



MONEY BALL IN THE NBA

Gaining insight in NBA statistics and helping franchises make better choices. The Moneyball question: Does it work?

Justin Huang

CONTEXT

- National Basketball Association composed of 29 American teams and 1 Canadian team was founded in New York on June 6, 1946. It has grown from an American based game to an globalized sport surpassing soccer over the last few years.
- The game is played with 10 players, 5 on each side, on the court trying to score more points than their opponent. Up to 15 players can be in uniform at any time on a team.
- The agreement between the 30 teams owners and the players association is the collective bargaining agreement. Through the 2000-2018 NBA season there has been 4 CBA's. 1999, 2005, 2011 and 2017.
- Unlike other professional American sports, the NBA runs on a soft cap instead of a hard cap. This allows salary cap specialists on teams to make the most out of the current seasons salary cap.
- What mainly affects the salary cap each year is basketball related income. Basketball related income is all revenue including television and merchandising related to the NBA. If the amount of money available rises the salary cap and thus the amount players can get rises as well.
 - [Tv Money](#)
- For the 30 teams there are 30 general managers. His or her job is to oversee basketball operations, hiring and firing coaches and assistants. They have a financial role as well as predicting talent by choosing the right coaches, player trades and direction of the team.

MONEYBALL AND THE NBA

- Moneyball is the real life story of Billy Jeanne who managed the Oakland Athletics. With a tight budget he produced a winning team using advanced player analytics to replace overpaid players with a combination of cheaper ones.
- So will Moneyball work in the NBA? Many general managers believe so. The one leading the charge in advanced player analytics and team play statistics is Darryl Morey the general manager of the Houston Rockets. In the 2018 NBA season they won a league best 65 games out of 82 games.
- They built their team around 2 superstars, 3 point shooting while playing small ball. Where teams would spread the floor and play at a fast up and down pace to outscore the other team.
- Differences between Baseball and Basketball
 - 15 active players vs 25 in baseball.
 - Soft cap vs Hard cap
 - 5 man vs 9 man playing at once.
 - 9 innings vs 48 minutes played.

METHODOLOGY

- Data Exploration And Analysis
 - Looking at Trends in the NBA
 - 3 pointers vs 2 pointers
 - The NBA is it becoming smaller
 - Plus Minus advanced analytics problem
 - Scouting
 - International
 - Colleges
 - Drafting number 1 and second round picks
 - Best vs Worst teams 2000-2018
 - Salary Trends
 - Best overall and worst overall and what they did
 - Avg Joe VS
 - A look at if we created an average player and comparing him against the top players, role players and award categories.
- Regression Analysis
 - Multiple linear regression of the 2000-2018 season
 - Multiple linear regression splitting it up into 3 CBAs, 1999, 2005 and 2011

DATA SOURCE

- NBA STATS
 - The official NBA statistics site for player and team statistics
 - <https://stats.nba.com>
- Basketball Reference
 - A sports reference website that has collected historical data on all sports.
 - <https://www.basketball-reference.com>
- Patricia Bender
 - An individual who analyzed the state of the last 17 years of the CBA and provided salary information to Basketball Reference
 - <https://www.eskimo.com/~pbender/>
- CBA
 - Information on the changes and agreements of the collective bargaining agreements.
 - <http://www.cbafaq.com/salarycap.htm>
- NBA Contract
 - Analyzing an NBA contract explaining the different types of contracts as well as NBA player pensions.
 - <https://atlhawksfanatic.github.io/NBA-CBA/uniform-player-contract.html#bonuses>.
- Hoops Hype
 - NBA news site, and salary reference
 - <https://hoopshype.com>
- Land of Basketball
 - Website containing historical information on awards and nominations of the NBA
 - <https://www.landofbasketball.com>

DATA WRANGLING AND CLEANING

Full process can be found here:

https://github.com/jvhuang1786/capstoneproject/blob/master/nba_data_wrangle.html

▪ Hoopshype webscrap for salary data

```
salary_2018 <- read_html("https://hoopshype.com/salaries/players/2017-2018/") %>%
  html_nodes("table") %>% html_table
```

▪ Patricia Benders, Basketball reference and NBA STATS csv file

▪ Problem with different spellings of American and Passport names as well as suffixes

```
#getting rid of periods in name
grep("\\\\*", age$player_name, value = T)
grep("\\\\*", stat$player_name, value = T)
stat$player_name <- gsub("\\\\*", "", stat$player_name)
grep("Ray Allen", stat$player_name, value = T)
grep("\\\\*", salary$player_name, value = T)
grep("\\\\.", age$player_name)
age$player_name <- gsub("\\\\.", "", age$player_name)
grep(".Jr$", age$player_name, value = TRUE) # 6 Jrs here.
grep("Glen Rice", age$player_name, value = TRUE)
grep(".Jr$", stat$player_name, value = TRUE)
#look through all stats for which one junior is missing for a total of 13 was missing in stats.
grep(".Jr$", salary$player_name, value = TRUE)
##### Filling all the juniors
#stat
grep("Roger Mason", stat$player_name, value = TRUE)
stat$player_name<-gsub("Roger Mason", "Roger Mason Jr", stat$player_name)
#testing salary
grep("Roger Mason", salary$player_name, value = TRUE)
salary$player_name<-gsub("Roger Mason", "Roger Mason Jr", salary$player_name)
#-----
```

```
#nene hilario weird spelling
grep("^Ne", stat$player_name, value = TRUE)
grep("^Nene", age$player_name, value = TRUE) # located in age
grep("^Ne", age$player_name)
age$player_name <- gsub("Nene", "Nene Hilario", age$player_name)
grep("^Nen", salary$player_name, value = TRUE)
salary$player_name <- gsub("Nenê", "Nene Hilario", salary$player_name)
```

Ran full join by name and searched for duplicated names and removed them,

```
#time to combine all of them
data <- full_join(salary, age, by = "player_name")
data <- full_join(data, stat, by = "player_name")

#have to fill in the N/As now.
colnames(data)[2] <- "salary"
colnames(data)[39] <- "ppg"
sum(is.na(data$salary))
which(is.na(data$salary))
sum(is.na(data$TEAM))
sum(is.na(data$ppg))

#find duplicated rows
which(duplicated(data$player_name))
```

Put in Award Data for each year

```
#Nba first team

a1<-grep("Tim Duncan", test$player_name)
test[a1, 48] <- "YES"
a2<-grep("Karl Malone", test$player_name)
test[a2, 48] <- "YES"
a3<-grep("Allen Iverson", test$player_name)
test[a3, 48] <- "YES"
a4<-grep("Jason Kidd", test$player_name)
test[a4, 48] <- "YES"
a5<-grep("Alonzo Mourning", test$player_name)
test[a5, 48] <- "YES"
```

Changed different columns to numeric, characters or factors

```
#Making salary into numeric
data$salary <- str_replace(data$salary, "\\$", "")
data$salary <- str_replace_all(data$salary, ",", "")
data$salary <- as.numeric(data$salary)
rownames(data) <- NULL
```



DATA EXPLORATION

Full code is here:

https://github.com/jvhuang1786/capstoneproject/blob/master/nba_data_visV4.html

- Initial observations when cleaning and preparing the data
 - Contractual Obligated players who didn't play but still got paid
 - Dealing with released players, injured players, players with contract disputes
 - Huge salary cap increase
 - With the newly signed tv deal and globalization of the game major increase in basketball related income.
 - Right skew in data
 - Most of the NBA data distribution is right skewed. With a few players taking more of the shots or getting more of the pay.

Contractual Obligated players who didn't play but still got paid

Because of the structure of an NBA contract that is legally binding. Teams are held accountable for players that are injured, released, retired or chose not to play that season.

Season	Age	Tm	Lg	Pos	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
2002-03 *	22	HOU	NBA	C	82	72	29.0	4.9	9.8	.498	0.0	0.0	.500	4.9	9.8	.498	.499	3.7	4.5	.811	2.4	5.8	8.2	1.7	0.4	1.8	2.1	2.8	13.5
2003-04 *	23	HOU	NBA	C	82	82	32.8	6.5	12.5	.522	0.0	0.0	.000	6.5	12.5	.523	.522	4.4	5.4	.809	2.4	6.6	9.0	1.5	0.3	1.9	2.5	3.3	17.5
2004-05 *	24	HOU	NBA	C	80	80	30.6	6.7	12.2	.552	0.0	0.0		6.7	12.2	.552	.552	4.9	6.2	.783	2.6	5.8	8.4	0.8	0.4	2.0	2.5	3.7	18.3
2005-06 *	25	HOU	NBA	C	57	57	34.2	8.2	15.8	.519	0.0	0.0	.000	8.2	15.8	.519	.519	5.9	6.9	.853	2.6	7.6	10.2	1.5	0.5	1.6	2.6	3.4	22.3
2006-07 *	26	HOU	NBA	C	48	48	33.8	8.8	17.1	.516	0.0	0.0	.000	8.8	17.0	.518	.516	7.4	8.6	.862	2.1	7.3	9.4	2.0	0.4	2.0	3.5	3.3	25.0
2007-08 *	27	HOU	NBA	C	55	55	37.2	7.9	15.5	.507	0.0	0.0	.000	7.9	15.5	.508	.507	6.3	7.4	.850	3.1	7.7	10.8	2.3	0.5	2.0	3.3	3.1	22.0
2008-09 *	28	HOU	NBA	C	77	77	33.6	7.4	13.4	.548	0.0	0.0	1.000	7.3	13.4	.548	.549	4.9	5.7	.866	2.6	7.2	9.9	1.8	0.4	1.9	3.0	3.3	19.7
2009-10	29	Did Not Play (injury—foot)																											
2010-11 *	30	HOU	NBA	C	5	5	18.2	3.6	7.4	.486	0.0	0.0		3.6	7.4	.486	.486	3.0	3.2	.938	1.4	4.0	5.4	0.8	0.0	1.6	1.4	2.6	10.2

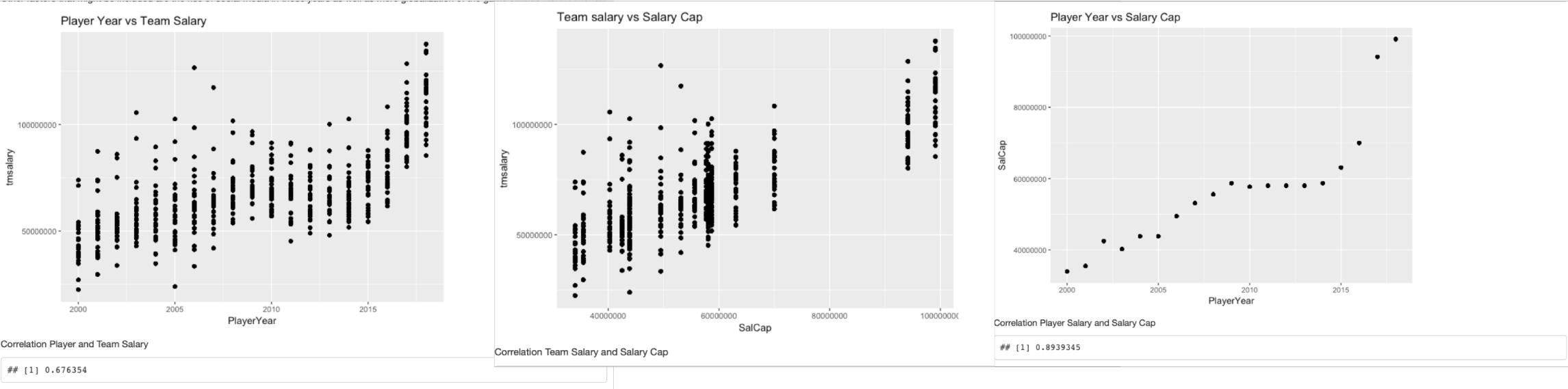
```
lessthanfive<-which(df$Games <5)
df <- df[-lessthanfive,]
lessthanfiveminGame <- which(df$MinGames < 5)
df <- df[-lessthanfiveminGame,]
lessthanoneppg<-which(df$ppg < 1)
df <- df[-lessthanoneppg,]
#decided not to use the incomplete 2019 dataset
twennineteen<-which(df$PlayerYear == 2019)
df <- df[-twennineteen,]
rownames(df) <- NULL
```

Ended up removing players who

- Did not play more than 5 games
- Less than 5 minutes a game
- Did not score one point per game
- Also removed the incomplete 2019 season

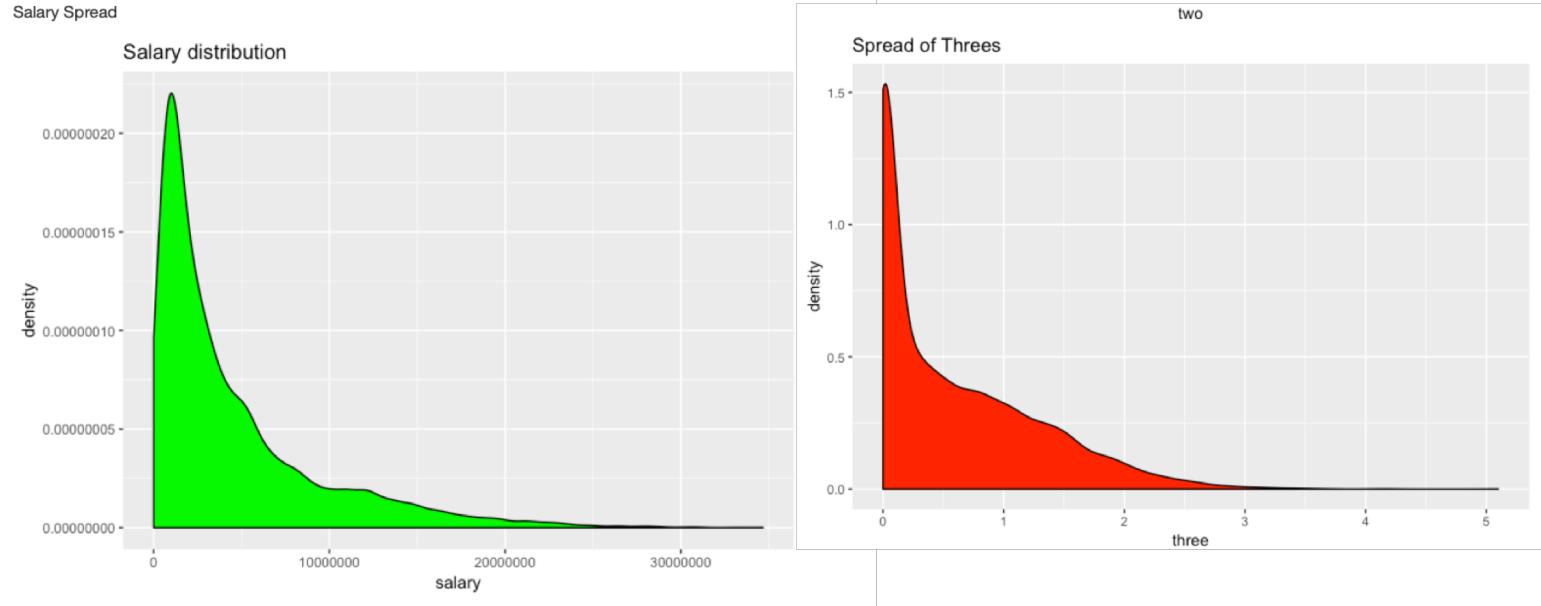


The Huge Salary Cap increase

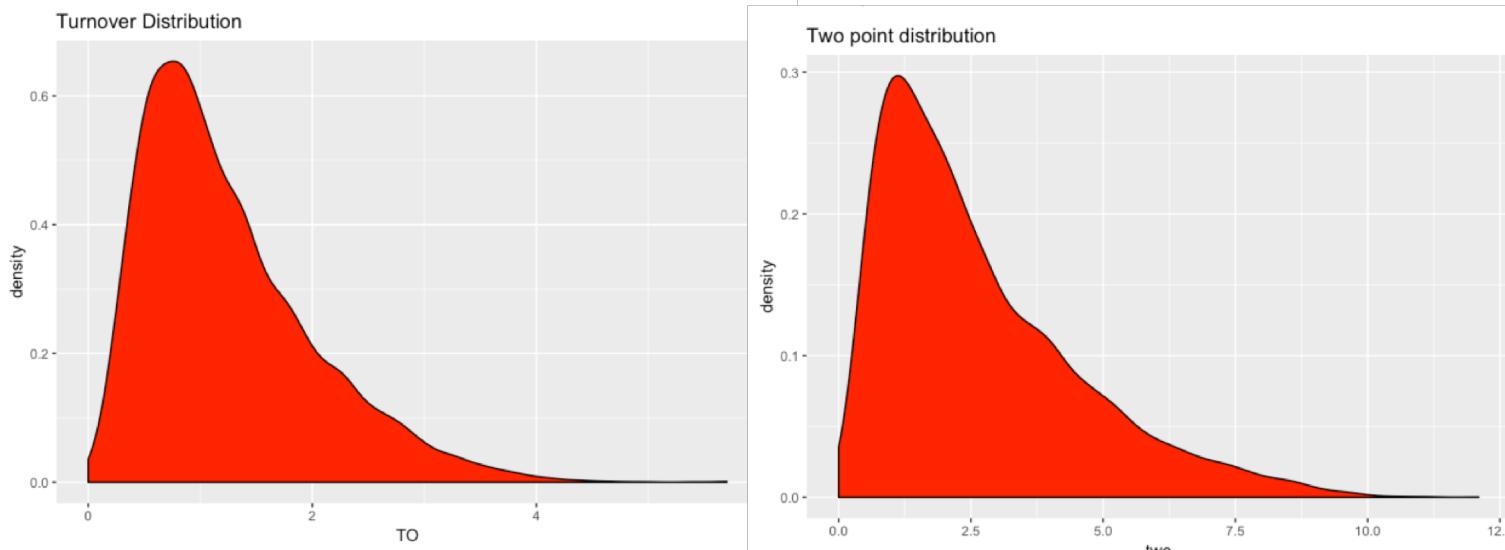


- Huge increase after the 2015 season that is not explained by inflation alone.
- New TV contract and a more global game with more global brands in the game of basketball which increases BRI.

Right Skewed Data



- With how max contracts are structured. They take 25 to 35 percent of the teams salary cap. At most a team can sign up to two without having either player take a discounted contract.



- With the top paid players of course they will be taking the most shots as well and putting up more of the statistics



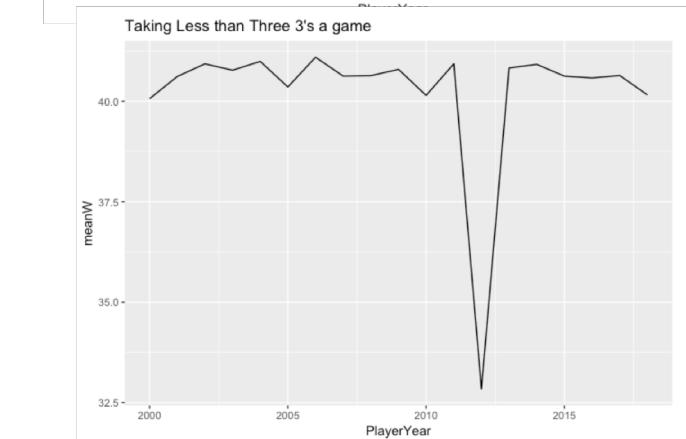
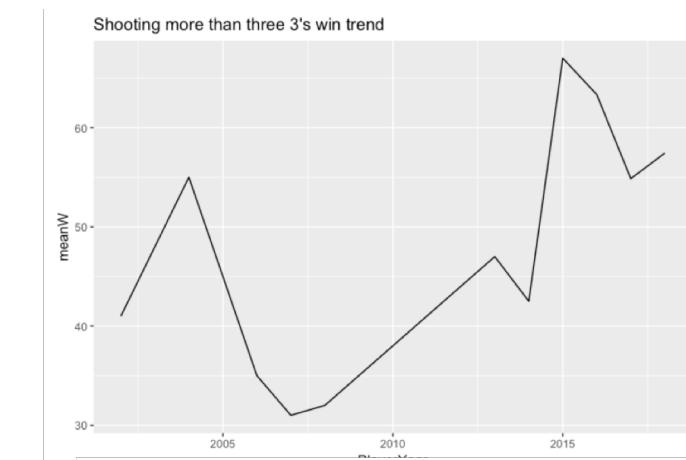
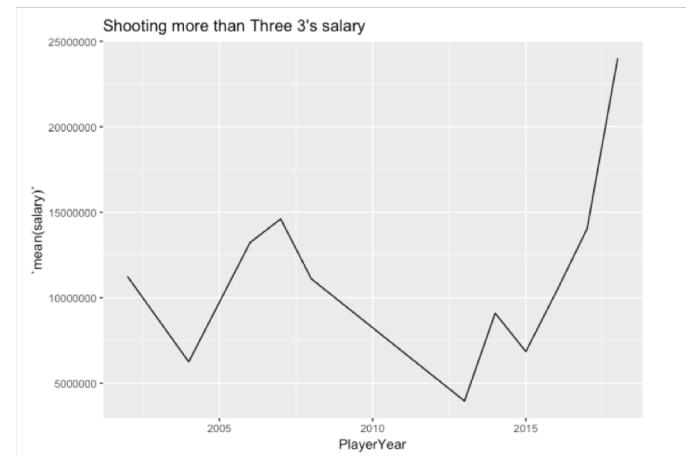
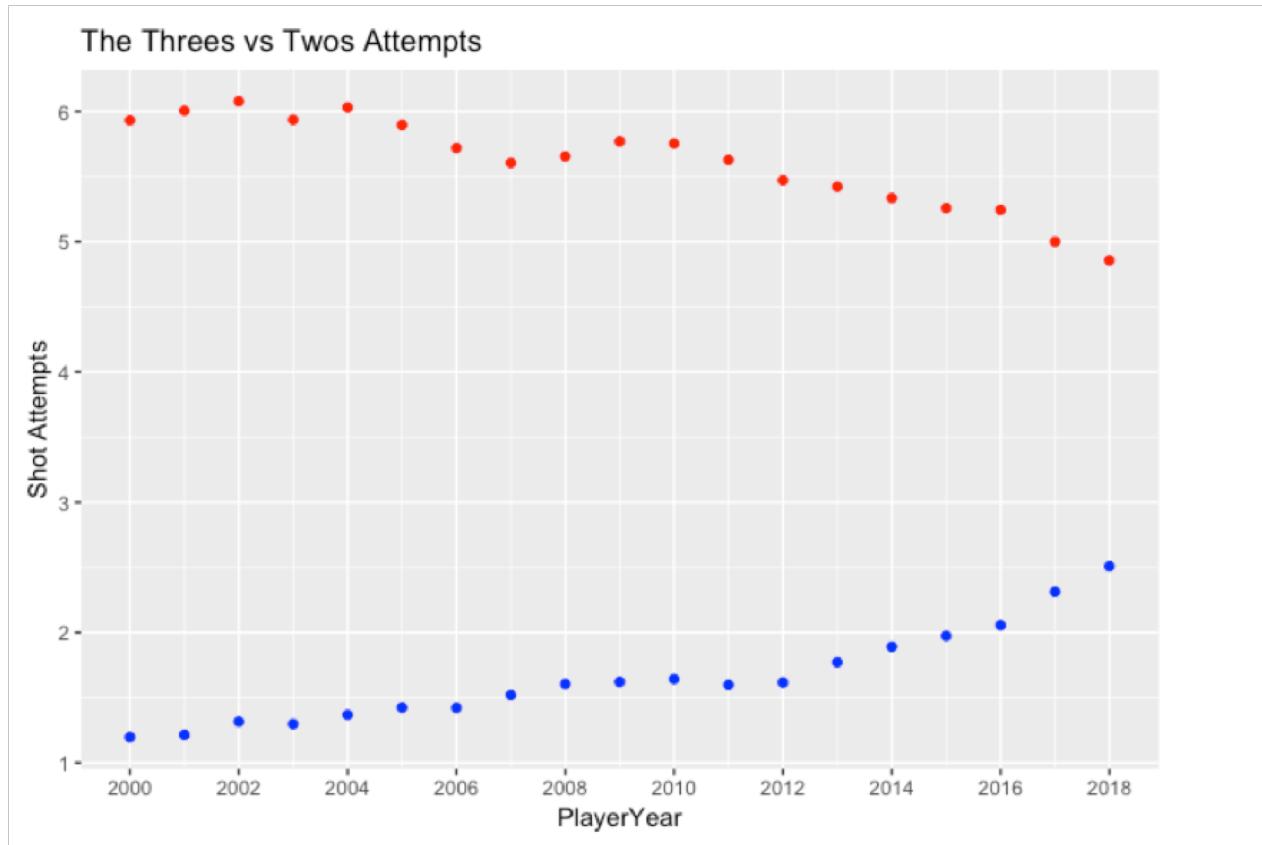
NBA TRENDS

Full code here:

https://github.com/jvhuang1786/capstoneproject/blob/master/nba_data_story.html

- 3 point shot
- Player height
- Plus Minus and player analytics

More 3's the better?



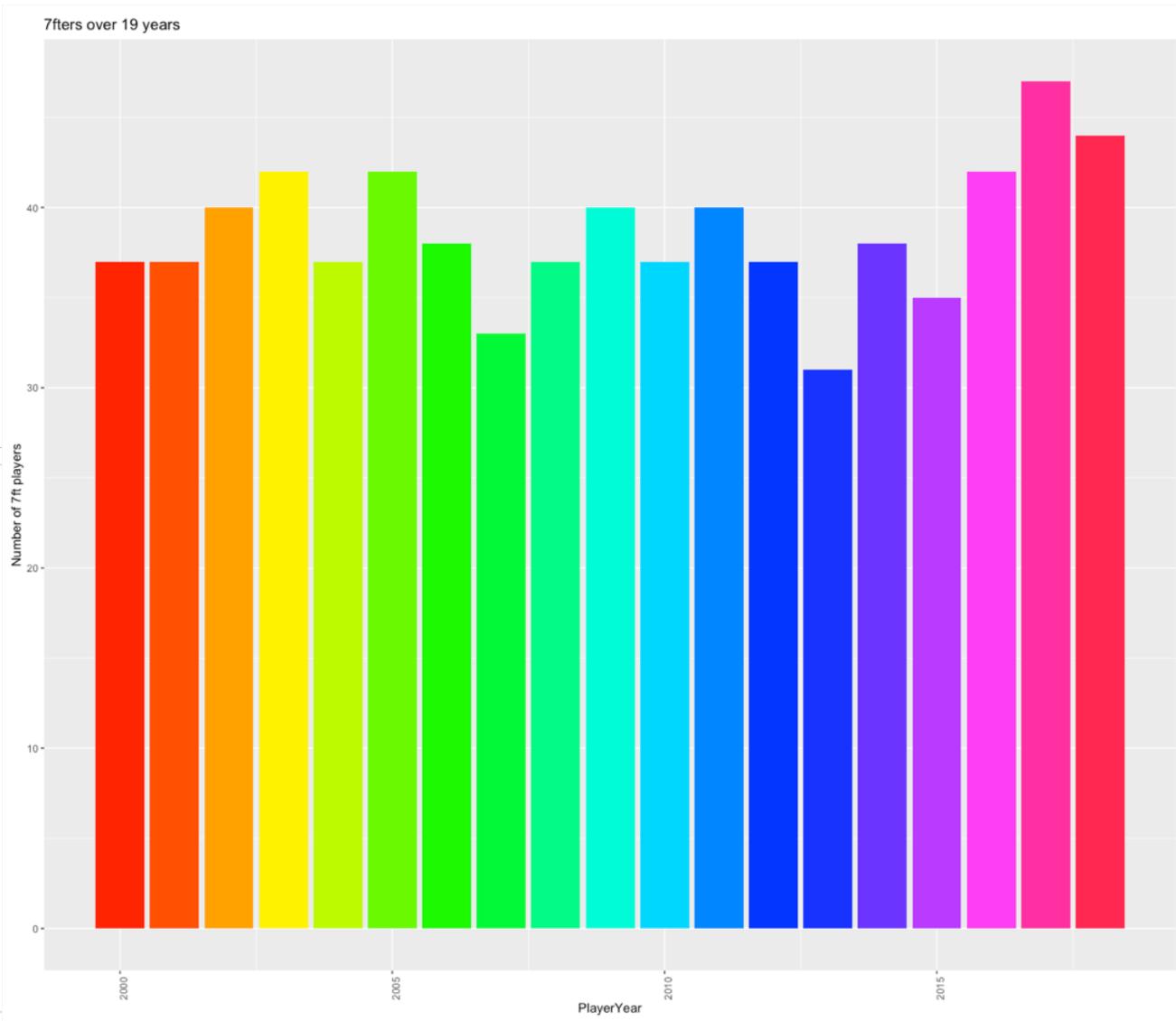
- Huge trend in more 3 attempts over the years with decrease in 2 attempts.
- Salary increase in 2015, need more years to see if 3 pointers and salary has stronger correlation
- Having players who make more than three 3's past 2015, 50 or more wins.
- Players less than 3 threes average 40 win team



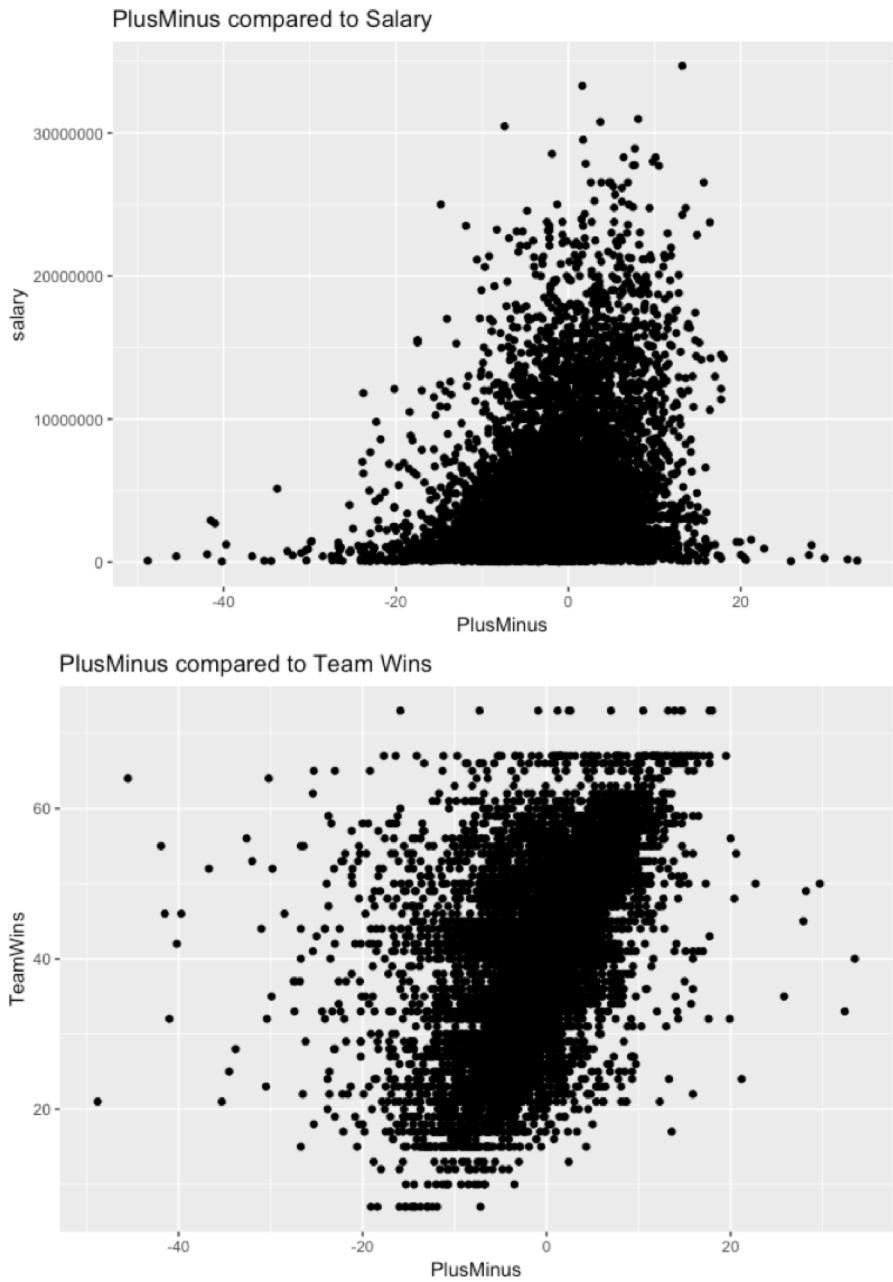
Death of the 7fters?

- Average height overall did not change that much over 19 seasons
- Number of 7fters actually increased still integral part of the game.

```
##   PlayerYear    height    salary
## 1      2000 200.9356 2880067
## 2      2001 200.9545 3398885
## 3      2002 201.5609 3487920
## 4      2003 201.6292 3783693
## 5      2004 201.3472 3817195
## 6      2005 201.4734 3911320
## 7      2006 201.0049 4091303
## 8      2007 200.6675 4140794
## 9      2008 200.8610 4574357
## 10     2009 201.2897 4715669
## 11     2010 200.9824 4829387
## 12     2011 201.3957 4560710
## 13     2012 200.9750 4348460
## 14     2013 200.8822 4447435
## 15     2014 200.9120 4402763
## 16     2015 200.8978 4430783
## 17     2016 201.1982 5097982
## 18     2017 201.1186 6218880
## 19     2018 200.5735 6450838
```



Plus Minus misconception



##	name	salary	PlusMinus	PlayerYear
## 1	Troy Daniels	158587	20.6	2014
## 2	Draymond Green	14260870	18.0	2016
## 3	Ben Wallace	14500000	17.7	2009
## 4	Stephen Curry	11370786	17.7	2016
## 5	Stephen Curry	12112359	17.7	2017
## 6	Andrew Bogut	12972973	17.0	2015
## 7	Tim Duncan	14260641	16.5	2005
## 8	Kevin Garnett	23750000	16.4	2008
## 9	Stephen Curry	10629213	16.4	2015
## 10	Draymond Green	915243	16.1	2015
## 11	Manu Ginobili	6603500	15.9	2005
## 12	Kyle Lowry	1011720	15.9	2007
## 13	Kevin Durant	26540100	15.7	2017
## 14	Zaza Pachulia	2898000	15.7	2017
## 15	Lamar Odom	14148596	15.4	2009
## 16	Klay Thompson	3075880	15.3	2015
## 17	Draymond Green	15330435	15.2	2017
## 18	Delonte West	3850000	15.0	2009
## 19	Chris Paul	22868827	14.9	2017
## 20	LeBron James	14410581	14.8	2009

- Plus minus is an advanced player statistic indicating if the player with the other 4 people score more than the opposing 5.
 - Problem is a bad player could be with 4 good players and outscore the second unit and get a huge positive plus minus
 - Need to look at the rotation that teams are playing to truly determine positive or negative points scored for players

