Midterm Project PM566 - Relationship Between Minutes Played and Total RAPTOR Score Among NBA Players During the Regular Season (2014-2022)

AUTHOR
Derek Walker

Introduction

The National Basketball Association is the apogee of professional basketball worldwide and as a result copious amounts of data have been, and are being collected on all the commodities and moving parts of the league including but not limited to: players, coaches and teams.

With such extensive amounts of big data at our will statisticians across the globe have made it their duty to assess and express in layman terms the different trends, associations and predictions regarding team and player performance. One organization, FiveThirtyEight (538) specializes in assessing team and player performance each season and predicting said statistics for the new year. Recently the company created their own statistic called RAPTOR (**R**obust **A**lgorithm (using) **P**layer **T**racking (and) **O**n/Off **R**atings) which is used to accurately reflect how modern NBA teams actually evaluate their players. "RAPTOR, consists of two major components that are blended together to rate players: a"box" (as in "box score") component, which uses individual statistics (including statistics derived from player tracking and play-by-play data), and an "on-off" component, which evaluates a team's performance when the player and various combinations of his teammates are on or off the floor" (Silver, 2019).

The company website goes further into detail about how they created their metric and the sort of confounding factors the metric takes into account as a means of improving on player and team assessment statistics of the past. In general, a higher, or positive, RAPTOR score for a player indicates that that player improves their team's performance on both the offensive and defensive ends.

These data provided by "538", include the RAPTOR data for every player by season since 2014, as well as standard NBA metrics such as number of minutes played and number of possessions per player. Through use of these data we chose to test the hypothesis of whether there is an association between the number of minutes a player plays within a season and the total RAPTOR score (points above average per 100 possessions added by player on both offense and defense, relative to a league-average player) for said player. Additionally, by assessing this relationship it opens us up to the opportunity to explore confounding factors (number of seasons played between 2014-2022, team, coach, etc.) that relate to our independent and dependent variable as well as evaluating the effect size and bias these could potentially impose.

Methods

These data were acquired and loaded into R via the Github repository (https://github.com/fivethirtyeight/data/tree/master/nba-raptor) provided by the FiveThirtyEight company website. As stated before these data consists of 7,289 observations of RAPTOR data for each player, that played in the regular season and/or the playoffs season, along with other common metrics such as minutes played and number of possessions per player.

First we assessed the variables of interest: minutes played and raptor total score. Our first potential bias arrived when looking at the distribution of minutes played in the regular season versus the playoffs. Not only is there a significantly larger maximum for the potential number minutes played in the regular season than there is in the playoffs, but also due to the elimination style of the playoff season a player on a team that is knocked out in the first round will have lower cap on potential minutes played than a player on a team that makes it to the final. This could skew our analysis either in favor of those players that make it further in the playoff rounds or those that make early exits. To nullify this potential bias we decided to focus only on the regular season for our analysis.

The next form of potential bias we observed comes from those players that played very few minutes (<20). Many of the players playing very few minutes had extreme raptor total scores due to the small sample size of minutes played (i.e., a bunch of good plays within a very few amount of minutes will make the player appear to have a much greater impact based on the raptor metric and vise versa for a player making only bad plays in the few minutes on the court). To combat the wide variance observed for players with low numbers of minutes played we decided to subset our data to only include those players that played at least 20 minutes in a given season. In regard to missingness, our original dataset had very few missing values (<1%) and once we subset our data based on the above criterion there were no more missing values.

Using our final dataset of 5,039 player data points, we generated summary statistics for our independent and dependent variables (minutes played and Raptor_total, respectively). We produced histograms, inclusive of the mean and median values, to visualize the distribution for both variables. Next, a scatter plot, with a LOESS smooth curve, and a correlation plot were used to assess the association between the two variables. In addition, we created scatterplots and a correlation table grouping by team as a way of analyzing the confounding effects of "team" on the relationship between minutes played and total RAPTOR score. Lastly to provide a more up close analysis and better determine the impact of minutes played we assessed the trends in minutes played across seasons as well as the Total RAPTOR Score across seasons for two of the greatest players of all time, 4x NBA Finals Champion and 2x MVP award winner Stephen Curry, and 4x NBA Finals Champion and 4x MVP award winner LeBron James.

All analysis on these data were performed via R 4.3.1.

Preliminary Results

```
library(dplyr)
library(data.table)
library(ggplot2)
library(hrbrthemes)
library(cowplot)
library(Hmisc)
library(knitr)
library(corrplot)
library(gridExtra)
library(kableExtra)

nba <- fread(file='/Users/gazallab/Downloads/nba-raptor/modern_RAPTOR_by_team.csv')

dim(nba)
str(nba)
head(nba)
tail(nba)
mean(is.na(nba))#missingness

nba %>%
  summarise(n=n(),
            mean_mp = mean(mp, na.rm=T),
            med_mp = median(mp, na.rm=T),
            sd_mp = sd(mp, na.rm=T),
            min_mp = min(mp),
            max_mp = max(mp),
            mean_raptot = mean(raptor_total, na.rm=T),
            sd_raptot = sd(raptor_total, na.rm=T),
            min_raptot = min(raptor_total,na.rm=T),
            max_raptot = max(raptor_total,na.rm=T))

nba %>% group_by(season_type) %>%
  summarise(n=n(),
            mean_mp = mean(mp, na.rm=T),
            med_mp = median(mp, na.rm=T),
            sd_mp = sd(mp, na.rm=T),
            min_mp = min(mp),
            max_mp = max(mp),
            mean_raptot = mean(raptor_total, na.rm=T),
            med_raptot = median(raptor_total, na.rm=T),
            sd_raptot = sd(raptor_total, na.rm=T),
            min_raptot = min(raptor_total,na.rm=T),
            max_raptot = max(raptor_total,na.rm=T))

nbaRS <- nba %>% filter(season_type=="RS", mp > 20) %>% select(player_name,player_id,se

dim(nbaRS)
str(nbaRS)
head(nbaRS)
tail(nbaRS)
mean(is.na(nbaRS))#missingness

nbaRSmpsumstats <- nbaRS %>%  summarise("N"=n(),
            "Mean" = round(mean(mp, na.rm=T),digits=2),
            "Median" = round(median(mp, na.rm=T),digits=2),
            "Std. Deviation"= round(sd(mp, na.rm=T),digits=2),
            "Minimum" = round(min(mp),digits=2),
            "Maximum" = round(max(mp),digits=2))
nbaRSraptotsumstats<- nbaRS %>% summarise("N"=n(),
            "Mean" = round(mean(raptor_total, na.rm=T),digits=2),
            "Median" = round(median(mp, na.rm=T),digits=2),
           "Std. Deviation" = round(sd(raptor_total, na.rm=T),digits=2),
           "Minimum" = round(min(raptor_total,na.rm=T),digits=2),
           "Maximum" = round(max(raptor_total,na.rm=T),digits=2))
```

```
nbaRSmpsumstats %>%  kbl(caption = "<center><b>Table 1: Summary Statistics of Minutes P
  kable_classic(full_width = F, html_font = "Cambria")
```

Table 1: Summary Statistics of Minutes Played in NBA Regular Season (2014-2022)

N	Mean	Median	Std. Deviation	Minimum	Maximum
5039	1030.42	879	802.34	21	3125

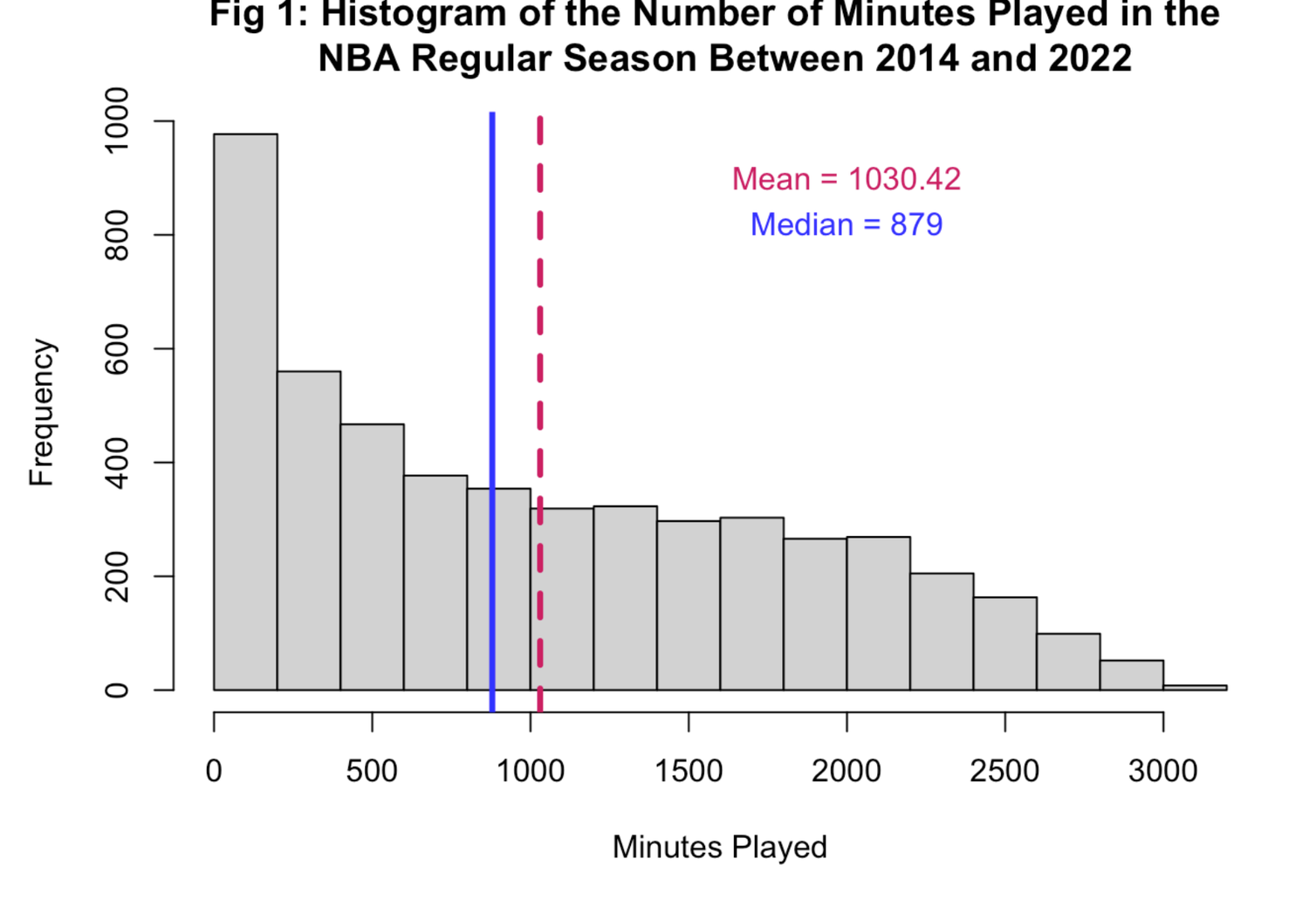
```
nbaRSraptotsumstats %>%  kbl(caption = "<center><b>Table 2: Summary Statistics of Total
  kable_classic(full_width = F, html_font = "Cambria")
```

Table 2: Summary Statistics of Total RAPTOR Score in NBA Regular Season (2014-2022)

N	Mean	Median	Std. Deviation	Minimum	Maximum
5039	-1.5	879	4.32	-31.68	20.48

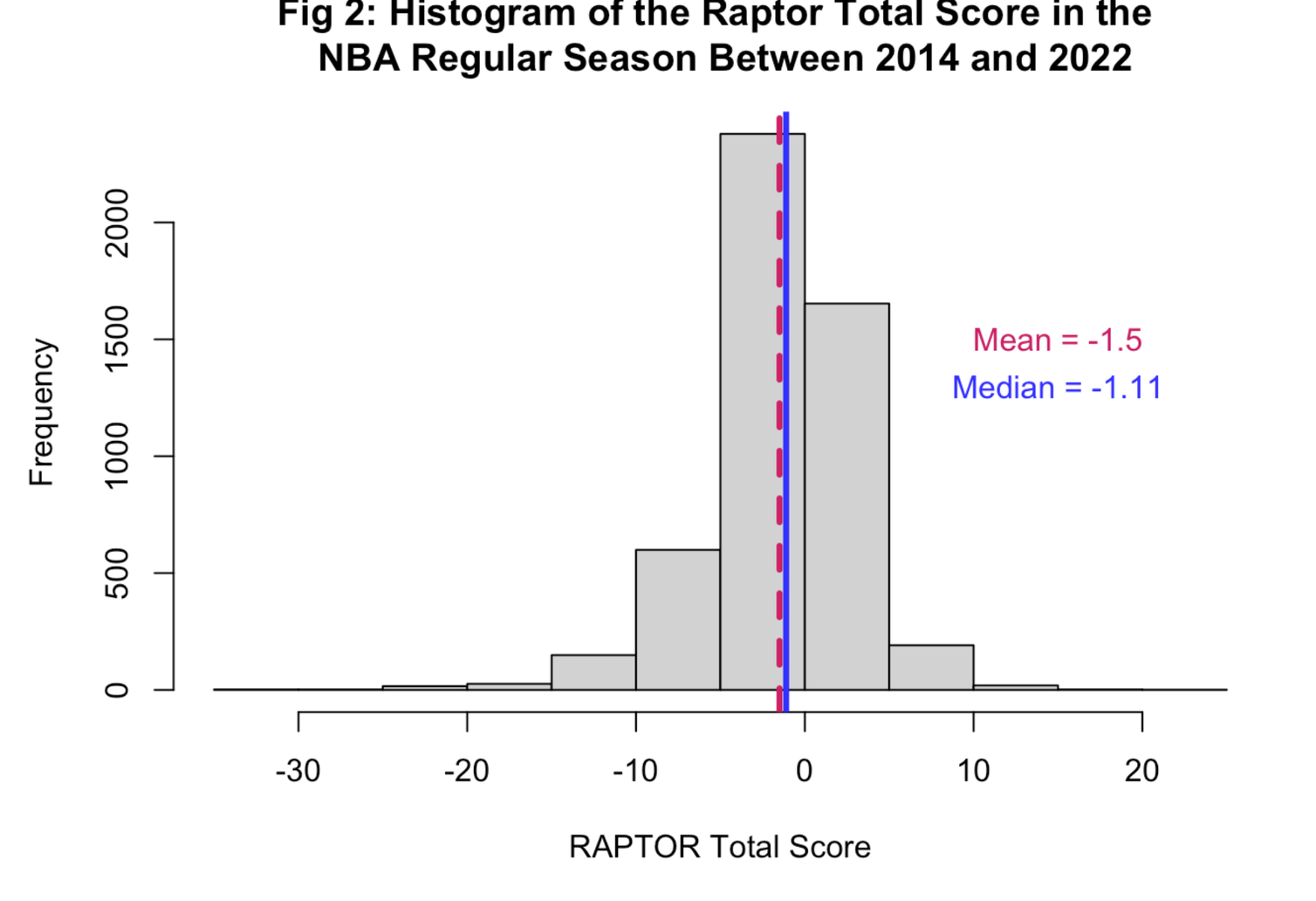
```
hist(nbaRS$mp, main="Fig 1: Histogram of the Number of Minutes Played in the \n NBA Reg
abline(v = mean(nbaRS$mp), col = "#CC0066", lty=2, lwd = 3)
abline(v = median(nbaRS$mp), col = "#3333FF", lwd = 3)
text(x = 2000, y = 900, paste("Mean =", round((mean(nbaRS$mp)),digits=2)), col = "#CC00
text(x = 2000, y = 820, paste("Median =", round((median(nbaRS$mp)),digits=2)), col = "#
```

Fig 1: Histogram of the Number of Minutes Played in the NBA Regular Season Between 2014 and 2022




```
hist(nbaRS$raptor_total, main="Fig 2: Histogram of the Raptor Total Score in the \n NBA
abline(v = mean(nbaRS$raptor_total), col = "#CC0066", lty=2, lwd = 3)
abline(v = median(nbaRS$raptor_total), col = "#3333FF", lwd = 3)
text(x = 15, y = 1500, paste("Mean =", round((mean(nbaRS$raptor_total)),digits=2)), col
text(x = 15, y = 1300, paste("Median =", round((median(nbaRS$raptor_total)),digits=2)),
```

Fig 2: Histogram of the Raptor Total Score in the NBA Regular Season Between 2014 and 2022



The distribution of minutes played (Figure 1) is skewed right and has a fairly large standard deviation (802.34). We notice a large portion of the players played less than the mean number of minutes. The distribution of the total RAPTOR score (Figure 2) is rather symmetric with only a small portion of extreme values. Based on the distribution it appears that a majority of the players had a RAPTOR total score that was less than zero.

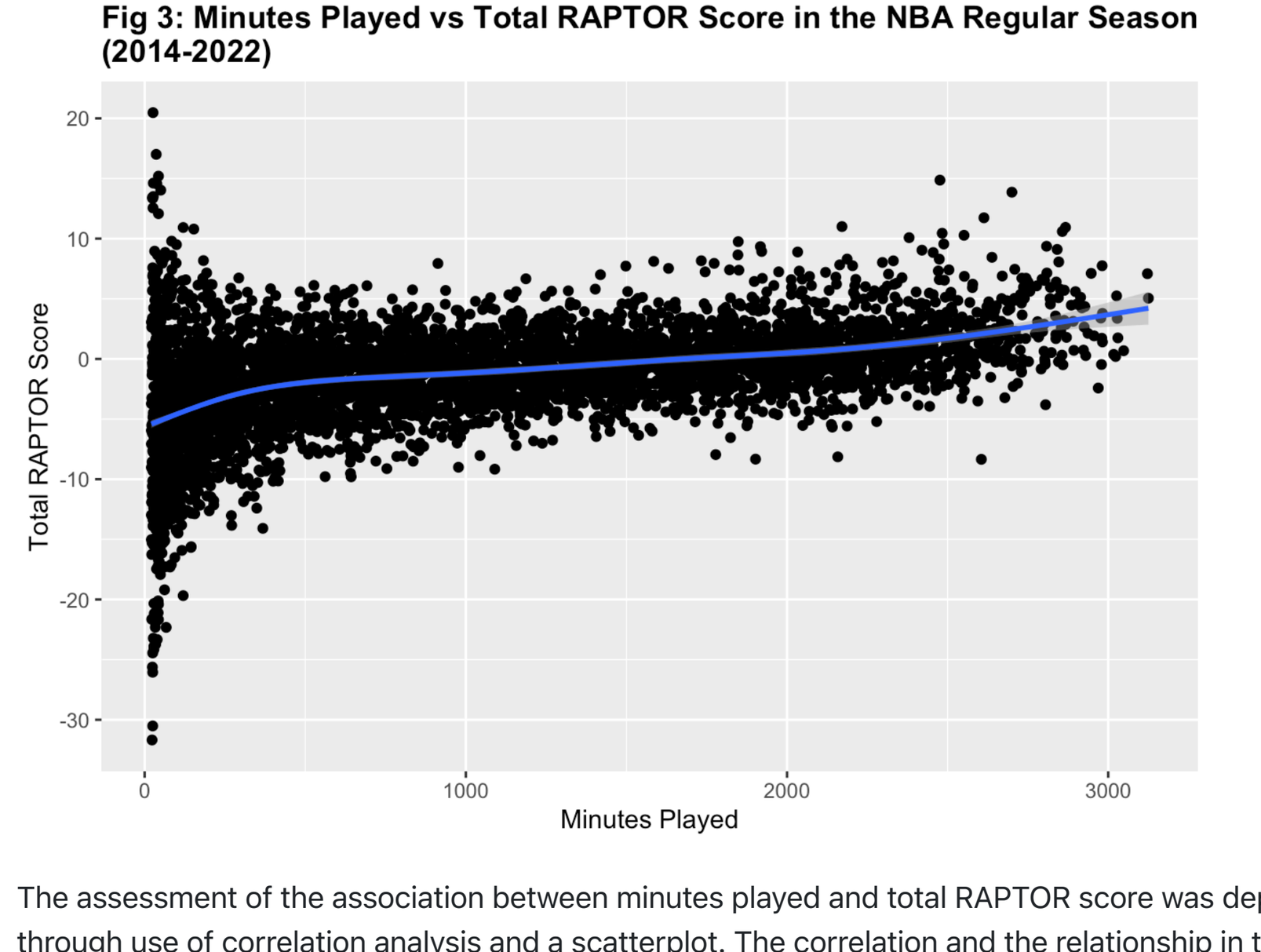
```
##Correlation Analysis: Minutes vs Rap_tot
nbaRScor <- cor(nbaRS[,c(5,6)])
nbaRScor_sub <- as.data.table(nbaRScor[2,1])
colnames(nbaRScor_sub) <- c("Corr (ρ)")
nbaRScor_sub %>% kbl(caption = "<center><b>Table 3: Correlation (ρ) of Minutes Played v
  kable_classic(full_width = F, html_font = "Cambria")
```

Table 3: Correlation (ρ) of Minutes Played versus Total RAPTOR Score in NBA Regular Season (2014-2022)

Corr (ρ)
0.4505772

```
##Scatterplot: Minutes vs Rap_tot
ggplot(data = nbaRS, mapping = aes(x = mp, y = raptor_total)) +
  geom_point() + geom_smooth(mapping = aes(x = mp, y = raptor_total)) +
  labs(x  = "Minutes Played", y = "Total RAPTOR Score") + ggtitle("Fig 3: Minutes Playe
```

Fig 3: Minutes Played vs Total RAPTOR Score in the NBA Regular Season (2014-2022)



The assessment of the association between minutes played and total RAPTOR score was depicted through use of correlation analysis and a scatterplot. The correlation and the relationship in the scatterplot were moderately positive with our correlation being 0.45. This is a clear indication that the more minutes a player plays, the higher their total RAPTOR score will end up being. In addition, one can observe the large variance that was mentioned in the players with few minutes played.

```
#Correlation: Minutes vs Rap_tot by Team
nbaRScorteam <- nbaRS %>% group_by(team) %>%
  summarise(cor=round(cor(mp,raptor_total),digits=2)) %>%
  arrange(-cor) %>% rename("Team"=team, "Corr (ρ)"=cor)

nbaRScorteam %>% kbl(caption = "<center><b>Table 4: Correlation (ρ) of Minutes Played v
  kable_classic(full_width = F, html_font = "Cambria")
```

Table 4: Correlation (ρ) of Minutes Played versus Total RAPTOR Score in NBA Regular Season (2014-2022) Grouped by Team

Team	Corr (ρ)
GSW	0.65
MIA	0.59
POR	0.58
HOU	0.57
OKC	0.56
UTA	0.54
SAS	0.53
ATL	0.52
CHA	0.50
TOR	0.49
DEN	0.48
IND	0.48
MIL	0.48
LAC	0.47
WAS	0.47
PHO	0.46
ORL	0.45
NOP	0.44
CLE	0.42
MIN	0.42
MEM	0.41
DAL	0.39
PHI	0.38
BOS	0.37
DET	0.33
SAC	0.33
CHI	0.30
LAL	0.30
BRK	0.25
NYK	0.22

```
#Scatterplot: Minutes vs Rap_tot by Team
ggplot(data=nbaRS, aes(mp, raptor_total, col=team)) +
  geom_point(mapping = aes(x = mp, y = raptor_total, color=team)) + stat_smooth(method=
  facet_wrap(~ team, nrow = 3) + xlab("Minutes Played") + ylab("Total RAPTOR Score") +
```

Fig 4: Linear Regression of Minutes Played vs Total RAPTOR Score by Team

To assess whether we the same patterns can be observed across teams we analyzed our data grouping our correlated analysis and scatter plot by team. Based on this analysis we observe the same moderately positive association across all teams. Certain teams such as the Golden State Warriors and the Miami Heat had extremely high correlation between minutes played and total RAPTOR score as compared to teams with lower correlation such as the New York Knicks and the Brooklyn Nets. This observed pattern does in fact directly correlate with the success of these teams and their players within recent years.

```
#One player Evaluation (Stephen Curry)
dt <- nbaRS %>% filter(nbaRS$player_name =="Stephen Curry") %>% select (player_name,sea

p1 <- ggplot(data=dt, aes(season, mp)) + geom_point( mapping = aes(x = season, y = mp),

p2 <- ggplot(data=dt, aes(season, raptor_total)) + geom_point( mapping = aes(x = season

plot_grid(p1,p2)
```

Fig 5.1: Scatter Plot of Minutes Played for Stephen Curry each Season

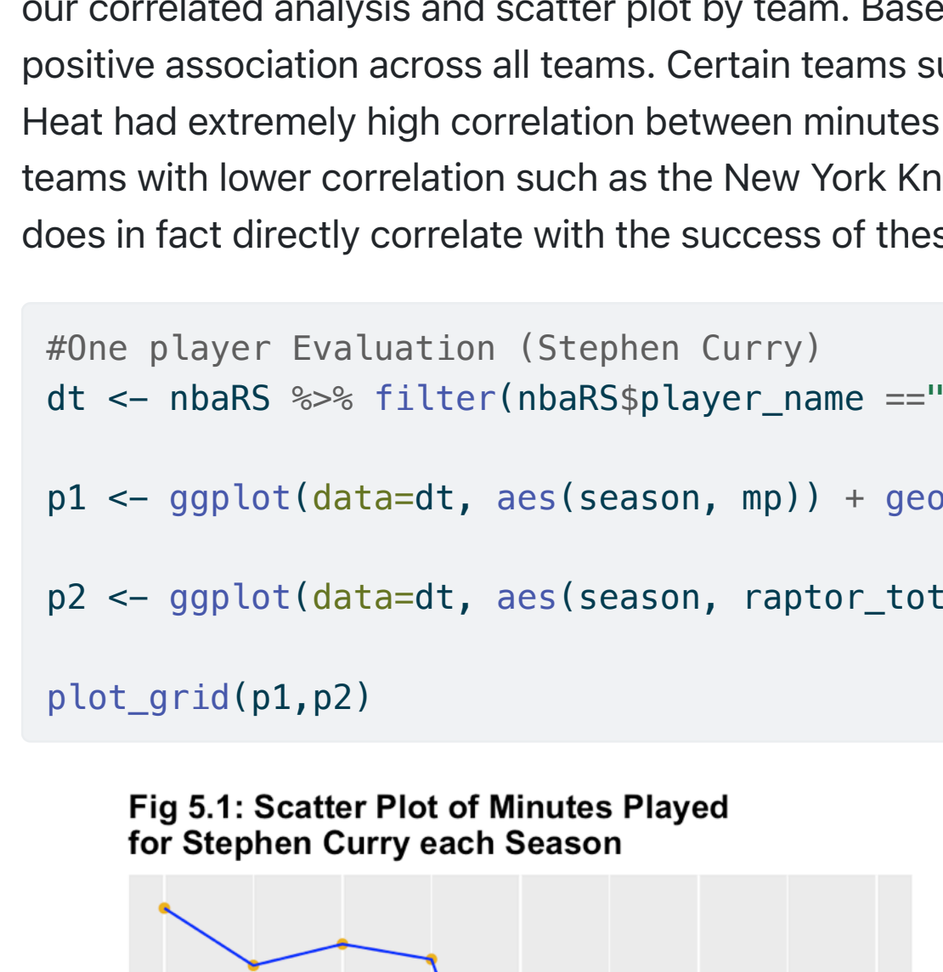
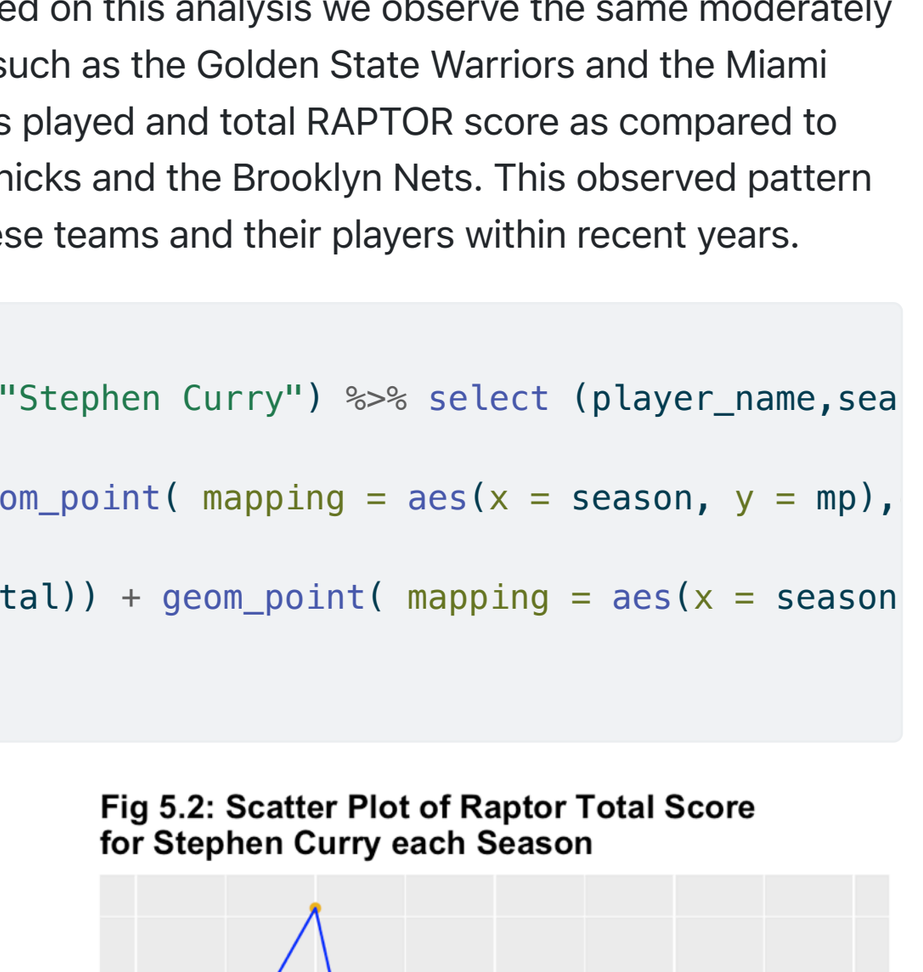


Fig 5.2: Scatter Plot of Raptor Total Score for Stephen Curry each Season




```
dt2 <- nbaRS %>% filter(nbaRS$player_name =="LeBron James") %>% select (player_name,tea

p3 <- ggplot(data=dt2, aes(season, mp)) + geom_point( mapping = aes(x = season, y = mp)

p4 <- ggplot(data=dt2, aes(season, raptor_total)) + geom_point( mapping = aes(x = seaso

plot_grid(p3,p4)
```

Fig 6.1: Scatter Plot of Minutes Played for LeBron James each Season

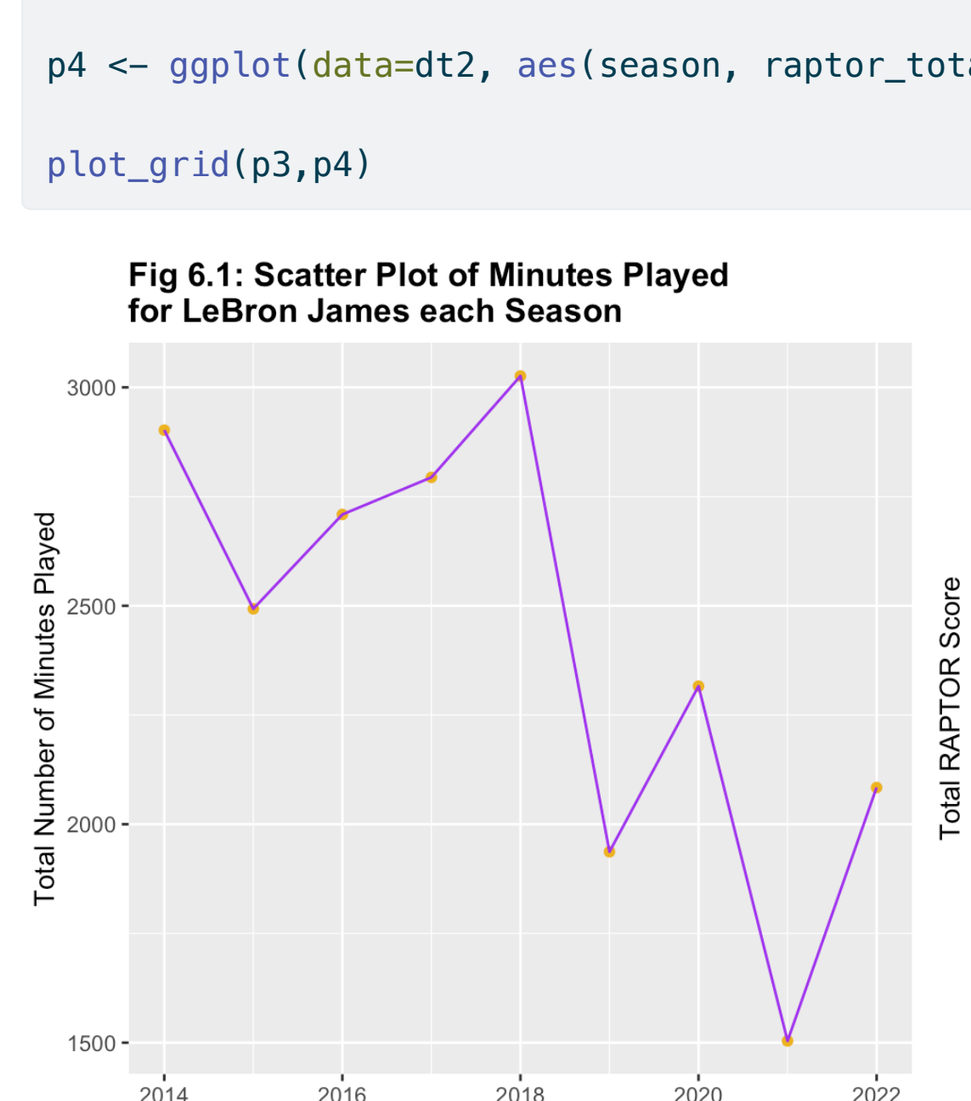
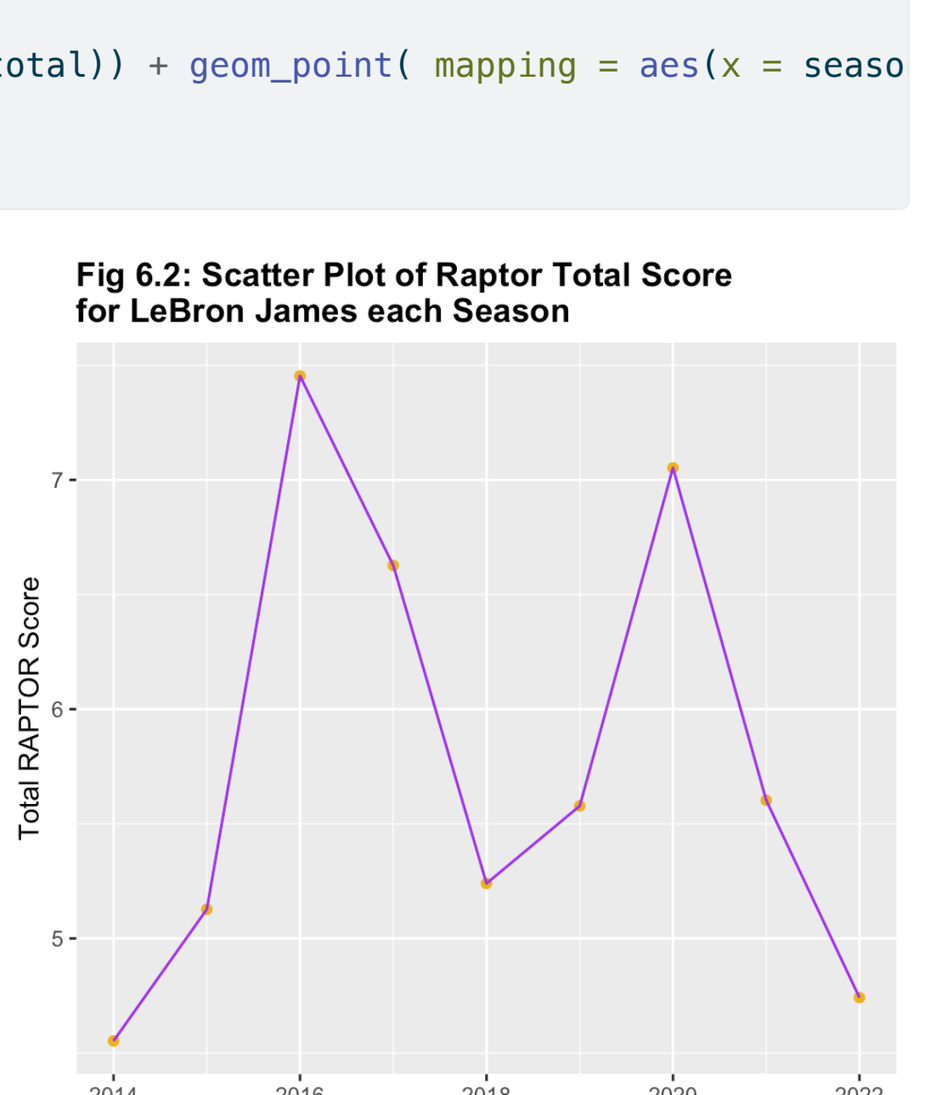


Fig 6.2: Scatter Plot of Raptor Total Score for LeBron James each Season



Our assessment at the player level looking at both Stephen Curry and LeBron James indicates that there is an association between minutes played and total RAPTOR score at the individual level for these players. In addition, when observing our variables at the individual player level (Stephen Curry, LeBron James) across all 9 NBA seasons, we observed more similarities between the two plots for Stephen Curry, a player who has been on the same team for all 9 seasons, than for LeBron where we observed a few more examples of indirect correlation between minutes played and total RAPTOR score for him from season to season. This indicats that confounding effects such as team chemistry can have an effect on the overall relationship between minutes played and total RAPTOR score for an individual player. Generally, both graphs at the individual player level do show a positive association between minutes played and total RAPTOR score.

Conclusion

Within the NBA regular season, between the years of 2014 and 2022, there is a moderate positive association and a moderate positive correlation ($\rho = 0.45$) between the number of minutes played by a player (mp) and the total RAPTOR score (RAPTOR_total). This indicates that the more game experience a player has, the higher the chances they will have a positive impact on their team's performance.

Based on our team level analysis and player level analysis, certain teams and certain players have higher correlation between minutes played and total RAPTOR score than other teams and/or players. This shows that a players team and many other potential confounding factors such as team chemistry, coaching style and/or injury can possibly have a rather large effect on the relationship between minutes played and total RAPTOR score for a player. This may have been overlooked in our analysis given these sorts of confounding data were not provided.

For future work in this topic it would be beneficial to assess the relationship between minutes played and total RAPTOR score while adjusting for these potential confounding factors: team chemistry, coaching style/history, player position, and injuries. These, as well as many other potential factors, all have the potential to impact a players number of minutes played as well as their total RAPTOR score for a season. With more information on these confounding variables we can assess other trends and patterns related to player performance as well. Overall, there is still much to be discovered in the statistics of basketball and with countless amounts of data being presented the opportunity to perform analysis and produce predictions will always be there to profit.

Cited Works

natesilver538. "How Our Raptor Metric Works." FiveThirtyEight, FiveThirtyEight, 10 Oct. 2019, fivethirtyeight.com/features/how-our-raptor-metric-works/.