Midrange

Kevin Chen

December 25, 2018

Webscraping Data

```r
library(rvest)
```

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

```r
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
```

```r
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_shooti

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)

#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)
```

Select relevant columns for mid range data

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

teams_midrange<-teams_shooting[,c(1,2,5,10,11,12,15,16)]
teams_midrange$Year<-as.vector(teams_shooting%>%select(., i))
teams_midrange[,3:8]<-sapply(teams_midrange[,3:8],as.numeric)
teams_midrange$` % of FGA Midrange` <- rowSums(teams_midrange[,c(4,5)], na.rm=TRUE)
teams_midrange$` Midrange FG%` <- rowSums(teams_midrange[,c(6,7)]) #have to weigh the FG% by distance f
```

Remove unwanted characters from team stats data

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

Arranging team stats each season in alphabetical order

```r
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

```r
teams_midrange$ORtg <- as.numeric(teams_stats$ORtg) #adding ORtg from team_stats to team_shooting
teams_midrange<-teams_midrange[,c(1:8,10,11,12,9)] #rearrange columns
```

sort according to ortg in desc order

```r
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

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

```
##
## Call:
## lm(formula = teams_midrange$ORtg ~ teams_midrange$` % of FGA Midrange`)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

3

```
## -13.5273   -2.4515   -0.0895    2.2113    8.1452
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      112.0352     0.7317 153.111  < 2e-16
## teams_midrange$` % of FGA Midrange` -16.5268   2.4500  -6.746 5.34e-11
##
## (Intercept)                       ***
## teams_midrange$` % of FGA Midrange` ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.329 on 401 degrees of freedom
## Multiple R-squared:  0.1019, Adjusted R-squared:  0.09967
## F-statistic:  45.5 on 1 and 401 DF,  p-value: 5.344e-11
```

```
plot(offrtg)
```


Residuals vs Fitted

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

Normal Q–Q

Standardized residuals

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

Scale–Location

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

Residuals vs Leverage

lm(teams_midrange$ORtg ~ teams_midrange$` % of FGA Midrange`)
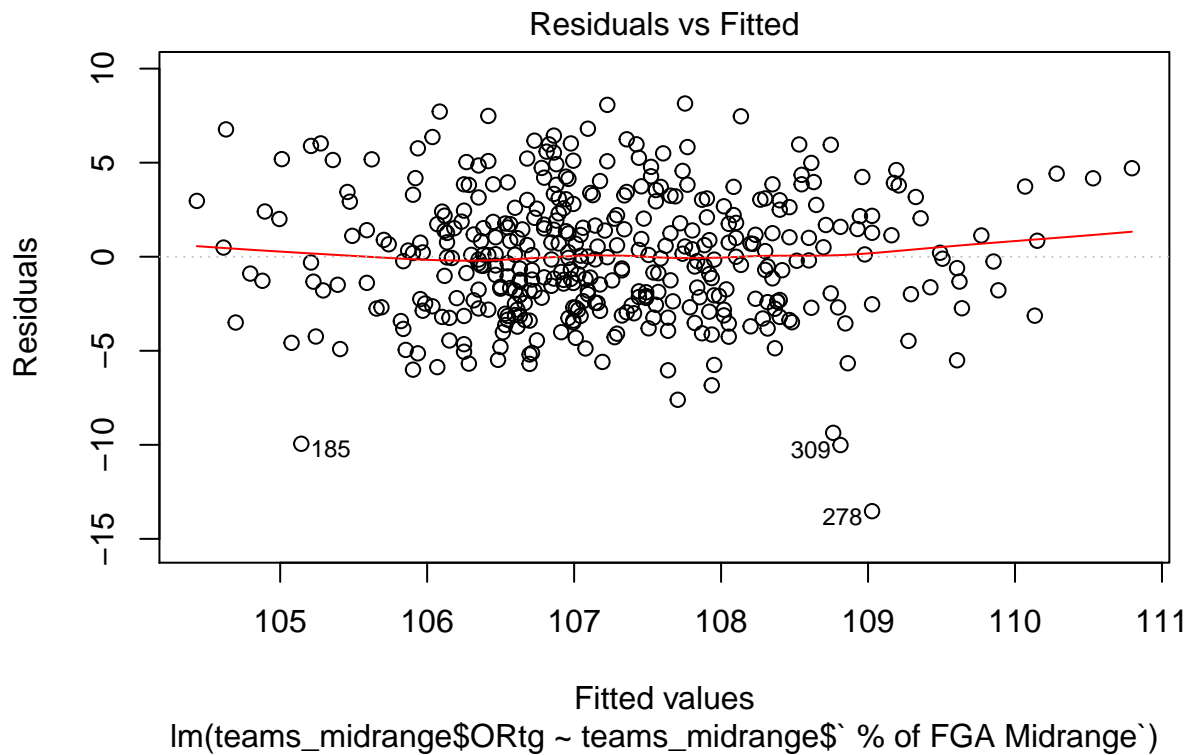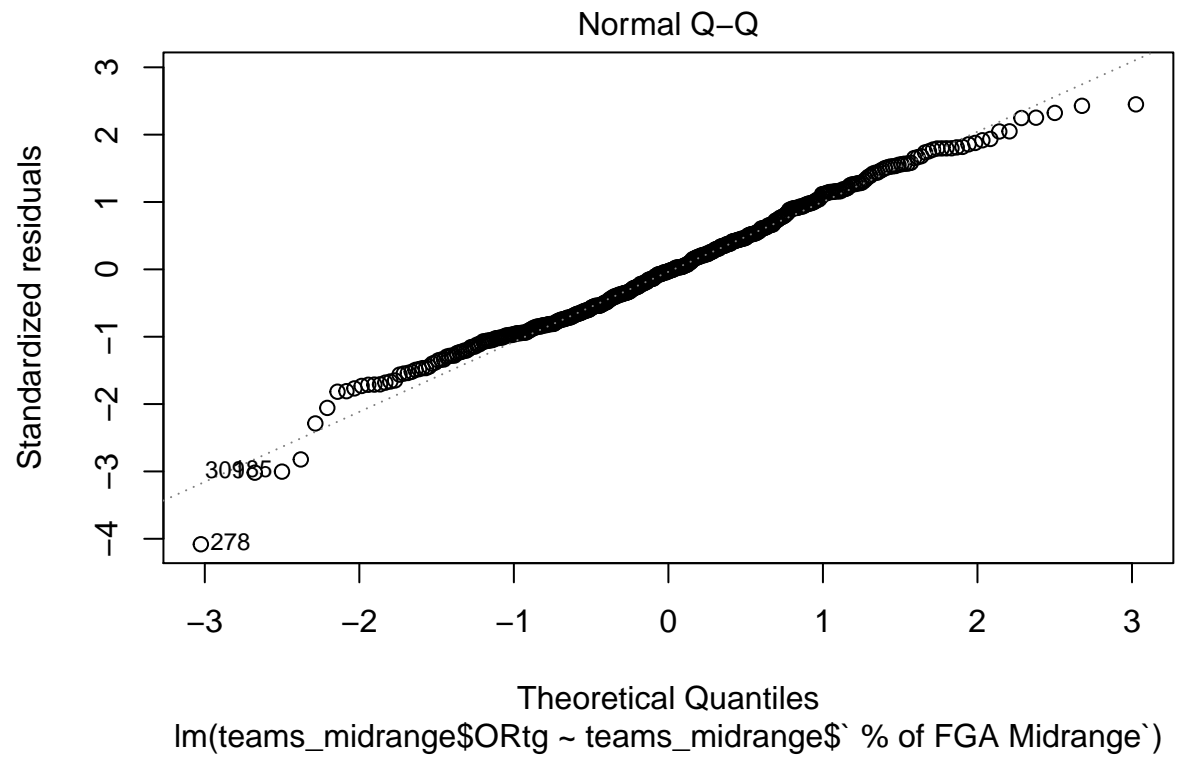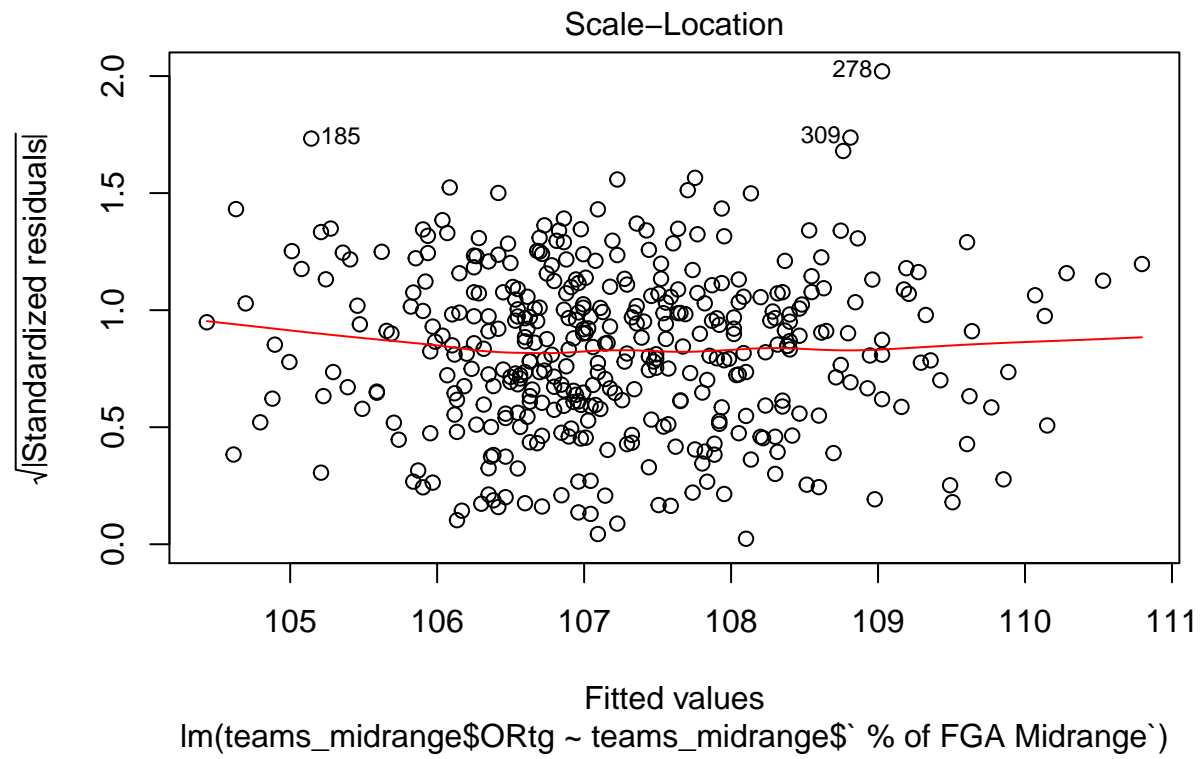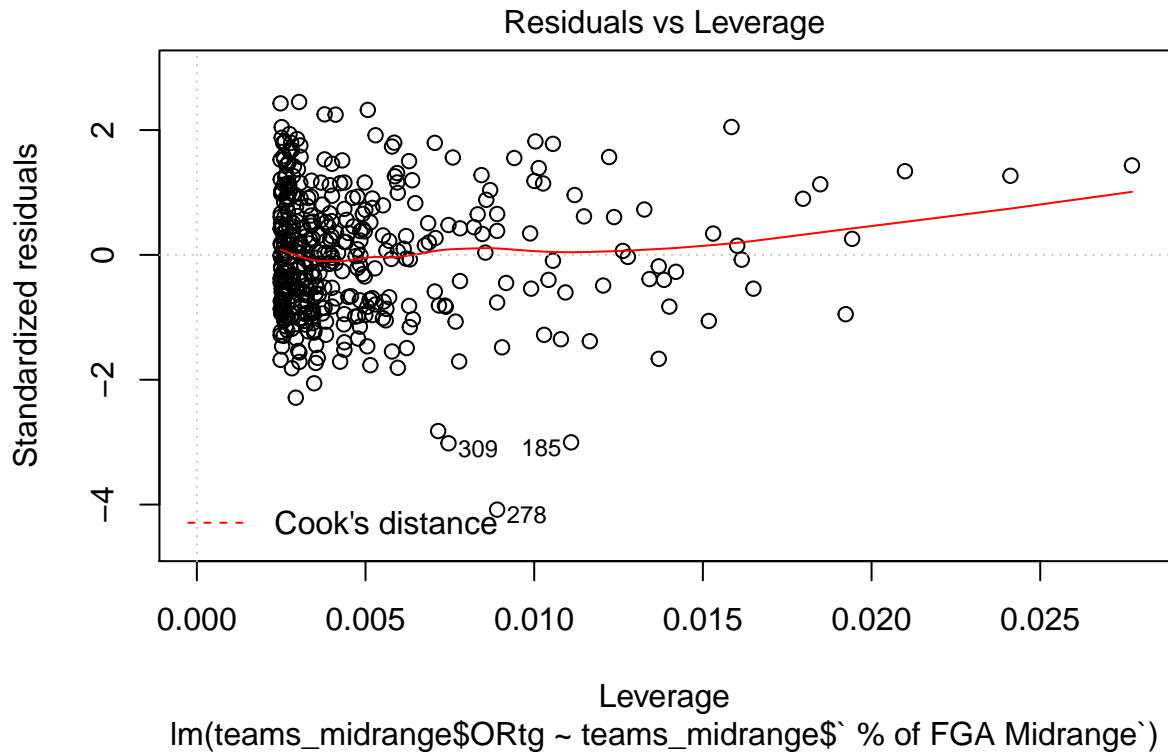
% field goal attempts from 3 by year

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

Off rating leaders 2007-19

```r
Off_Rank<- data.frame(teams_midrange[seq(1,403,31),])
Off_Rank$Year<-unlist(Off_Rank$Year)
```
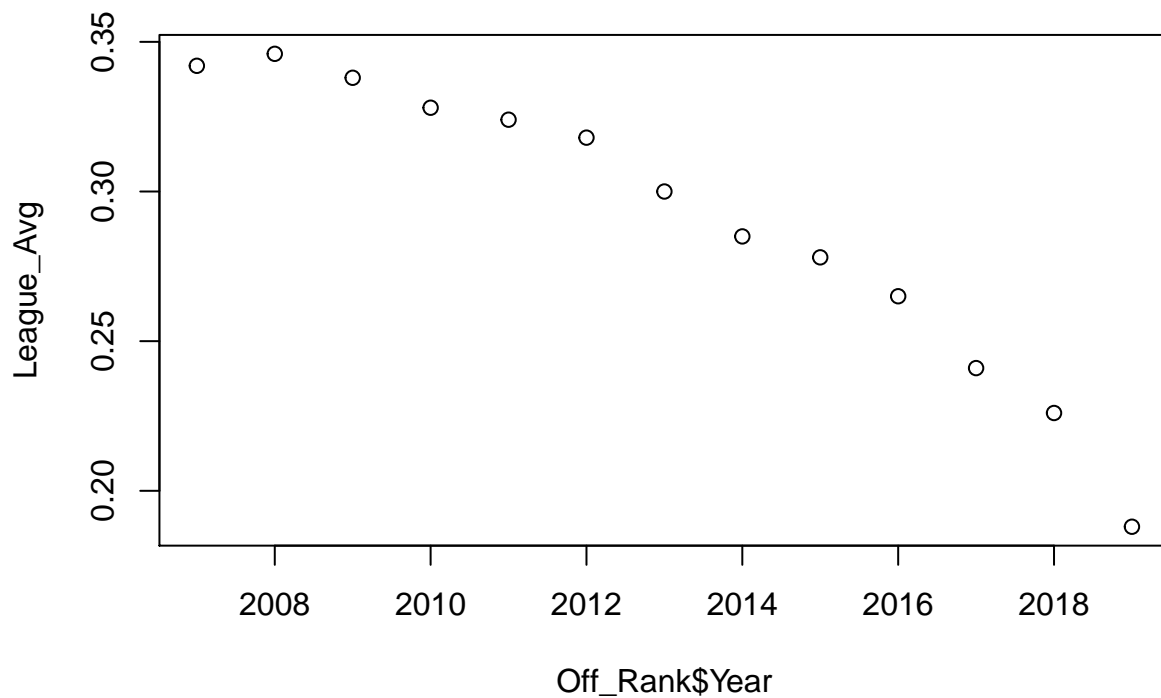
% FGA 3p compared to league average for ortg leader

```r
Off_Rank<-cbind(Off_Rank,League_Avg3) ##Difference in Off Rtg leaders 3P FGA to league avg
```

% 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)
plot(League_Avg~Off_Rank$Year)
```


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

```
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
teams_midrange$` % of FGA Midrange`
```

```
##    [1] 0.299 0.409 0.370 0.348 0.311 0.398 0.292 0.314 0.307 0.249 0.296
##   [12] 0.390 0.307 0.332 0.289 0.317 0.253 0.337 0.449 0.358 0.413 0.331
##   [23] 0.335 0.305 0.438 0.344 0.433 0.332 0.384 0.359 0.342 0.360 0.313
##   [34] 0.306 0.239 0.340 0.448 0.228 0.413 0.313 0.425 0.308 0.329 0.337
```

```
## [45] 0.432 0.343 0.357 0.302 0.381 0.344 0.327 0.327 0.271 0.295 0.327
## [56] 0.346 0.402 0.328 0.444 0.401 0.421 0.346 0.340 0.258 0.315 0.363
## [67] 0.332 0.404 0.297 0.350 0.344 0.320 0.256 0.350 0.359 0.397 0.277
## [78] 0.317 0.326 0.238 0.342 0.361 0.460 0.426 0.368 0.303 0.332 0.349
## [89] 0.301 0.329 0.386 0.320 0.338 0.291 0.324 0.312 0.226 0.349 0.294
## [100] 0.388 0.312 0.311 0.371 0.312 0.313 0.358 0.351 0.295 0.254 0.274
## [111] 0.331 0.396 0.355 0.375 0.314 0.323 0.283 0.390 0.307 0.408 0.325
## [122] 0.356 0.348 0.328 0.260 0.273 0.369 0.269 0.344 0.317 0.265 0.324
## [133] 0.309 0.321 0.331 0.307 0.332 0.260 0.354 0.311 0.383 0.367 0.373
## [144] 0.371 0.357 0.264 0.329 0.303 0.333 0.332 0.350 0.376 0.300 0.350
## [155] 0.324 0.263 0.305 0.202 0.317 0.357 0.322 0.314 0.337 0.344 0.322
## [166] 0.313 0.309 0.322 0.327 0.215 0.321 0.266 0.272 0.353 0.412 0.248
## [177] 0.323 0.333 0.248 0.411 0.323 0.336 0.374 0.369 0.417 0.318 0.316
## [188] 0.291 0.271 0.284 0.195 0.154 0.259 0.256 0.247 0.329 0.347 0.254
## [199] 0.342 0.343 0.306 0.270 0.335 0.281 0.304 0.294 0.241 0.305 0.366
## [210] 0.367 0.310 0.375 0.322 0.323 0.350 0.361 0.300 0.305 0.318 0.308
## [221] 0.114 0.252 0.298 0.271 0.261 0.265 0.284 0.333 0.269 0.291 0.272
## [232] 0.306 0.230 0.337 0.247 0.225 0.294 0.325 0.285 0.329 0.340 0.305
## [243] 0.288 0.328 0.334 0.335 0.198 0.285 0.313 0.258 0.216 0.250 0.276
## [254] 0.231 0.251 0.267 0.302 0.278 0.358 0.115 0.309 0.193 0.278 0.275
## [265] 0.257 0.296 0.241 0.280 0.302 0.368 0.327 0.363 0.356 0.304 0.266
## [276] 0.262 0.371 0.182 0.278 0.212 0.268 0.220 0.338 0.250 0.249 0.146
## [287] 0.308 0.302 0.248 0.270 0.340 0.220 0.310 0.221 0.245 0.222 0.275
## [298] 0.275 0.299 0.216 0.299 0.249 0.327 0.344 0.255 0.287 0.247 0.293
## [309] 0.195 0.265 0.236 0.091 0.207 0.186 0.278 0.273 0.182 0.277 0.307
## [320] 0.283 0.239 0.223 0.185 0.296 0.301 0.213 0.225 0.236 0.304 0.251
## [331] 0.255 0.285 0.298 0.221 0.305 0.243 0.243 0.147 0.266 0.192 0.241
## [342] 0.106 0.172 0.283 0.279 0.211 0.164 0.243 0.174 0.241 0.208 0.287
## [353] 0.216 0.230 0.288 0.238 0.208 0.287 0.226 0.223 0.299 0.145 0.199
## [364] 0.182 0.242 0.220 0.227 0.232 0.252 0.274 0.222 0.226 0.259 0.075
## [375] 0.199 0.119 0.173 0.171 0.321 0.206 0.211 0.223 0.205 0.220 0.162
## [386] 0.187 0.137 0.201 0.182 0.238 0.132 0.153 0.147 0.232 0.130 0.158
## [397] 0.219 0.166 0.196 0.207 0.167 0.225 0.188
```

```
teams_midrange3 <- teams_midrange3 %>% arrange(Year, rank)
```

Plotting model (Incomplete)

```
plot(teams_midrange$ORtg~teams_midrange$` % of FGA Midrange`, xlab = "% of FGA Midrange", ylab = "Offens
```
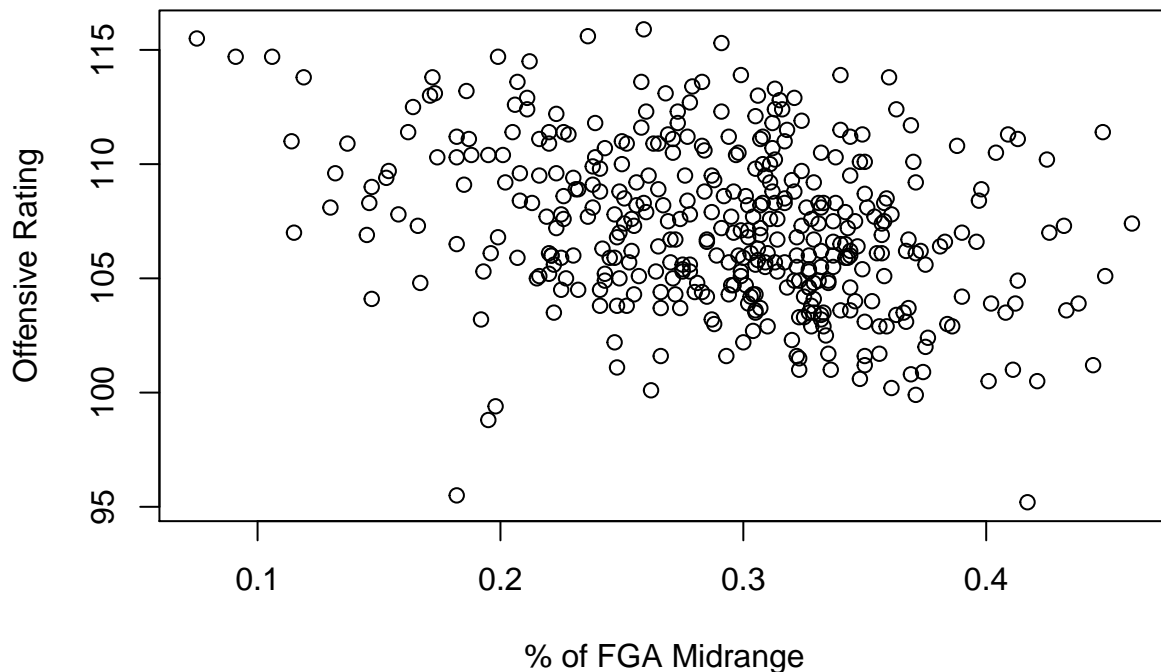
How many times did team with highest FGA from mid range finished inside the top 10 in offense (In progress)

```
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
}
```

Further Work to be Done

```
#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