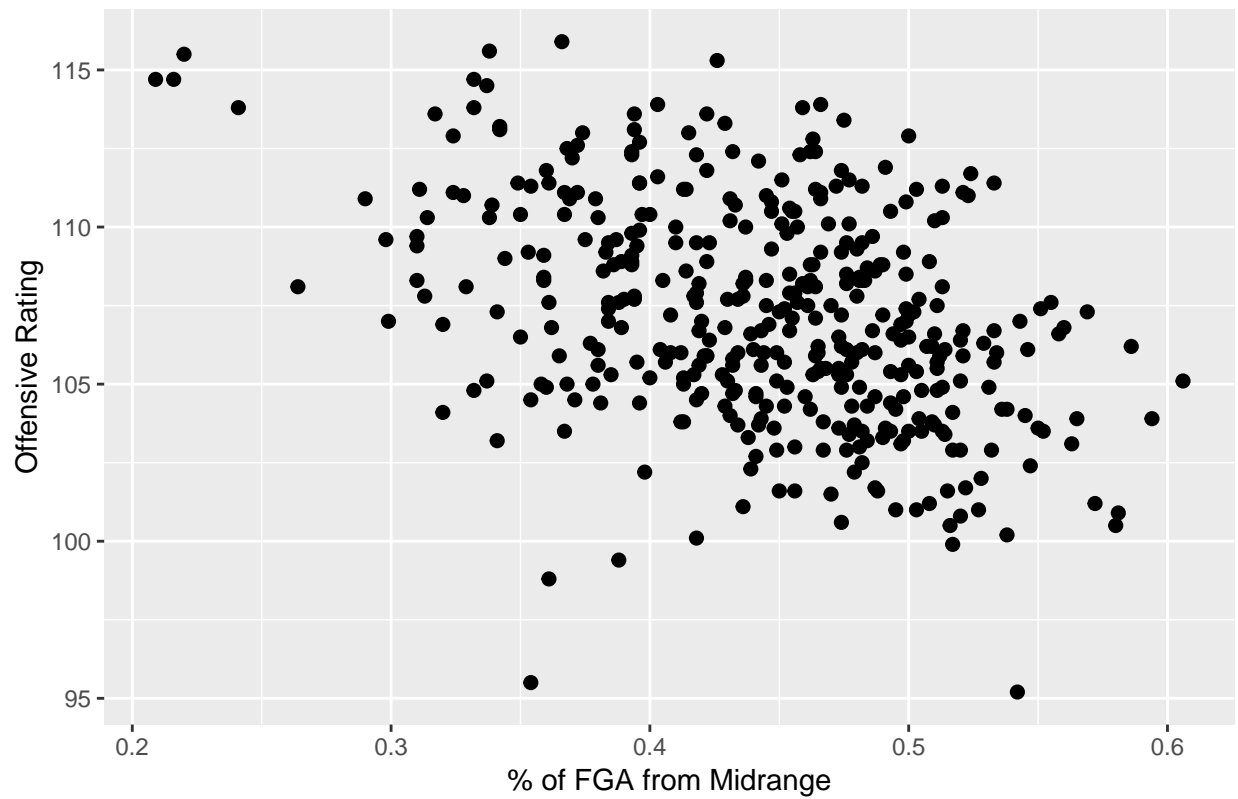


`lm(teams_midrange$ORtg ~ teams_midrange$`% of FGA by Midrange`)`

`##Plot offensive rating and % of FGA from midrange`

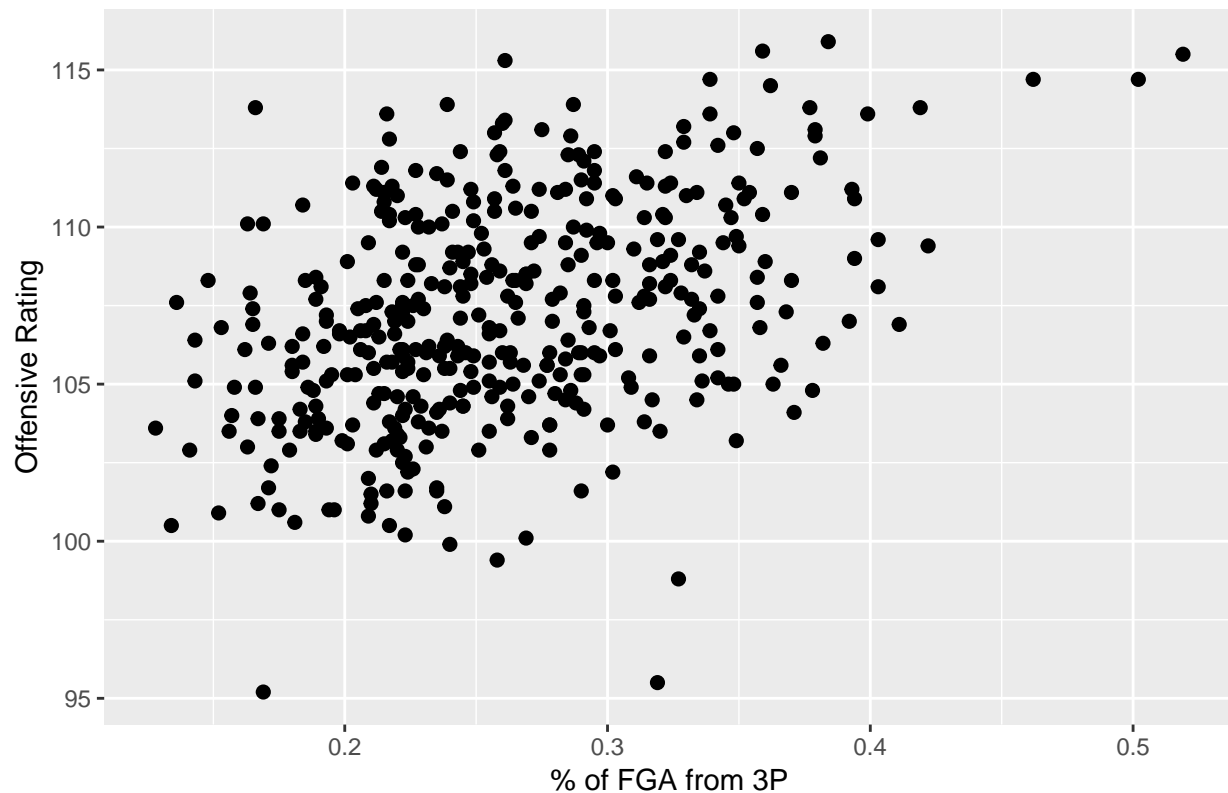
```
library(ggplot2)
ggplot(teams_midrange, aes(x = `% of FGA by Midrange`, y = ORtg)) + geom_point(size = 2) + labs(title =
```


Offensive Rating vs % of FGA from Midrange



```
ggplot(teams_midrange, aes(x = `% of FGA by Distance 3P`, y = ORtg)) + geom_point(size = 2) + labs(tit.
```

Offensive Rating vs % of FGA from 3P



% FGA from 3 by year

```
League_Avg3<-teams_midrange[seq(31,403,31), 6] #taking the league avg every year of FGA from 3P
```

Off rating leaders 2007-19

```
Off_Rank<- data.frame(teams_midrange[seq(1,403,31),])
Off_Rank$Year<-unlist(Off_Rank$Year)
names(Off_Rank) <- names(teams_midrange)
```

% FGA 3p compared to league average for ortg leader

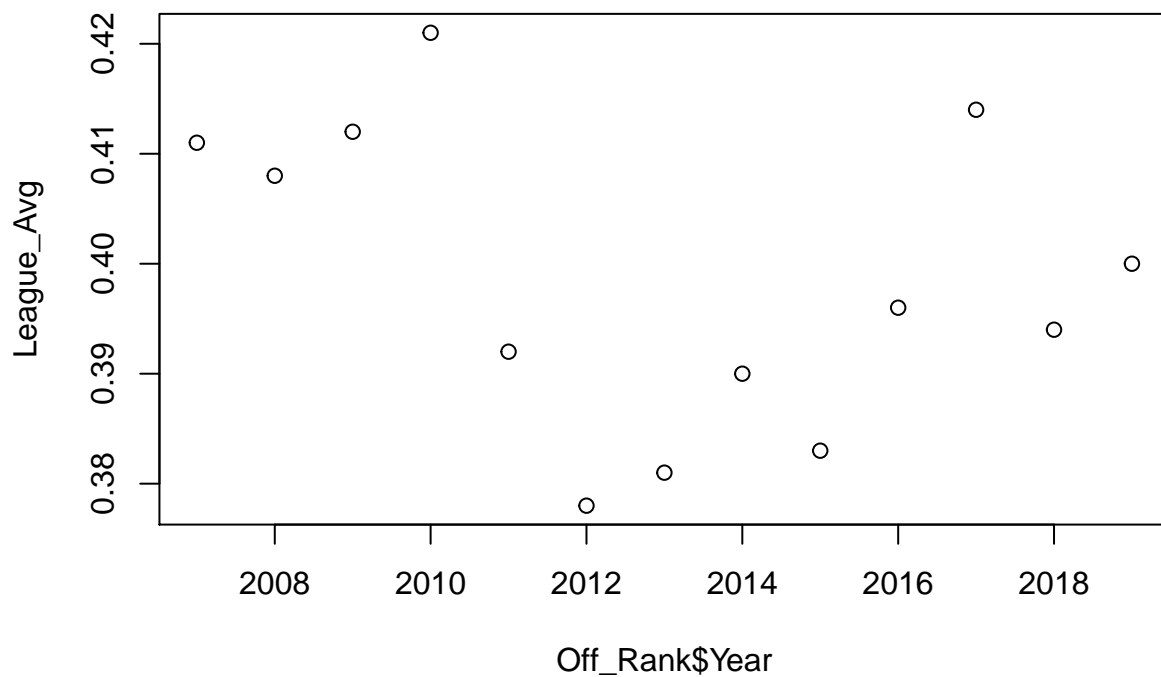
```
Off_Rank<-cbind(Off_Rank,League_Avg3) ##Difference in Off Rtg leaders 3P FGA to league avg
Off_Rank <- Off_Rank %>% select(Team, ` % of FGA by Distance 3P`, ORtg, Year, League_Avg3)
```

% field goal attempts from mid range by year

```
League_Avg<- teams_midrange[seq(31,403,31), 9]
```

Compare ortg leaders to FGA from midrange avg for league

```
Off_Rank<-cbind(Off_Rank,League_Avg) %>% select(Team, `"% of FGA by Distance 3P"`, ORtg, Year)
plot(League_Avg~Off_Rank$Year)
```



assign corresponding ranks according to %FGA mid and FG% mid

```
“{r} teams_midrange$Year<- unlist(teams_midrange$Year) teams_midrange2<-teams_midrange
%>% select(Team,% of FGA Midrange‘, Year)
```

```
teams_midrange3 <- teams_midrange %>% dplyr::group_by( Year) %>% dplyr::mutate(rank =
dplyr::dense_rank(% of FGA Midrange)) teams_midrange$% of FGA Midrange teams_midrange3 <-
teams_midrange3 %>% arrange(Year, rank)
```

```
##Plotting model (Incomplete)
#``{r}
plot(teams_midrange$ORtg~teams_midrange$` % of FGA Midrange`, xlab = "% of FGA Midrange", ylab = "Offense")
```

How many times did team with highest FGA from mid range finished inside the top 10 in offense (In progress)

“{r}”

```
teams_midrange_att<-teams_midrange names(teams_midrange_att)[9]<- "MidrangeFGAP" end =
30 start = 1 c=0 while(c<13){ teams_midrange_att[start:end,<-teams_midrange_att[start:end,]>%
arrange(.,-MidrangeFGAP) start = end+2 end = start+29 c=c+1 }
is.valu
```

```
#Further Work to be Done
#``{r}
#leaders from 2007-2019
```

```
##group by year and team, top n fn in dplyr, mutate (if in top 10, )
teams_midrange3$top <- ifelse(teams_midrange3$rank %in% 1:10, 1,0)
##plyr, sum top by team tapply(teams_midrange3$top, teams_midrange3$` Team`,sum)
```

Further investigative analysis ()

Among teams in top 10 midrange FGA every year, how many of them ranked in the top 10 in offense that year?

- out of all the teams in that span, how many times did team with highest % of midrange FGA ranked in the top 10 in offense? creat new df and order acc to highest FGA mid -> or find max FGA midrange every season (more challenging) -> nested loop to see if the team is in the top 10 in offense that season (use a counter)
- among the teams in top 10 midrange FGA every year, how many of them ranked in the top 10 in offense? same method, except we are now comparing 10 teams
- Shooting 3's should not be done blindly, even if the stats may suggest so. Hypothesis is more threes, less midrange give higher probability of winning. But, one thing this hypothesis doesnt account for is the psychological effect of mised shots. Missed shots are frustrating, and taking more threes just because they're more "efficient" than long twos doesn't alwas improve chance to win. Can expect to see frustration in the form of more fouls, more contested jumpers, more turnovers, and less energy on defense (def rtg) Ho: Lower 3p% has no effect on team/player psyche Ha: Lower 3p% has a negative effect on players psyche Examine games where teams shoot lower than their season av 3p%. In those games, expect to see lower def rtg, more TOs, more fouls and more

Multi level regression and average of all teams offense across 13 years