

A Study on NBA: What was changed by the rule?

Junyu Chen, Jinghong Zou, Xinyi Wang

1. Introduction and Background

A. The objective

Following the elimination of illegal defense and implementation of the defensive 3-second rule in 2001, NBA games have become more free flowing. From slow-moving games that heavily rely on dominant centers who consistently stationed themselves inside the paint, NBA have embraced the rise of “small ball,” which sacrifices height and low-post tactics in favor of a line-up of smaller players for speed and scoring from the three point line.

These two rule modifications led to revolutionary changes to the way of scoring and execution of game strategies. Shaquille O’Neal, a super star center who was publicly thought to be the main driving force for the rule change, is one focus of our exploration and analysis. How was he affected by this rule change? Did he earn fewer blocks since he could not stay inside the paint longer than three seconds?

In addition, other players could also be affected by these two rule changes. How did traditional centers adapt to these two new rules? How did players of other positions evolve? Fascinated by the impact of these two rule changes, we will analyze the statistics from the scope of the whole NBA league to specific players to investigate the gravity of impact of these two rule changes in breadth and depth.

We seek to explore how stats and performance of NBA players have changed due to the two rule changes in 2001, which will further provide us insights of changes in the dynamic of the game.

B. Data collection & general information

The data set that we will examine, “NBA Players stats since 1950,” was posted by Omri Goldstein on Kaggle. This data-set consists of aggregate statistics for individual players over 67 NBA seasons. The data-set contains variables from basic box-score attributes such as points, assists, rebounds etc., to more advanced features such as PER (player efficiency rating) and WS(Win share).

Among the three data sets given in “NBA Players stats since 1950,” Players.csv, Seasons_Stats.csv, and player_data.csv, we will only examine Seasons_Stats.csv because it has a holistic list of variables that are relevant to our studies. With the provided variables in Seasons_Stats.csv, we are able to generate substantial and in-depth analysis.

C. Origin of the data source

This data set on Kaggle is originally scraped from Basketball Reference, and Basketball-Reference uses statistics from Sportradar, the official statistics partner of the NBA.

D. Background knowledge of NBA basketball

a. Positions in basketball

Typically, there are five positions on the court, and each of which is assigned with a number for categorization: the point guard (PG) or 1, the shooting guard (SG) or 2, the small forward (SF) or 3, the power forward (PF) or 4, and the center (C) or 5.

Traditionally, point guard is usually the “leader” of the team, who facilitates the set-up of offence and pace of the game. Shooting guards are usually the leading scorers in the team, who have a variety of offensive tools and are capable of shooting with longer range. Forwards are usually excellent finishers who often have a more aggressive approach to scoring and who are known to make cuts to open spaces. The power forwards and centers are known for rebounders and shot-blockers. They usually score with low-post moves or receive passes to take close-range shots.

Positions can be very flexible as players usually play one or multiple positions. Due to the rise of “small ball,” the designated roles of different positions also became more fluid as players started to use different offensive tactics to make themselves more versatile to game situations.

b. Zone Defense & Defensive 3-second rule

Zone defense Zone defense is a basketball defensive formation in which a coach assigns each player to cover a specific area of the court. In such formation, a defender begins guarding an opponent when the opponent enters the defender’s designated zone. However, when the offensive player leaves the defender’s zone, the defender continues to guard their zone rather than following the offensive player as in a man-to-man defense.

Man-to-Man defense In basketball, man-to-man defense is a defensive formation in which a coach assigns each player on the line-up a specific offensive player to follow and guard on the court.

Defensive 3-second rule The defensive 3-second rule simply states that a defender cannot be camped inside the paint area for longer than 3 seconds. The team committing a defensive three-second violation will be given a team technical foul along with a free throw, after which the offensive team retains possession on the sideline.

The defensive three-second count is suspended when:

- (1) A player is in the act of shooting
- (2) There is a loss of team control
- (3) The defender is actively guarding an opponent,
- (4) The defender completely clears the 16-foot lane
- (5) It is imminent the defender will become legal.

E. Variables description

All variables listed are measured.

a. Identifying variables

Year: Year that the season occurred. Since the NBA season stretches over two calendar years, the year given is the last year for that season. For example, the year for the 1999-00 season would be 2000.

We will only manipulate the data sets from 1990 to 2017 for two primary reasons. Firstly, there are many missing values in the early years of the data set, as the statistical system and rules were not as developed. Secondly, we want to focus on the impact of the rule changes in 2001. There would be too much uncertainty with the correlation of the data if we trace back to the 1960s, providing that many other rules were implemented within that time span.

We will ignore all the data in 1999 since there was a lockout in that year. During a lockout, the value of all variables will be fewer than normal value since all players only played fifty games in that season.

Player: the name of the players, in the order of first name and last name.

Pos: The role that the player plays on the court. C is center, F is forward, G is guard, SG is shooting guard, SF is small forward, PF is power forward, PG is point guard. When there are two positions listed, the one in the former is the position that the player normally plays

b. Box Score Variables

AST: AST is the number of assists by a player.

BLK: BLK is the number of blocks by a player.

3PA: 3PA stands for 3-Point Field Goal Attempts

c. Advanced Variables

G: The number of matches played by this player in this season. We use this to measure per game value for different variables.

PER: PER stands for player efficiency rating, which is calculated by summing up a player's positive accomplishments, subtracting the negative accomplishments, which then returns a per-minute rating of a player's performance.

We will look at this variable because it is a direct indicator of the player's overall contribution, and we can analyze whether it is correlated with the position, and how the relationship has changed over time.

USG percentage: USG percentage represents usage percentage, which is an estimate of the percentage of team plays used by a player while he was on the floor.

WS.48: Win Shares Per 48 Minutes is an estimate of the number of wins contributed by the player per 48 minutes (league average is approximately 0.100). This includes credits to both defensive and offensive ratings.

d. Ignored variables or data

Identifying variables: Age: Age of the player is ignored because it is not relevant to our study of the two implemented rules.

Tm: Tm is the abbreviation of team and we will ignore it because we are more focused on performance of individual players of different positions.

Double Position: Since there are not many double-position players, Players with double positions can be ignored for simplicity.

Career statistics: The variables listed below are all ignored for two primary reasons: either they have no potential correlation with our topic or they are not relevant to the variables we have chosen. Since there are 55 columns in the data set, we need to narrow our focus to the most reflective and worth exploring variables. Moreover, many of the variables listed below, for instance, 2P percentage and 3P percentage, can be calculated using the raw data and given formula. Thus, we can create new columns if we discover important correlations with these variables.

GS(number of games that the player plays in the starting line-up), TS percentage(the percentage of true shooting), FTr (Free Throw rate), ORB percentage (Offensive Rebound percentage), DRB percentage (Defensive Rebound percentage), TRB percentage (Total Rebound percentage), STL percentage (stealing percentage), TOV percentage (Turnover percentage), Blakl (blank column), OBPM (offensive box plus/minus), DBPM (defensive box plus/minus), BPM (box plus/minus), VORP (value over replacement), FG (field goals), FGA (field goal attempts), FG percentage (field goal percentage), eFG percentage (Effective Field Goal Percentage), FT (Free throws), FTA (Free Throw Attempts), FT percentage (Free throw shot percentage), ORB (Offensive Rebounds), DRB (Defensive Rebounds), TRB (Total Rebounds), 3P percentage (3-Point Field Goal percentages), 2P percentage (2-Point Field Goal percentages), 3PAr(3-Point Attempt Rate), AST percentage (Assist Percentage), BLK percentage(Block Percentage), MP(Minute Played), 2P(2-Point Made), 2PA(2-Point Attempt), PTS(Points Scored)

2. Set up and Loading data

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.2      v dplyr   1.0.10
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1
```

```
## Warning: package 'dplyr' was built under R version 4.1.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

nba_1950 <- read.csv("Seasons_stats.csv")
```

Here we read the data we want to use for our project and name it “nba_1950.”

3. Cleaning the data

```
nba <- nba_1950 %>%
  select(Year, Player, Pos, G, MP, PER, WS, WS.48, FGA, X3P, X3PA, X2P, X2PA, AST, BLK, PTS, USG.) %>%
  filter(Year != 1999)
```

The 1999-2000 NBA season had a lockout due to salary cap issues, which forced teams to only play 50 games. We decided to delete the statistics of this year because many lockout-related factors, such as strikes and protests, might affect a player’s game performance and mindset. Secondly, this issue also reminds us to calculate per-game box scores instead of sum of box scores because total value can be affected by missed games, which does not reflect game performance accurately.

We clean the data by selecting only the variables that will contribute to our following exploration and analysis. Also, we delete all the data in year 1999 as mentioned above. Finally, this cleaned data is named “nba.”

```
nba_raw <- nba %>%
  filter(!grepl("-", Pos)) %>%
  drop_na(Year) %>%
  mutate(blk.game = BLK/G) %>%
  mutate(ast.game = AST/G) %>%
  mutate(pa3.game = X3PA/G) %>%
  mutate(pa2.game = X2PA/G) %>%
  mutate(mp.game = MP/G) %>%
  mutate(pat.game = FGA/G)
```

The second step of cleaning this data set includes eliminating all the rows with empty values or double-position player. In addition, new variables, such as block per game and 3-point attempts per game etc, are created for our further analysis.

4. Exploratory Data Analysis

For the graphs below, we want to plot box score statistics (including blocks per game, assists per game, 3-point attempts) against Year to explore whether performance and statistics of players have changed after the rule change. Most of the graphs below will be plotted within a ten year time-span: five years before 2001 and five years after for comparative analysis, except for the variable three points attempts, which we find interesting after initial plotting. Therefore, we plot the trend of three point attempts between year 1995 and 2010 for further analysis. In all plots, categorical variable Year is on x-axis while numeric variable box score stats will be on the y-axis.

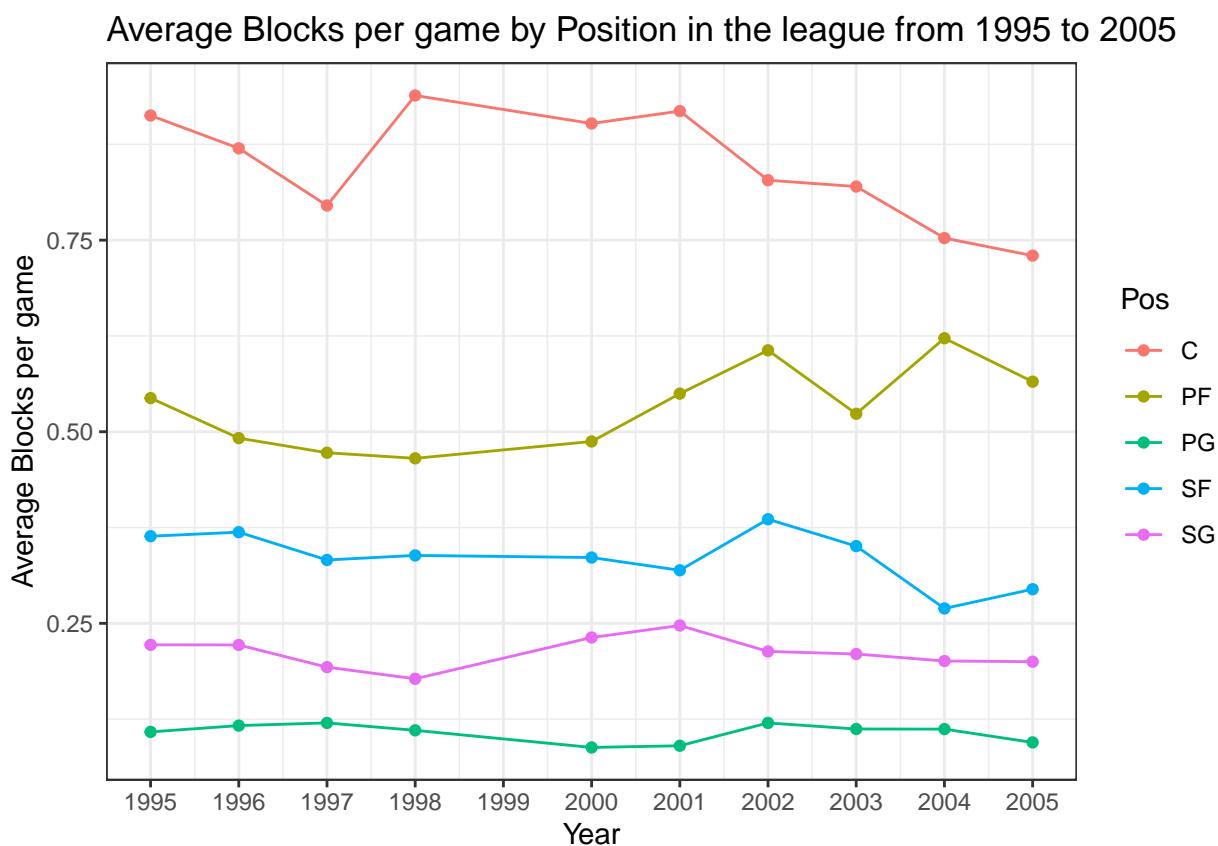
Expectations about the change of average blocks per game: Before the rule change, players can station themselves inside the paint to block opponents who drives into the paint to score. However, after the implementation of Defensive 3 second rule, players can no longer stay inside the paint for more than 3 seconds after 2001. Thus, it becomes reasonable for us to speculate that the average blocks per game for players would decrease. Since centers tend to be top shot-blockers of the game, we speculate that their statistics with regards to average blocks per game would decrease more dramatically than other positions.

Categorical variable Year is on x-axis while numeric variable will be the average block per game, on the y-axis.

```
nba_5yrs <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2005) %>%
  group_by(Year,Pos) %>%
  summarise(mean_blk = mean(blk.game, na.rm = TRUE))
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```

```
ggplot(nba_5yrs, aes(x = Year, y = mean_blk, color = Pos)) +
  geom_point() +
  geom_line(stat = "identity", position = "identity") +
  theme_bw() +
  ggtitle("Average Blocks per game by Position in the league from 1995 to 2005") +
  xlab("Year") + ylab("Average Blocks per game") +
  scale_x_continuous(breaks = seq(1995, 2005, by = 1))
```



According to the graph, no obvious trends can be seen. The trend-lines for PG,SF,SG have negligible fluctuation. The average blocks per game for centers tend to decrease after 2001, though the declining pattern is not very obvious. The average blocks per game for PF tend to increase slightly after year 2000. Overall, in terms of the whole NBA league, the association between average blocks per game and rule change is not strong.

Another variable that we have chosen to analyze is the assist per game. Due to the establishment of zone-defense, which discourages plays and low-post moves, it becomes harder for players to get through the defense on their own. Therefore, we can reasonably speculate that teams will develop more game tactics that involve teams plays and passing. Consequently, the average assists per game would increase.

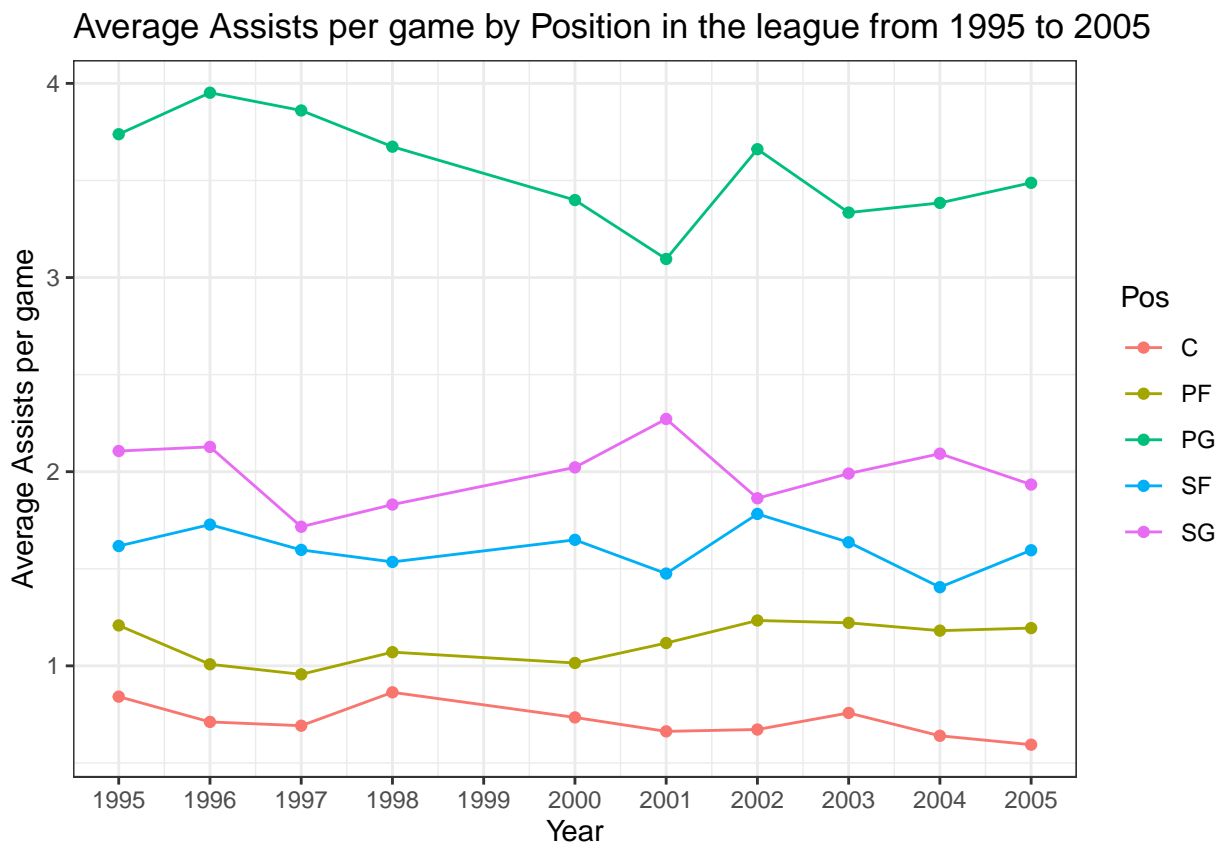
Categorical variable Year is on x-axis while numeric variable, the average assist per game, will be on the y-axis.

```
nba_5yrs <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2005) %>%
```

```
group_by(Year, Pos) %>%
  summarise(mean_ast = mean(ast.game, na.rm = TRUE))
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```

```
ggplot(nba_5yrs, aes(Year, mean_ast, color = Pos)) +
  geom_line(stat = "identity", position = "identity") + geom_point() +
  theme_bw() +
  ggtitle("Average Assists per game by Position in the league from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Assists per game") +
  scale_x_continuous(breaks = seq(1995, 2005, by = 1))
```



Unfortunately, there is no general trend shown in the graph. According to the graph, the average assists per game for all positions tend to have negligible fluctuations.

The average three-point attempts per game is another important variable that we expect to be affected by the implementation of two rules. Interestingly, the effect of these two rules actually counter each other. For zone defense, after zone-defense was allowed, players would find it more challenging to drive into the lane, thus they would seek more opportunities to score outside the perimeter, which is three pointers. Thus, average three-point attempts would increase. However, the implementation of Defensive 3-second rule frees up the lane for more scoring inside the paint, which would counter the effect of zone defense. The graph below will allow us to explore the gravity of impact of these two rules. If average three-point attempts increase, the effect of zone defense is greater. If average three-point attempts decrease, the effect of Defensive 3 second rule is greater.

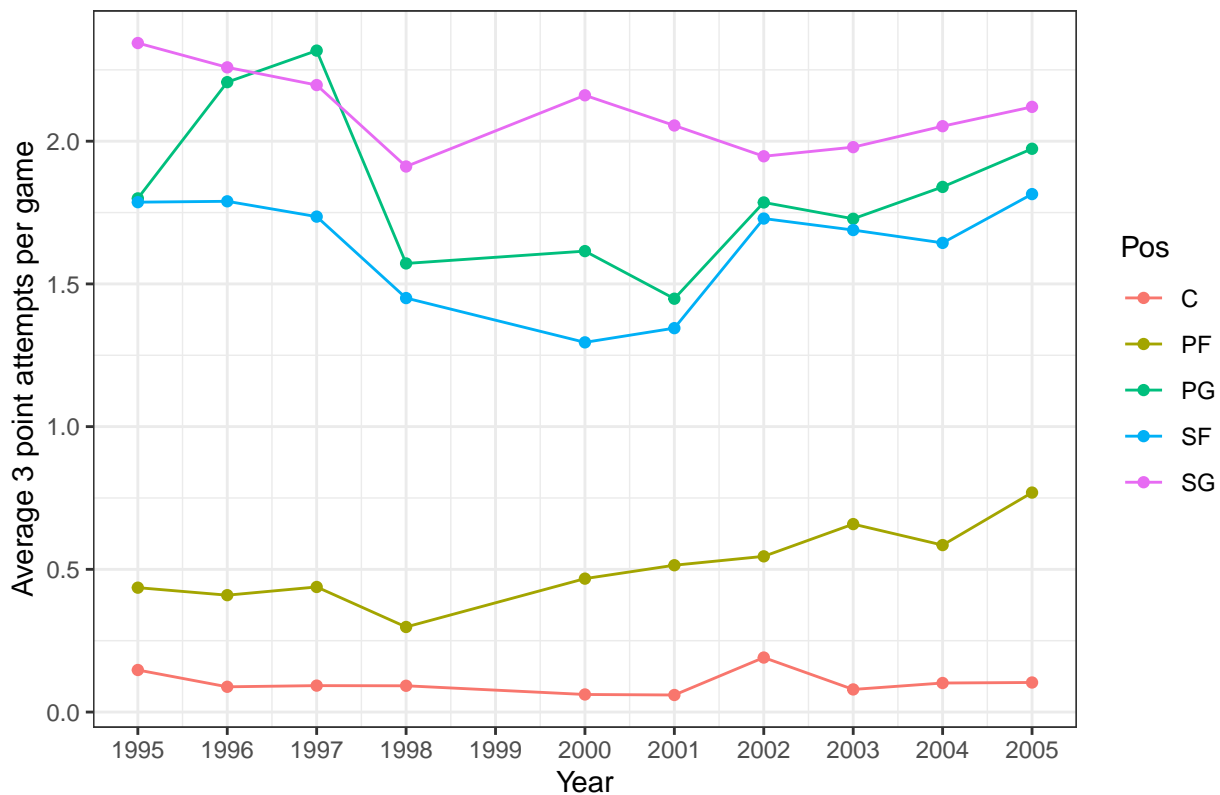
Categorical variable Year is on x-axis while numeric variable, the average three-point attempts per game, will be on the y-axis.

```
nba_5yrs <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2005) %>%
  group_by(Year,Pos) %>%
  summarise(mean_3pa = mean(pa3.game, na.rm = TRUE))
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```

```
ggplot(nba_5yrs, aes(Year, mean_3pa, color = Pos)) +
  geom_line(stat = "identity", position = "identity") + geom_point() +
  theme_bw() +
  ggtitle("Average 3pt attempt per game by Position in the league from 1995 to 2005") +
  xlab("Year") +
  ylab("Average 3 point attempts per game") +
  scale_x_continuous(breaks = seq(1995, 2005, by = 1))
```

Average 3pt attempt per game by Position in the league from 1995 to 2005

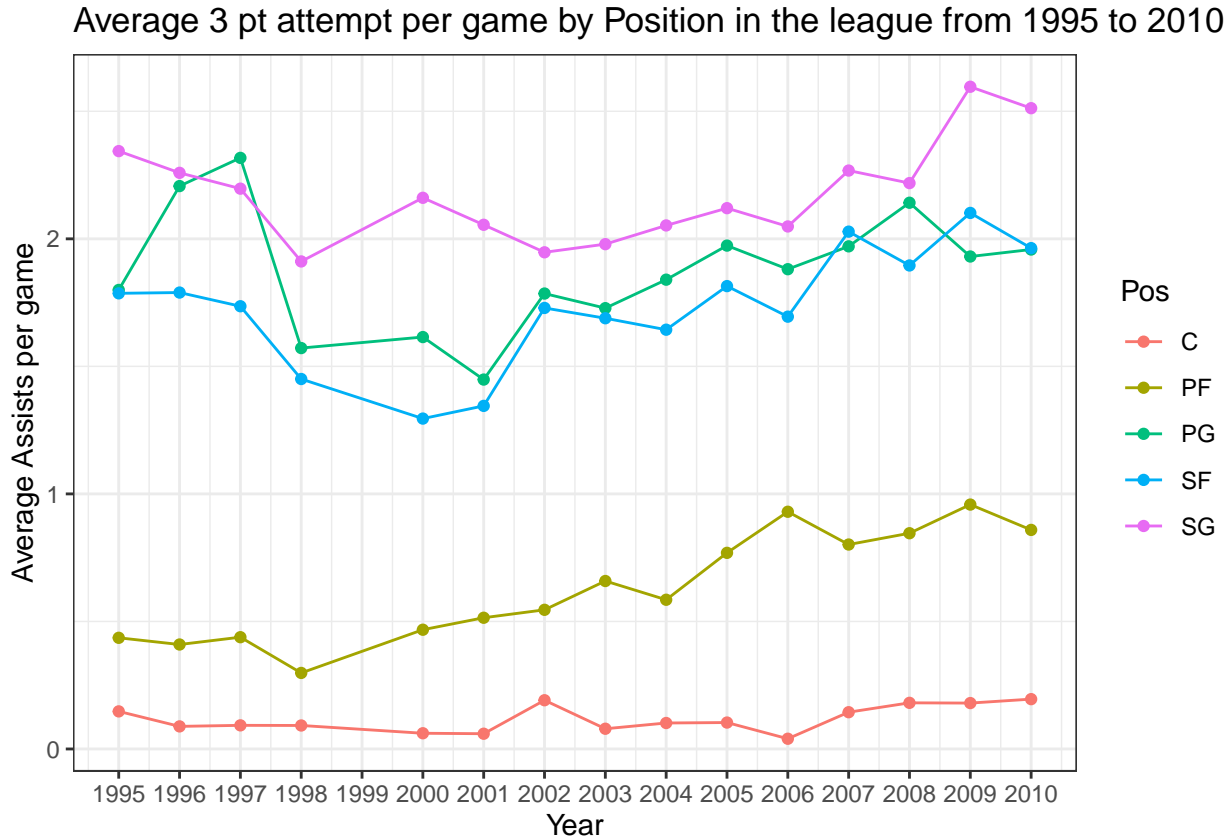


In this graph, we see potential rise of average three-point attempts for PF,PG, and SF after the year of rule change, 2001. However, the trend-lines for SG and C tend to be stagnant. To have more in-depth analysis, we expand our time-span to 10 years after 2001 to see if this trend continues to maintain or become more obvious in a larger time scope.

```
nba_10yrs <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2010) %>%
  group_by(Year,Pos) %>%
  summarise(mean_3pa = mean(pa3.game, na.rm = TRUE))
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```

```
ggplot(nba_10yrs, aes(Year, mean_3pa, color = Pos)) +
  geom_line(stat = "identity", position = "identity") + geom_point() +
  theme_bw() +
  ggtitle("Average 3 pt attempt per game by Position in the league from 1995 to 2010") +
  xlab("Year") +
  ylab("Average Assists per game") +
  scale_x_continuous(breaks = seq(1995, 2010, by = 1))
```



In this graph, we can see that the trend is stable between 1996 to 2001. However, after the rule change, the trend-lines start to rise gradually and reach an unprecedented peak for all positions except C. Hence, we believe there is an association between the rule change and 3-point attempts. Moreover, this potentially indicates that the impact of zone defense is much greater than the impact of defensive three second rule.

There are a few possible explanations. First, after the rule change, teams realize that zone defense is very effective against dominant players. Thus, this strategy becomes more widely used. To counter zone defense, players of most positions started to shoot more threes. Second, the effect of defensive 3-second rule is not strong because centers are not entirely prohibited from staying inside the paint. They can stay very close to the lines of the paint and still block the opponents whenever opponents attack the paint.

5. Data Modeling

The trend without obvious tendency as we observed above may give us a false understanding of how the rule changed the players. This is because most of the players did not have enough time to play on the court. Consequently, it is reasonable that they have mediocre and low performing stats, for example zero blocks and zero assist. Therefore, to fully analyze the effect of rule change, we should also look at the top players and their stats before and after rule change.

In the following, we focus our attention on variables block per game, assist per game, and 3-point attempts per game. At the same time, we analyzed five positions separately and tried to observe potential effect on different positions. Among the three variables that we have explored, we find that the 3-points attempts per game among

top players shows some changes before and after rule change. Therefore, we go on to explore more top players in three point attempts.

```
highest_3pa_pre_5<-nba_raw%>%
  filter(Year>=1995&Year<=2000&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(5)

highest_3pa_post_5<-nba_raw%>%
  filter(Year>=2001&Year<=2005&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(5)

colnames(highest_3pa_pre_5)[2]="3pt attempt/game"
colnames(highest_3pa_post_5)[2]="3pt attempt/game"
cbind(highest_3pa_pre_5,highest_3pa_post_5)
```

| ## | Year | 3pt attempt/game | Player | Year | 3pt attempt/game | Player |
|------|------|------------------|----------------|------|------------------|------------------|
| ## 1 | 1997 | 5.253165 | Terry Mills | 2002 | 7.962963 | Antoine Walker |
| ## 2 | 1997 | 4.709091 | LaPhonso Ellis | 2003 | 7.461538 | Antoine Walker |
| ## 3 | 1998 | 4.524590 | Chuck Person | 2001 | 7.444444 | Antoine Walker |
| ## 4 | 1996 | 4.426829 | Danny Ferry | 2004 | 6.000000 | Rasheed Wallace |
| ## 5 | 1996 | 4.426829 | Sam Perkins | 2005 | 5.584615 | Donyell Marshall |

From 1995 to 2000, the range of top three point attempts per game in one year is [4.43,5.25]. From 2001 to 2005, the range of top three point attempts per game in one year is [5.58,7.96]. There is an obvious difference between these two sets of data. Power forwards became more likely to shoot 3-point balls. Hence, we can say that there is an association between rule change and power forwards' willingness to shoot 3-point balls.

Seeing the change in 3-point attempts among the top 5 power forwards, we look into more top power forwards by selecting the top 30 of them. After visualizing the data of these 30 players with a qqnorm plot, we do a t-test on the data before and after rule change using average 3-point attempts as an effect size. Our null hypothesis is that the rule change does not affect the mean of three point attempts among the top 30 players in the league.

```
highest_3pa_pre<-nba_raw%>%
  filter(Year>=1995&Year<=2000&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(30)

highest_3pa_post<-nba_raw%>%
  filter(Year>=2001&Year<=2005&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(30)

colnames(highest_3pa_pre)[2]="Three point attempts Per Game"
colnames(highest_3pa_post)[2]="Three point attempts Per Game"
cbind(highest_3pa_pre,highest_3pa_post)
```

| ## | Year | Three point attempts Per Game | Player | Year |
|------|------|-------------------------------|-------------|------|
| ## 1 | 1997 | 5.253165 | Terry Mills | 2002 |

| | | | | |
|-------|----------------------|----------|---------------------|------|
| ## 2 | 1997 | 4.709091 | LaPhonso Ellis | 2003 |
| ## 3 | 1998 | 4.524590 | Chuck Person | 2001 |
| ## 4 | 1996 | 4.426829 | Danny Ferry | 2004 |
| ## 5 | 1996 | 4.426829 | Sam Perkins | 2005 |
| ## 6 | 1995 | 3.958333 | Terry Mills | 2005 |
| ## 7 | 1997 | 3.867925 | Charles Barkley* | 2005 |
| ## 8 | 2000 | 3.731707 | Dirk Nowitzki | 2003 |
| ## 9 | 1998 | 3.612903 | Keith Van | 2002 |
| ## 10 | 1998 | 3.560976 | Antoine Walker | 2004 |
| ## 11 | 1995 | 3.546875 | Robert Horry | 2003 |
| ## 12 | 2000 | 3.475610 | Antoine Walker | 2001 |
| ## 13 | 1997 | 3.463415 | Danny Ferry | 2004 |
| ## 14 | 1997 | 3.415094 | Henry James | 2004 |
| ## 15 | 1995 | 3.220588 | Charles Barkley* | 2005 |
| ## 16 | 2000 | 3.195122 | Rodney Rogers | 2005 |
| ## 17 | 2000 | 3.160714 | Matt Bullard | 2003 |
| ## 18 | 2000 | 2.951220 | Terry Mills | 2003 |
| ## 19 | 2000 | 2.850000 | Keith Van | 2003 |
| ## 20 | 1995 | 2.841463 | Danny Ferry | 2002 |
| ## 21 | 1998 | 2.740741 | Sam Perkins | 2004 |
| ## 22 | 2000 | 2.691358 | Sam Perkins | 2004 |
| ## 23 | 1995 | 2.685185 | Chris Webber | 2004 |
| ## 24 | 1995 | 2.592593 | Larry Johnson | 2002 |
| ## 25 | 1997 | 2.577465 | Matt Bullard | 2002 |
| ## 26 | 1995 | 2.540984 | Brad Lohaus | 2004 |
| ## 27 | 1996 | 2.524390 | Terry Mills | 2001 |
| ## 28 | 2000 | 2.485714 | Larry Johnson | 2004 |
| ## 29 | 1996 | 2.478261 | Brad Lohaus | 2001 |
| ## 30 | 1996 | 2.464789 | Charles Barkley* | 2001 |
| ## | Three point attempts | Per Game | Player | |
| ## 1 | | 7.962963 | Antoine Walker | |
| ## 2 | | 7.461538 | Antoine Walker | |
| ## 3 | | 7.444444 | Antoine Walker | |
| ## 4 | | 6.000000 | Rasheed Wallace | |
| ## 5 | | 5.584615 | Donyell Marshall | |
| ## 6 | | 5.222222 | Vladimir Radmanovic | |
| ## 7 | | 5.056604 | Antoine Walker | |
| ## 8 | | 5.024691 | Pat Garrity | |
| ## 9 | | 4.950000 | Pat Garrity | |
| ## 10 | | 4.896104 | Vladimir Radmanovic | |
| ## 11 | | 4.875000 | Dirk Nowitzki | |
| ## 12 | | 4.756098 | Dirk Nowitzki | |
| ## 13 | | 4.753247 | Walter McCarty | |
| ## 14 | | 4.515152 | Donyell Marshall | |
| ## 15 | | 4.428571 | Antoine Walker | |
| ## 16 | | 4.209877 | Shawn Marion | |
| ## 17 | | 4.174603 | Toni Kukoc | |
| ## 18 | | 4.148649 | Rasheed Wallace | |
| ## 19 | | 4.069444 | Vladimir Radmanovic | |
| ## 20 | | 4.012658 | Rasheed Wallace | |
| ## 21 | | 3.963415 | Donyell Marshall | |
| ## 22 | | 3.844444 | Rasheed Wallace | |
| ## 23 | | 3.829268 | Clifford Robinson | |
| ## 24 | | 3.800000 | Clifford Robinson | |
| ## 25 | | 3.739726 | Eddie Griffin | |
| ## 26 | | 3.719512 | Antoine Walker | |
| ## 27 | | 3.653846 | Rashard Lewis | |
| ## 28 | | 3.647059 | Rasheed Wallace | |
| ## 29 | | 3.491803 | Matt Bullard | |

```

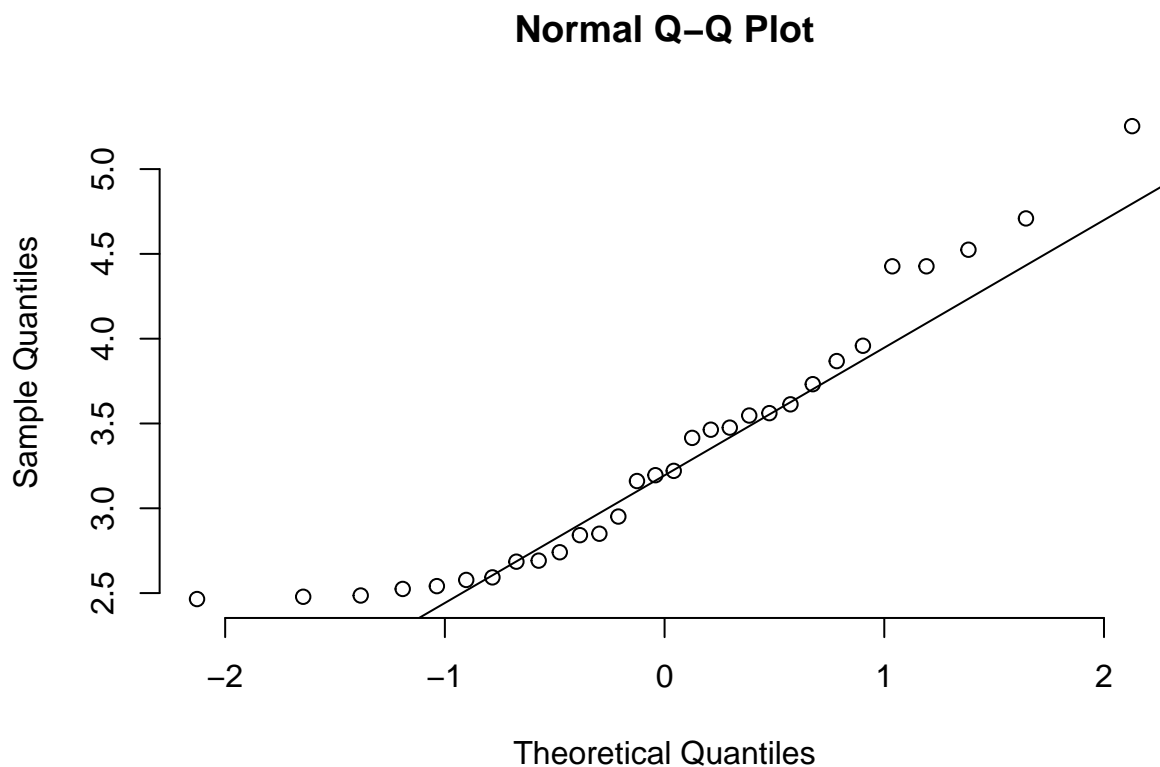
highest_3pa_pre<-nba_raw%>%
  filter(Year>=1995&Year<=2000&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(30)

highest_3pa_post<-nba_raw%>%
  filter(Year>=2001&Year<=2005&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(30)

combined <- c(highest_3pa_pre$pa3.game, highest_3pa_post$pa3.game)

qqnorm(highest_3pa_pre$pa3.game, frame = FALSE)
qqline(highest_3pa_pre$pa3.game)

```

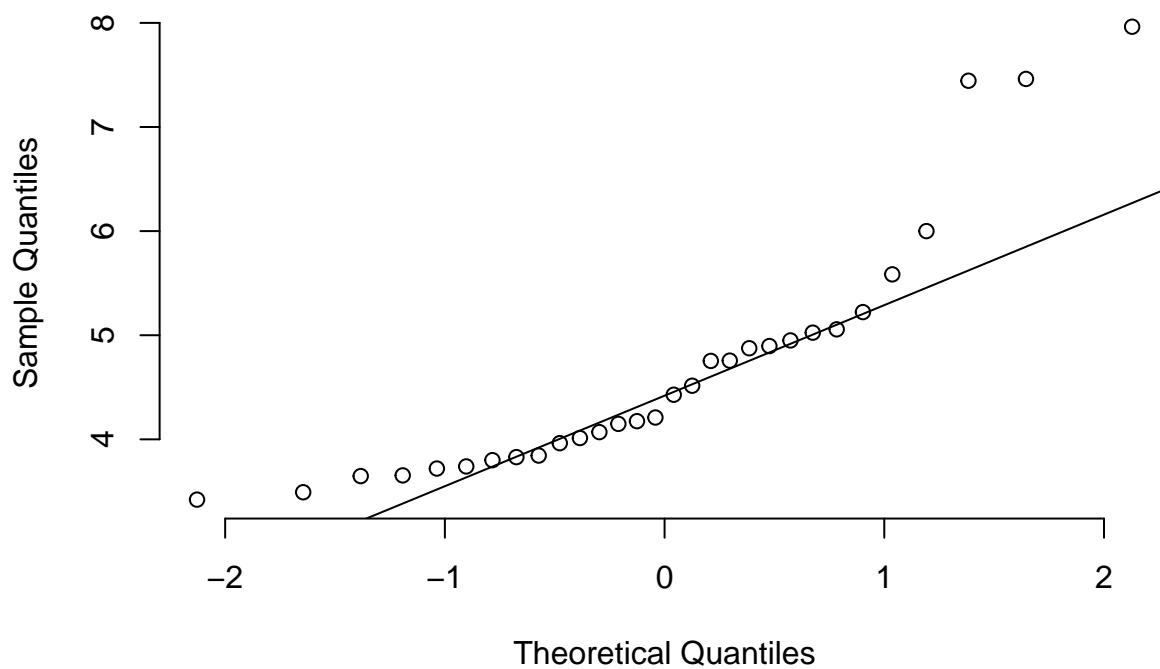


```

qqnorm(highest_3pa_post$pa3.game, frame = FALSE)
qqline(highest_3pa_post$pa3.game)

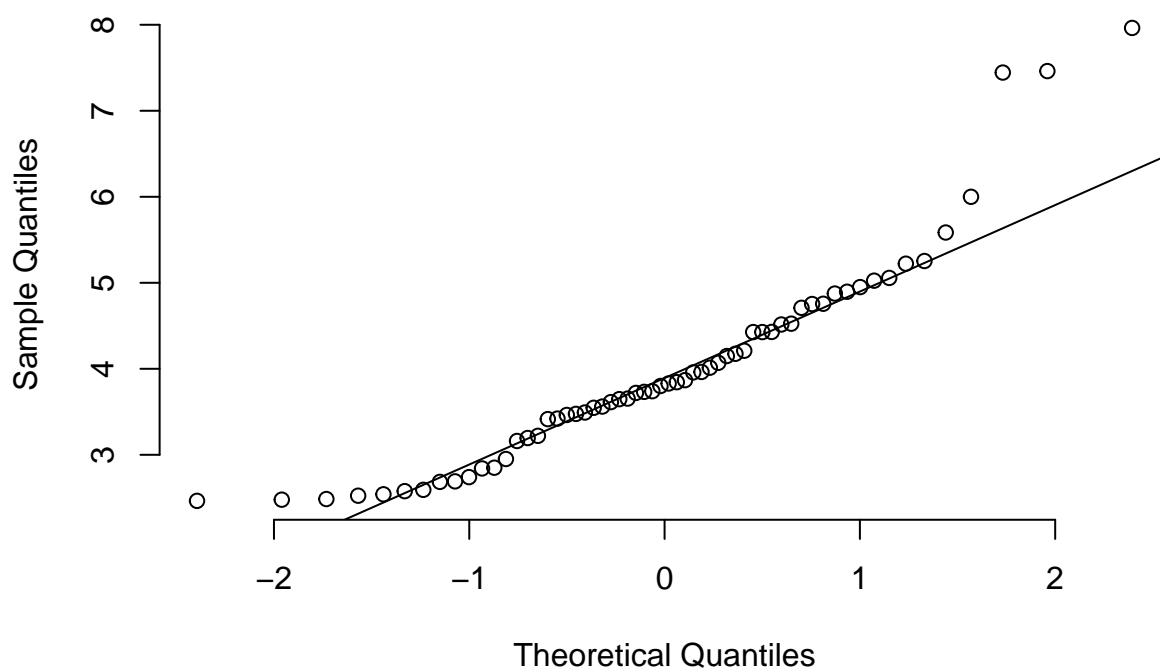
```

Normal Q-Q Plot



```
qqnorm(combined, frame = FALSE)  
qqline(combined)
```

Normal Q-Q Plot



The plots above show that the distribution of 3-point attempts before and after the rule change for the top 30 power forwards in the league is approximately normal. Therefore, we continue to do the t-test.

```
highest_3pa_pre<-nba_raw%>%
  filter(Year>=1995&Year<=2000&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(30)

highest_3pa_post<-nba_raw%>%
  filter(Year>=2001&Year<=2005&Pos=="PF")%>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))%>%
  head(30)

t_test <- t.test(highest_3pa_pre$pa3.game, highest_3pa_post$pa3.game)
t_test
```

```
##
## Welch Two Sample t-test
##
## data: highest_3pa_pre$pa3.game and highest_3pa_post$pa3.game
## t = -5.2678, df = 49.542, p-value = 2.993e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.8732740 -0.8389046
## sample estimates:
## mean of x mean of y
## 3.332464 4.688554
```

From the t test, we get a p value of 1.188e-05, which allows us to reasonably reject our null hypothesis that there is no difference between the average value of 3-point attempts before and after the rule change among top power forwards by random chance. Also, the 95 percent confidence interval that does not include zero provides further evidence for our rejection of the null hypothesis. In conclusion, our alternative hypothesis that there is statistically significant difference in 3-point attempts before and after the rule change among the top 30 power forwards is supported.

For other variables, there is no obvious change before and after rule change. Several sample tables are attached below. Hence, other than the 3-point attempt for power forward, we can not say that rule change affected other variables.

```
highest_block_pre<-nba_raw%>%
  filter(Year>=1995&Year<=2000&Pos=="PG")%>%
  select(Year,blk.game,Player)%>%
  arrange(desc(blk.game))%>%
  head(5)

highest_block_post<-nba_raw%>%
  filter(Year>=2001&Year<=2005&Pos=="PG")%>%
  select(Year,blk.game,Player)%>%
  arrange(desc(blk.game))%>%
  head(5)

colnames(highest_block_pre)[2]="Block Per Game"
colnames(highest_block_post)[2]="Block Per Game"
a<-cbind(highest_block_pre, highest_block_post)

highest_ast_pre<-nba_raw%>%
  filter(Year>=1995&Year<=2000&Pos=="SF")%>%
  select(Year,ast.game,Player)%>%
```

| Year | Block Per Game | Player | Year | Block Per Game | Player |
|------|------------------------|-------------------|------|------------------------|--------------------|
| 1995 | 0.6625000 | Nate McMillan | 2002 | 1.0000000 | Randy Brown |
| 1997 | 0.5932203 | Anfernee Hardaway | 2002 | 0.5731707 | Baron Davis |
| 1998 | 0.5853659 | Ron Harper | 2004 | 0.5573770 | Dwyane Wade |
| 1996 | 0.5000000 | Anfernee Hardaway | 2001 | 0.5319149 | Ron Harper |
| 1996 | 0.5000000 | Terrence Rencher | 2001 | 0.5217391 | Felipe Lopez |
| Year | Assist Per Game | Player | Year | Assist Per Game | Player |
| 1997 | 7.287500 | Grant Hill | 2005 | 7.212500 | LeBron James |
| 1996 | 6.850000 | Grant Hill | 2001 | 6.250000 | Grant Hill |
| 1998 | 6.802469 | Grant Hill | 2001 | 6.235294 | Toni Kukoc |
| 1996 | 5.870130 | Scottie Pippen* | 2001 | 6.041667 | Jalen Rose |
| 1998 | 5.772727 | Scottie Pippen* | 2002 | 5.896552 | Lamar Odom |
| Year | 3-pt attempts Per Game | Player | Year | 3-pt attempts Per Game | Player |
| 1995 | 7.637500 | John Starks | 2005 | 7.987342 | Quentin Richardson |
| 1995 | 6.890625 | Vernon Maxwell | 2004 | 7.656716 | Tracy McGrady |
| 1997 | 6.617284 | Reggie Miller* | 2003 | 7.655172 | Ray Allen |
| 1997 | 6.428571 | Nick Anderson | 2002 | 7.652174 | Ray Allen |
| 1996 | 6.358025 | Mitch Richmond* | 2005 | 7.314286 | Jamal Crawford |

```

arrange(desc(ast.game))>%
  head(5)

highest_ast_post<-nba_raw%>%
  filter(Year>=2001&Year<=2005&Pos=="SF")>%
  select(Year,ast.game,Player)%>%
  arrange(desc(ast.game))>%
  head(5)

colnames(highest_ast_pre)[2]="Assist Per Game"
colnames(highest_ast_post)[2]="Assist Per Game"
b<-cbind(highest_ast_pre,highest_ast_post)

highest_3pa_pre<-nba_raw%>%
  filter(Year>=1995&Year<=2000&Pos=="SG")>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))>%
  head(5)

highest_3pa_post<-nba_raw%>%
  filter(Year>=2001&Year<=2005&Pos=="SG")>%
  select(Year,pa3.game,Player)%>%
  arrange(desc(pa3.game))>%
  head(5)

colnames(highest_3pa_pre)[2]="3-pt attempts Per Game"
colnames(highest_3pa_post)[2]="3-pt attempts Per Game"
c<-cbind(highest_3pa_pre,highest_3pa_post)

kable(list(a,b,c))

```

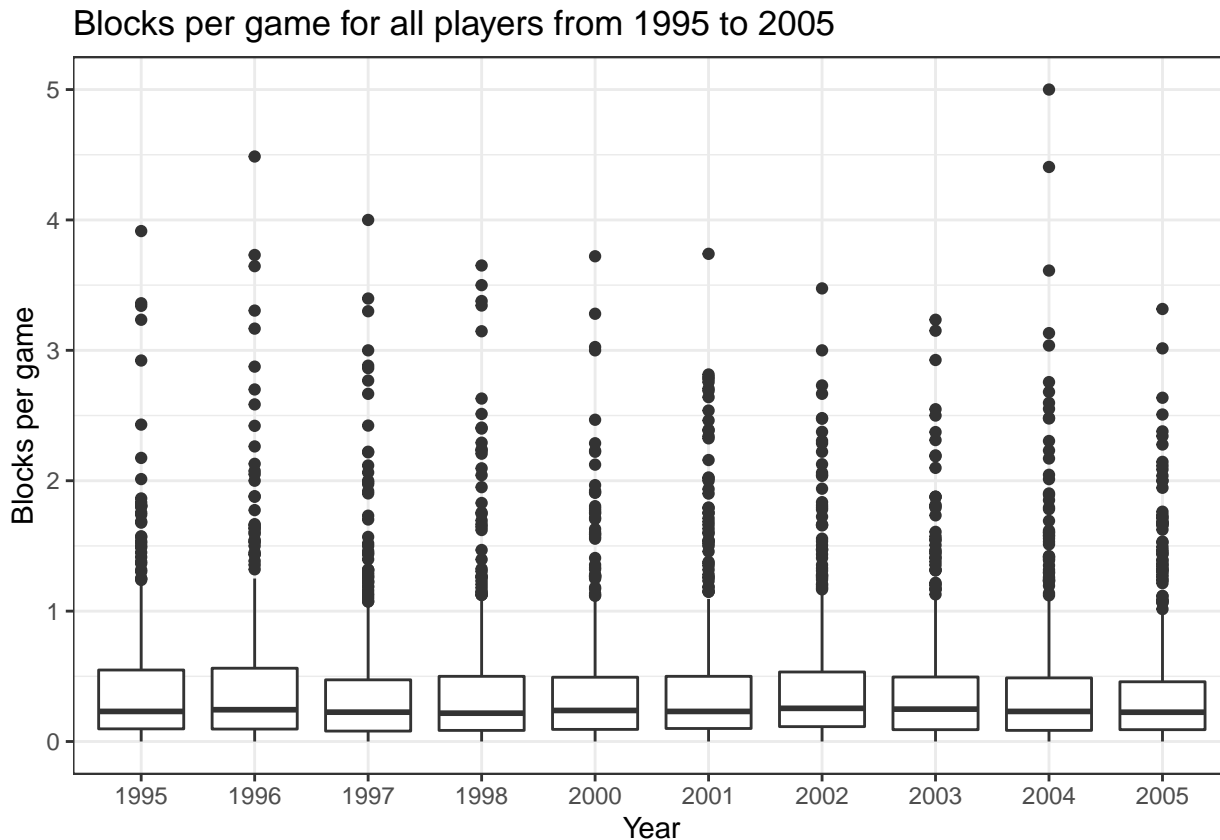
6. Selecting outliers

After analyzing the general trend in the whole league and top players of the league, we try to shift our focus to specific players. Other than Shaquille O'Neal, we will also analyze players who was not an outlier before the rule change but became an outlier after rule change. These players are the possible beneficiary of the rule change. Now,

we focus on three main variables of players: block per game, assist per game, and 3-point attempt per game. Three boxplots are made in the following.

```
nba_5yrs <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2005)

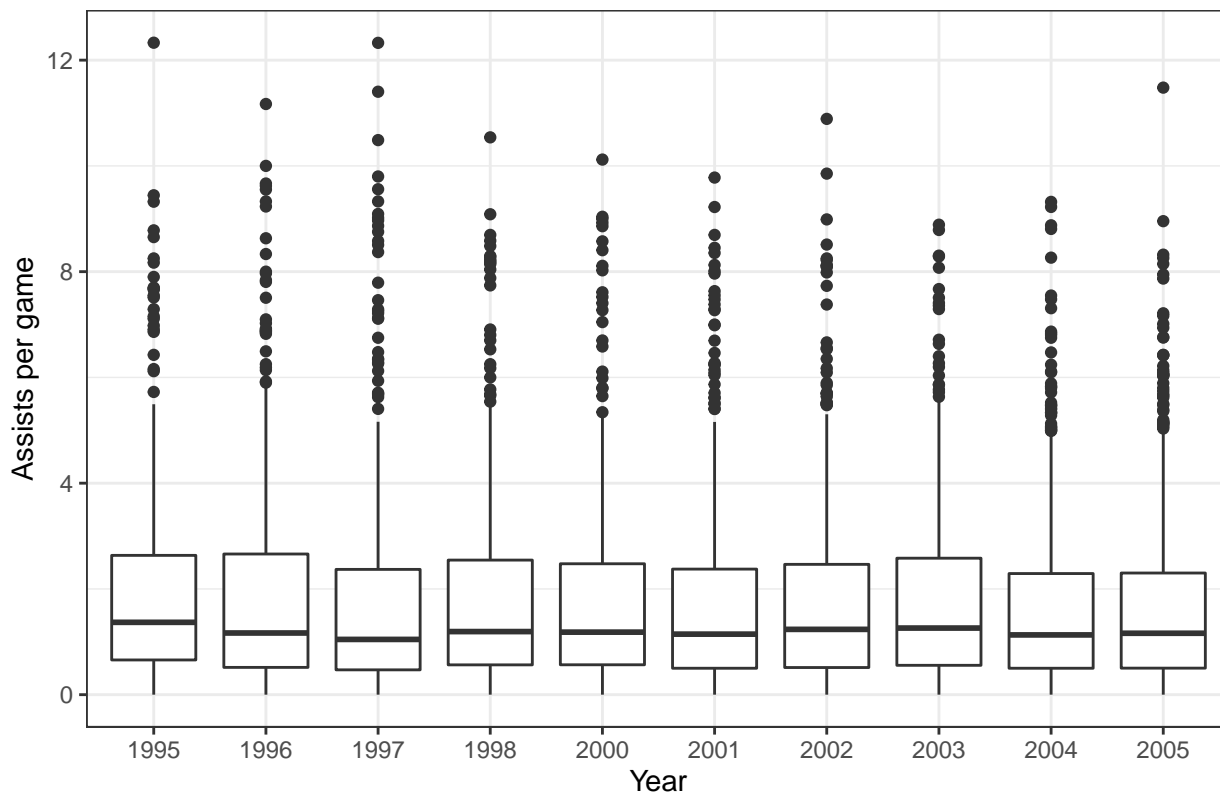
nba_5yrs$Year <- as.character(nba_5yrs$Year)
ggplot(nba_5yrs, aes(Year, blk.game)) +
  geom_boxplot() +
  theme_bw() +
  ggtitle("Blocks per game for all players from 1995 to 2005") +
  xlab("Year") + ylab("Blocks per game")
```



```
nba_5yrs <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2005)

nba_5yrs$Year <- as.character(nba_5yrs$Year)
ggplot(nba_5yrs, aes(Year, ast.game)) +
  geom_boxplot() +
  theme_bw() +
  ggtitle("Assists per game for all players from 1995 to 2005") +
  xlab("Year") +
  ylab("Assists per game")
```

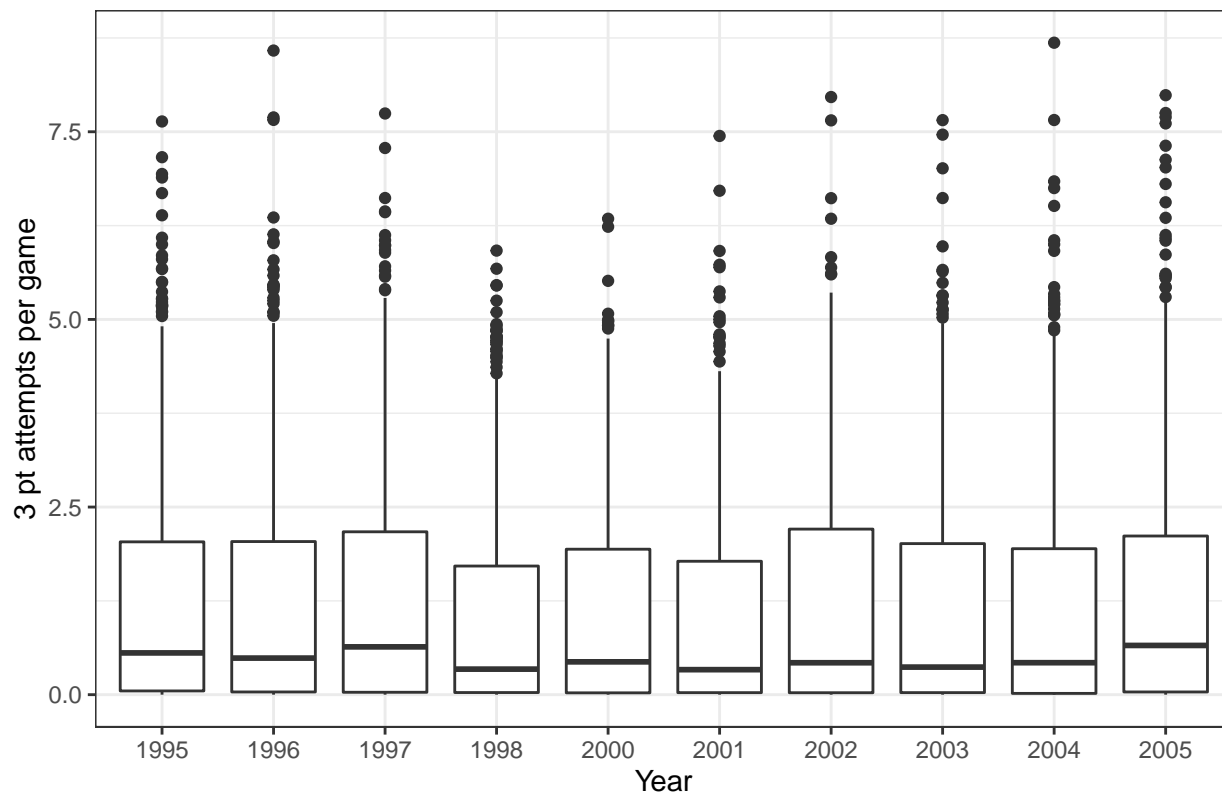
Assists per game for all players from 1995 to 2005



```
nba_5yrs <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2005)

nba_5yrs$Year <- as.character(nba_5yrs$Year)
ggplot(nba_5yrs, aes(Year, pa3.game)) +
  geom_boxplot() +
  theme_bw() +
  ggtitle("3 pt attempt per game for all players from 1995 to 2005") +
  xlab("Year") +
  ylab("3 pt attempts per game")
```


3 pt attempt per game for all players from 1995 to 2005



Then, we create and apply function to find outliers for each boxplot.

```
nba_outlier_5 <- nba_raw %>%
  filter(Year >= 2001 & Year <= 2005)

nba_outlier_p5 <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2000)

outlier <- function(data,var_name){
  vec <- data[,var_name]
  IQR <- IQR(vec)
  lower <- quantile(vec)[2] - 1.5*IQR
  upper <- quantile(vec)[4] + 1.5*IQR
  return(unique(data[vec>upper|vec<lower,"Player"]))
}

outlier_2001_2005 <- c(outlier(nba_outlier_5,"blk.game"))
outlier_1995_2000 <- c(outlier(nba_outlier_p5,"blk.game"))
yr5_BLK <- setdiff(outlier_2001_2005,outlier_1995_2000)
yr5_BLK
```

```
## [1] "Tony Battie"      "Mark Blount"      "Vince Carter"
## [4] "Shawn Marion"     "Kenyon Martin"    "Darius Miles"
## [7] "Dirk Nowitzki"    "Jermaine O'Neal"  "Scot Pollard"
## [10] "Robert Traylor"   "Jahidi White"     "Chris Andersen"
## [13] "Tyson Chandler"   "Pau Gasol"        "Eddie Griffin"
## [16] "Brendan Haywood"  "Jerome James"     "Andrei Kirilenko"
## [19] "Todd MacCulloch"  "Chris Mihm"       "Rasho Nesterovic"
## [22] "Joel Przybilla"   "Clifford Robinson" "Stromile Swift"
## [25] "Samaki Walker"    "Shane Battier"    "Jonathan Bender"
## [28] "Steven Hunter"    "Jamaal Magloire"  "Yao Ming"
```

```
## [31] "Mamadou N'Diaye"      "Kurt Thomas"      "Chris Bosh"
## [34] "Eddy Curry"           "Samuel Dalembert"  "Dan Gadzuric"
## [37] "Brad Miller"          "Amar'e Stoudemire" "Etan Thomas"
## [40] "David Harrison"       "Dwight Howard"     "Nazr Mohammed"
## [43] "Emeka Okafor"         "Josh Smith"        "Antoine Walker"
## [46] "Gerald Wallace"
```

```
nba_outlier_5 <- nba_raw %>%
  filter(Year >= 2001 & Year <= 2005)

nba_outlier_p5 <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2000)

outlier <- function(data,var_name){
  vec <- data[,var_name]
  IQR <- IQR(vec)
  lower <- quantile(vec)[2] - 1.5*IQR
  upper <- quantile(vec)[4] + 1.5*IQR
  return(unique(data[vec>upper|vec<lower,"Player"]))
}

outlier_2001_2005 <- c(outlier(nba_outlier_5,"ast.game"))
outlier_1995_2000 <- c(outlier(nba_outlier_p5,"ast.game"))
yr5_AST <- setdiff(outlier_2001_2005,outlier_1995_2000)
yr5_AST
```

```
## [1] "Travis Best"      "Baron Davis"      "Toni Kukoc"      "Jamal Mashburn"
## [5] "Jeff McInnis"     "Steve Nash"       "Antoine Walker"  "Chauncey Billups"
## [9] "Kobe Bryant"      "Lamar Odom"       "Jerry Stackhouse" "Jason Terry"
## [13] "Jamaal Tinsley"   "Alvin Williams"   "Ray Allen"       "Gilbert Arenas"
## [17] "Ricky Davis"      "Howard Eisley"    "Kevin Garnett"   "Troy Hudson"
## [21] "Tracy McGrady"    "Tony Parker"      "Chris Webber"    "Chucky Atkins"
## [25] "Brent Barry"      "Vlade Divac"      "T.J. Ford"       "Kirk Hinrich"
## [29] "LeBron James"     "Damon Jones"      "Bob Sura"        "Rafer Alston"
## [33] "Speedy Claxton"   "Marko Jaric"      "Tyronn Lue"      "Luke Ridnour"
## [37] "Dwyane Wade"    "Mo Williams"
```

```
nba_outlier_5 <- nba_raw %>%
  filter(Year >= 2001 & Year <= 2005)

nba_outlier_p5 <- nba_raw %>%
  filter(Year >= 1995 & Year <= 2000)

outlier <- function(data,var_name){
  vec <- data[,var_name]
  IQR <- IQR(vec)
  lower <- quantile(vec)[2] - 1.5*IQR
  upper <- quantile(vec)[4] + 1.5*IQR
  return(unique(data[vec>upper|vec<lower,"Player"]))
}

outlier_2001_2005 <- c(outlier(nba_outlier_5,"pa3.game"))
outlier_1995_2000 <- c(outlier(nba_outlier_p5,"pa3.game"))
yr5_3PA <- setdiff(outlier_2001_2005,outlier_1995_2000)
yr5_3PA
```

```
## [1] "Darrell Armstrong"  "Vince Carter"      "Stephon Marbury"
## [4] "Jerry Stackhouse"   "Antoine Walker"    "Baron Davis"
```

```
## [7] "Michael Dickerson" "Derek Fisher" "Cuttino Mobley"
## [10] "Paul Pierce" "Jason Terry" "Chauncey Billups"
## [13] "Pat Garrity" "Eddie Jones" "Tracy McGrady"
## [16] "Mike Miller" "Michael Redd" "Peja Stojakovic"
## [19] "Rafer Alston" "Derek Anderson" "Gilbert Arenas"
## [22] "Jamal Crawford" "Michael Finley" "Jim Jackson"
## [25] "Stephen Jackson" "Quentin Richardson" "Rasheed Wallace"
## [28] "Chucky Atkins" "Kobe Bryant" "Kirk Hinrich"
## [31] "Damon Jones" "Jason Kidd" "Kyle Korver"
## [34] "Rashard Lewis" "Donyell Marshall" "Vladimir Radmanovic"
## [37] "Jason Richardson"
```

After finding three groups of outliers, we find the players who were not outliers before but became outliers in multiple variables after 2001.

```
ast_3pa <- intersect(yr5_AST,yr5_3PA)
ast_3pa
```

```
## [1] "Baron Davis" "Antoine Walker" "Chauncey Billups" "Kobe Bryant"
## [5] "Jerry Stackhouse" "Jason Terry" "Gilbert Arenas" "Tracy McGrady"
## [9] "Chucky Atkins" "Kirk Hinrich" "Damon Jones" "Rafer Alston"
```

```
blk_3pa <- intersect(yr5_3PA,yr5_BLK)
blk_3pa
```

```
## [1] "Vince Carter" "Antoine Walker"
```

```
ast_blk <- intersect(yr5_AST,yr5_BLK)
ast_blk
```

```
## [1] "Antoine Walker"
```

```
all <- intersect(intersect(yr5_AST,yr5_BLK),yr5_3PA)
all
```

```
## [1] "Antoine Walker"
```

7. Case Study

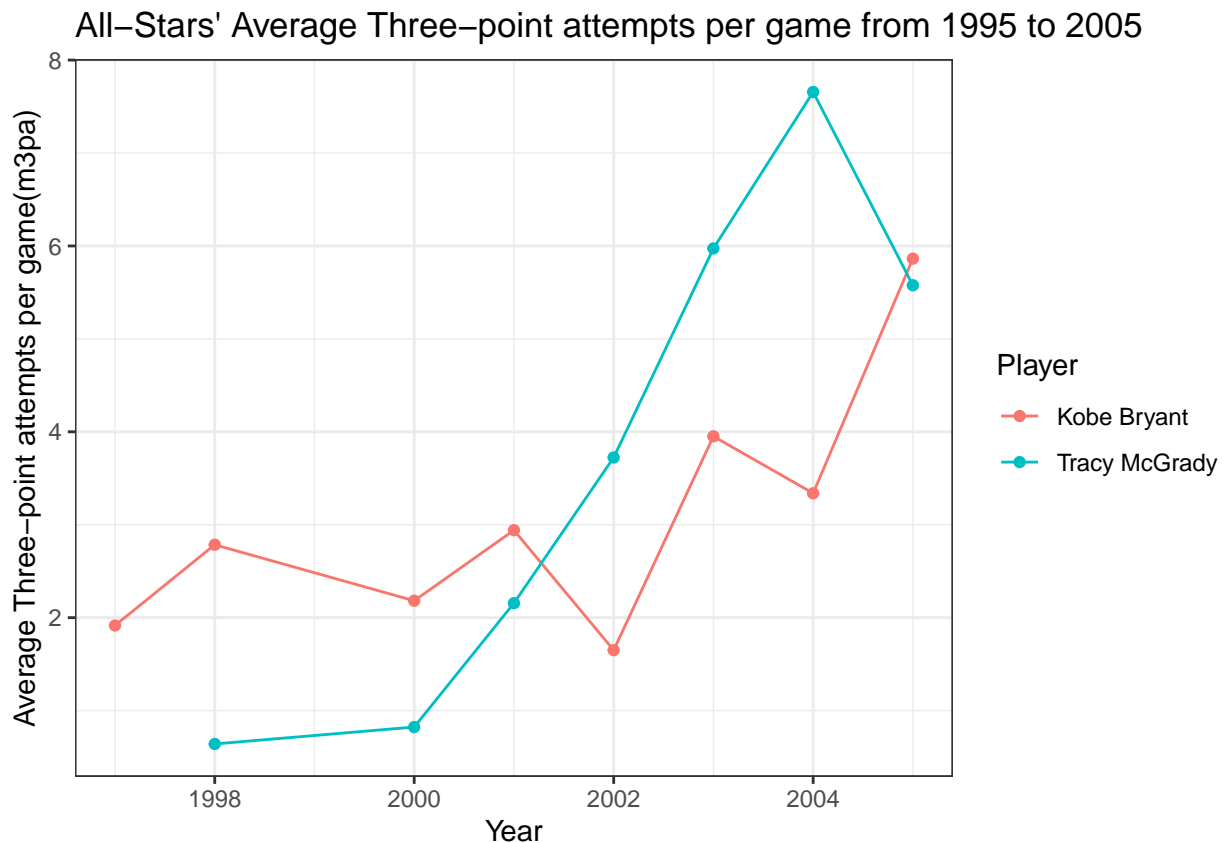
According to the outliers we get, we choose two top players, Kobe Bryant and Tracy McGrady, from the list and analyze in depth. Since they play similar positions, we will analyze them together.

```
All_Star <- nba_raw %>%
  filter(Player%in%c("Kobe Bryant","Tracy McGrady")&Year>=1995&Year<=2005)
```

Kobe and Tracy: In the following code, we investigate 7 different variables in total, including USG(usage percentage), PER(player efficiency rating), AST(assist per game), BLK(block per game), 2PA(2 point attempt per game), 3PA(3-point attempts per game), and WS.48(win share per game). One statistics stands out among these 7 variables: 3-point attempt per game. As shown in the graph, after 2001, both Kobe and Tracy shot more three point balls. Before, they only shot 2 to 3 three point balls per game. However, after the rule change, their three point attempts climbed up to about 6 to 7 shots per game. This rise matches our expectation: as zone defense

was established, players are more likely to shoot around three point line since it is sometimes hard for them to cut through defense. This corresponds to the decreasing tendency we find in the graph for 2-point attempts: after 2001, both Kobe and Tracy tended to shoot fewer two-point balls as we can see a gradual decrease in 2-point attempts in the graph.

```
All_Star_3PA_per_game<-All_Star%>%
  ggplot(aes(x=Year,y=pa3.game,color=Player))+
  geom_point()+geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("All-Stars' Average Three-point attempts per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Three-point attempts per game(m3pa)")
print(All_Star_3PA_per_game)
```

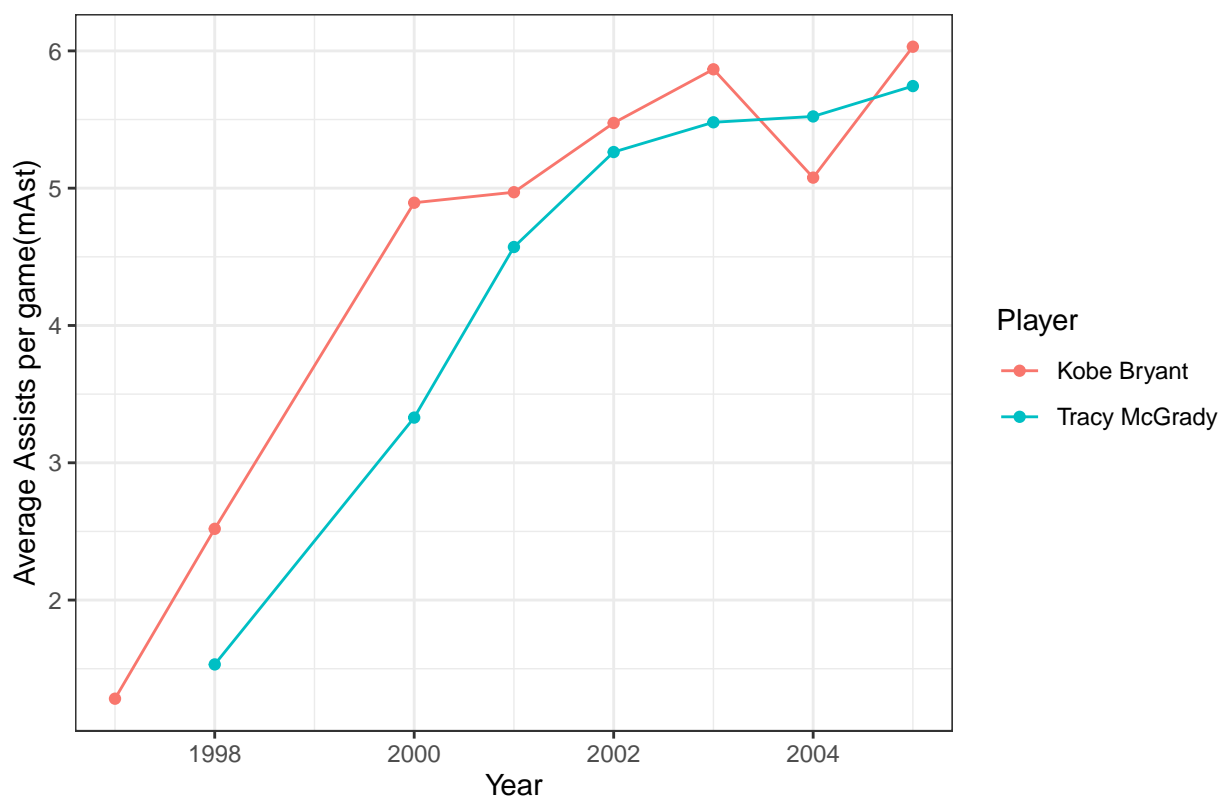


One interesting finding among these graph is that, for most variables, the value for both players increases from 1997/1998, when these two players started their careers. One reasonable explanation for this phenomenon is that they were quickly getting better and better after entering the league. At first, their stats started from a small value, but increased in a fast pace as they were getting familiar with the game. This can explain why these two phenomenal players are included in the list we found above. They were not outliers before rule change only because they just started their careers.

However, the analysis on them still provides insights for us since we can safely conclude that the rule change somewhat promoted players to shoot more three point balls.

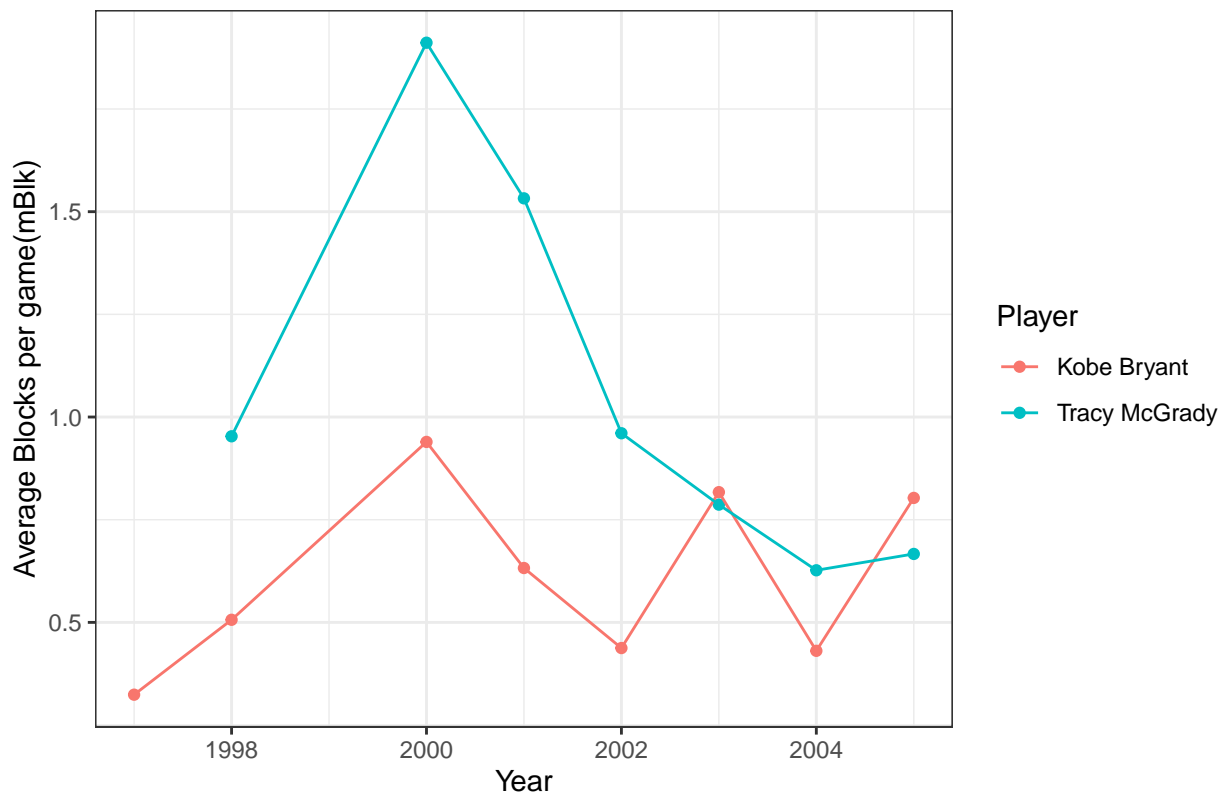
```
All_Star_AST_per_game<-All_Star%>%
  ggplot(aes(x=Year,y=ast.game,color=Player))+
  geom_point()+geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("All-Stars' Average Assists per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Assists per game(mAst)")
print(All_Star_AST_per_game)
```

All-Stars' Average Assists per game from 1995 to 2005



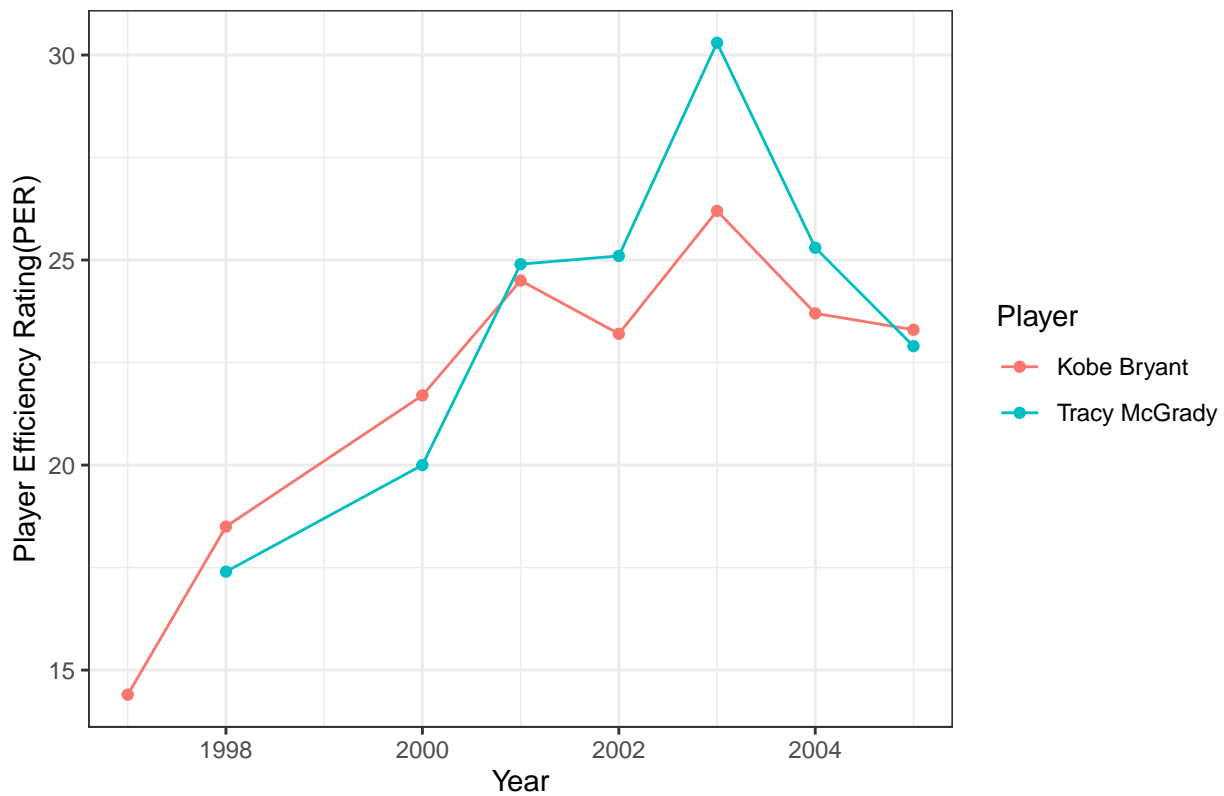
```
All_Star_BLK_per_game<-All_Star%>%
  ggplot(aes(x=Year,y=blk.game,color=Player))+
  geom_point()+geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("All-Stars' Average Blocks per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Blocks per game(mBlk)")
print(All_Star_BLK_per_game)
```

All-Stars' Average Blocks per game from 1995 to 2005



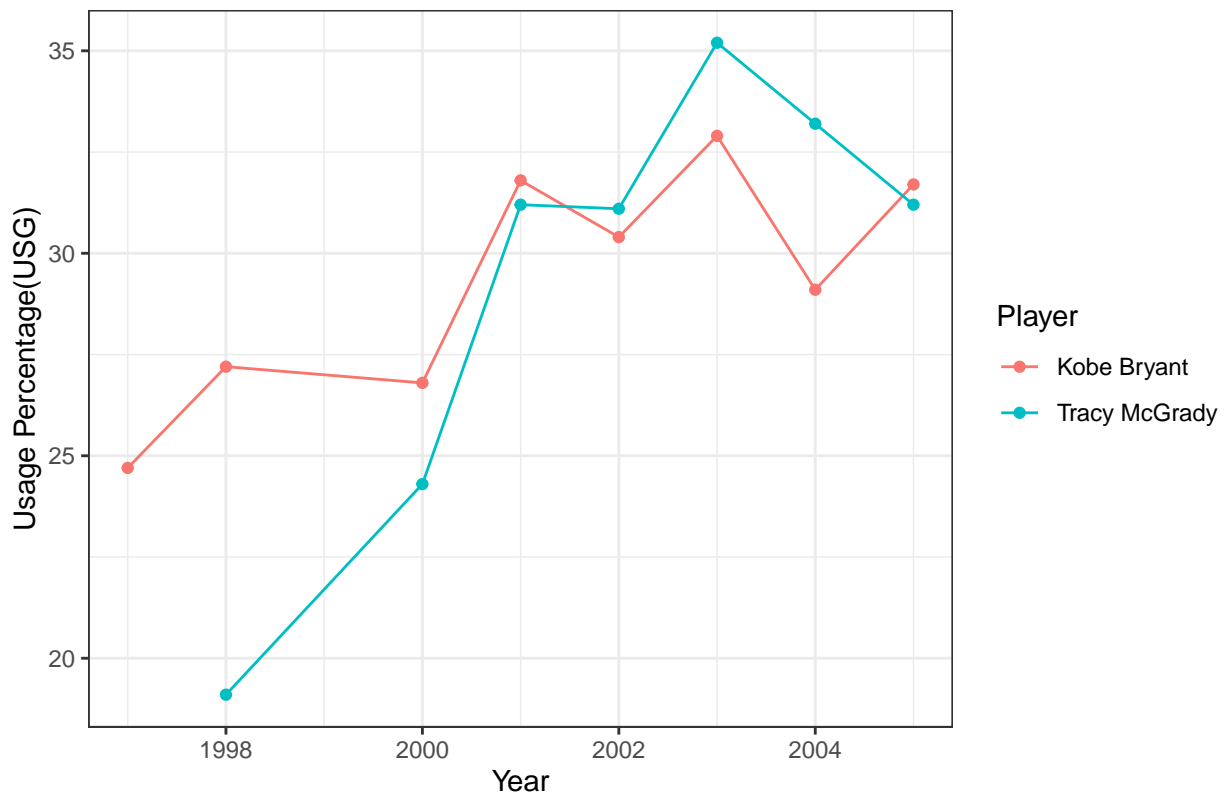
```
All_Star_PER<-All_Star%>%
  ggplot(aes(x=Year,y=PER,color=Player))+
  geom_point()+geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("All-Stars' Player Efficiency Rating(PER) from 1995 to 2005") +
  xlab("Year") +
  ylab("Player Efficiency Rating(PER)")
print(All_Star_PER)
```

All-Stars' Player Efficiency Rating(PER) from 1995 to 2005



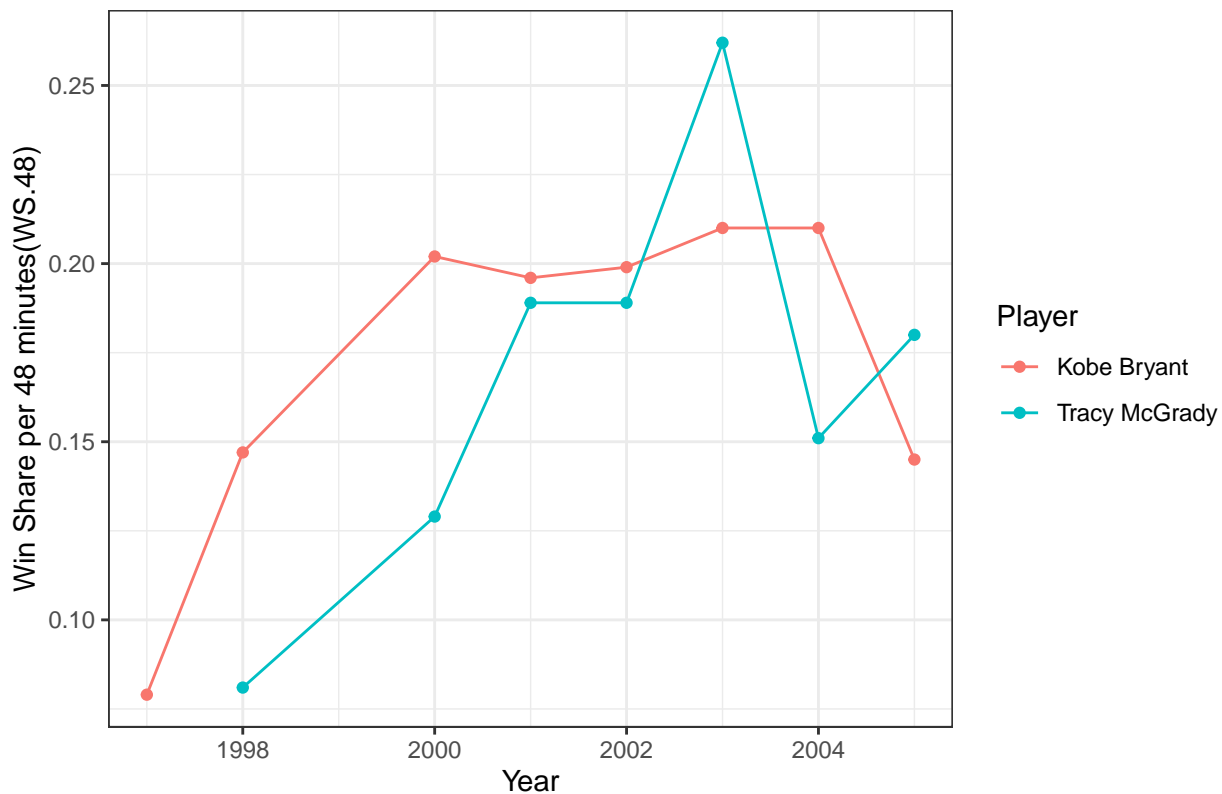
```
All_Star_USG<-All_Star%>%
  ggplot(aes(x=Year,y=USG.,color=Player))+
  geom_point()+
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("All-Stars' Usage Percentage(USG) from 1995 to 2005") +
  xlab("Year") +
  ylab("Usage Percentage(USG)")
print(All_Star_USG)
```

All-Stars' Usage Percentage(USG) from 1995 to 2005



```
All_Star_WS_48<-All_Star%>%
  ggplot(aes(x=Year,y=WS.48,color=Player))+
  geom_point()+
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("All-Stars' Win Shares per 48 minutes(WS.48) from 1995 to 2005") +
  xlab("Year") +
  ylab("Win Share per 48 minutes(WS.48)")
print(All_Star_WS_48)
```


All-Stars' Win Shares per 48 minutes(W.S.48) from 1995 to 2005



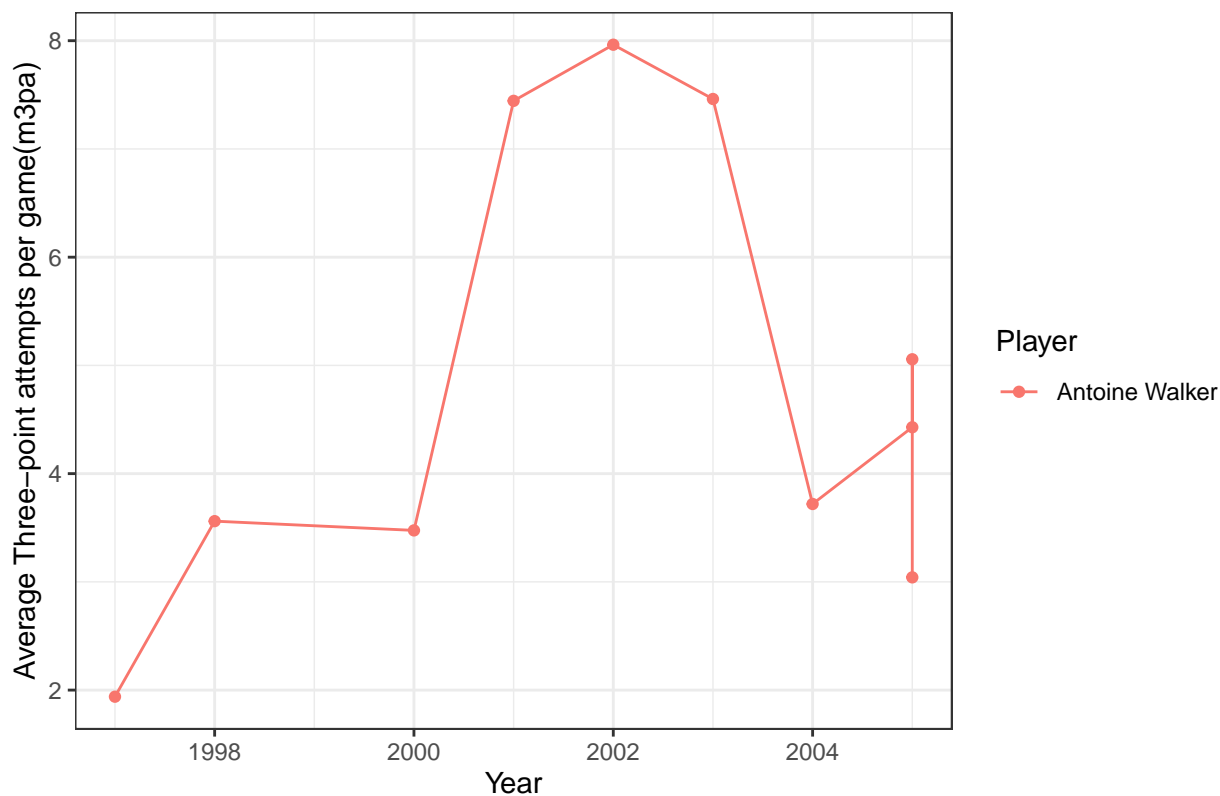
Antoine walker: During the filtering of outlier players, Antoine Walker stood out among all other players as he is the only player listed under the intersection of AST, 3PA, and BLK. Similarly, when we ranked the top 5 power forwards with most 3PA, Antoine Walker is at the top of the list. Intrigued by his performance, we conducted a case study to analyze whether the rule changes actually had an impact on his box score stats and performance.

In the following code, we investigate 6 different variables in total, including MP(minutes played per game), USG(usage percentage), PER(player efficiency rating), AST(assist per game), BLK(block per game), 3PA(3-point attempts per game), and WS.48(win share per game).

```
Antoine_Walker <- nba_raw %>%
  filter(Player%in%c("Antoine Walker")&Year>=1995&Year<=2005)

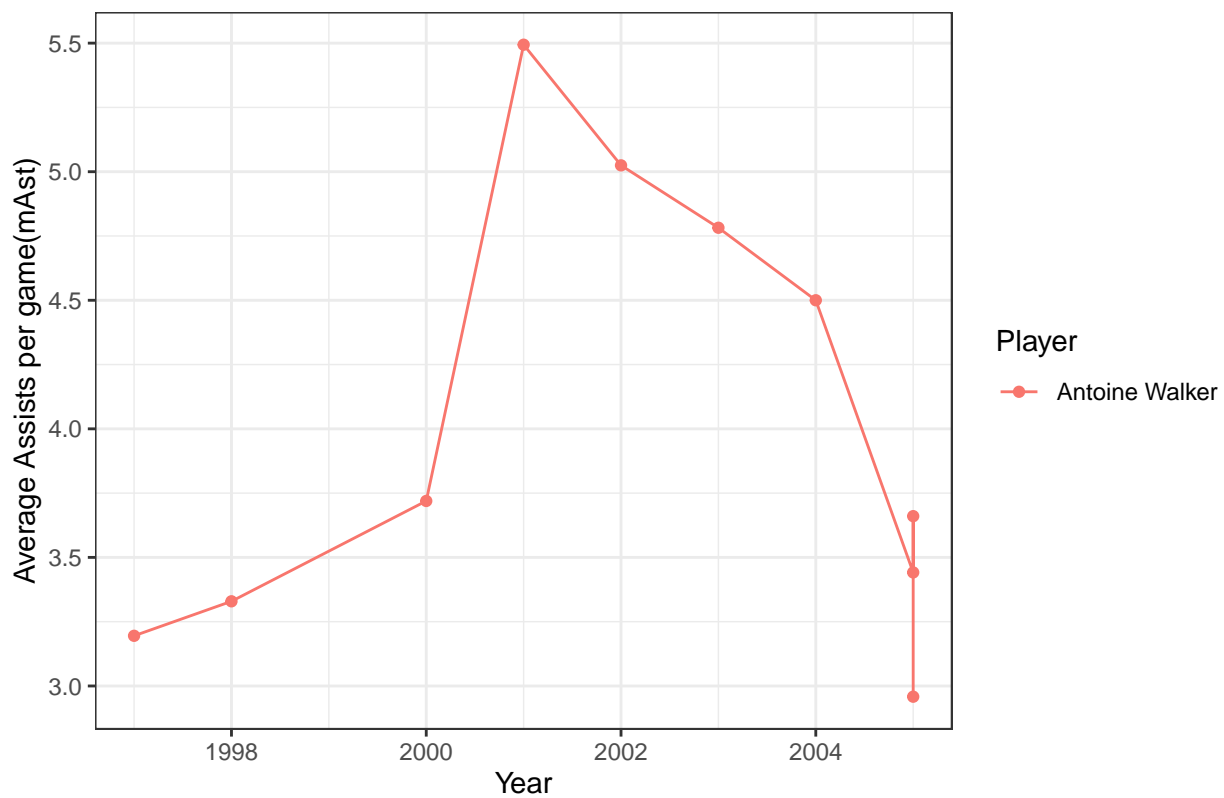
Antoine_Walker_3PA_per_game<-Antoine_Walker%>%
  ggplot(aes(x=Year,y=pa3.game,color=Player))+
  geom_point()+
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Antoine Walker's Average Three-point attempts per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Three-point attempts per game(m3pa)")
print(Antoine_Walker_3PA_per_game)
```

Antoine Walker's Average Three-point attempts per game from 1995 to 2005



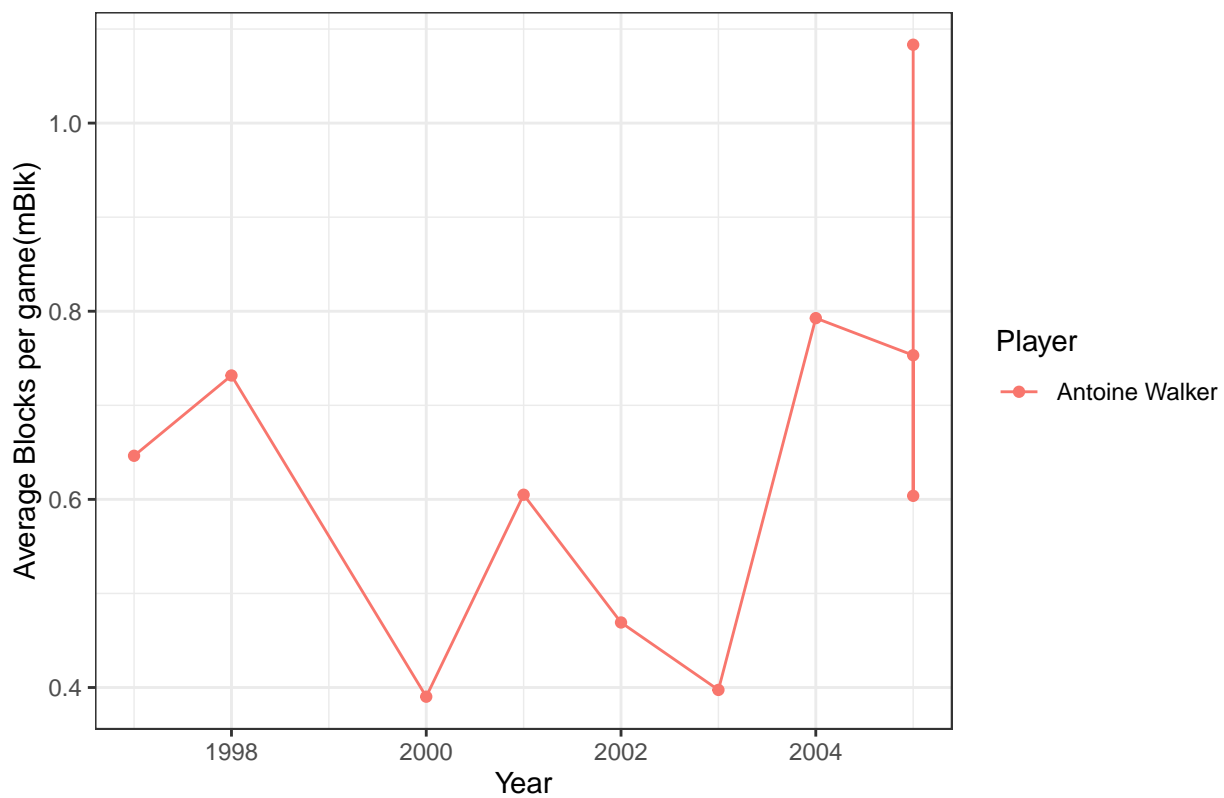
```
Antoine_Walker_AST_per_game<-Antoine_Walker%>%
  ggplot(aes(x=Year,y=ast.game,color=Player))+geom_point()+geom_line(stat = "identity", position = "identity")
  ggtitle("Antoine Walker's Average Assists per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Assists per game(mAst)")
print(Antoine_Walker_AST_per_game)
```

Antoine Walker's Average Assists per game from 1995 to 2005



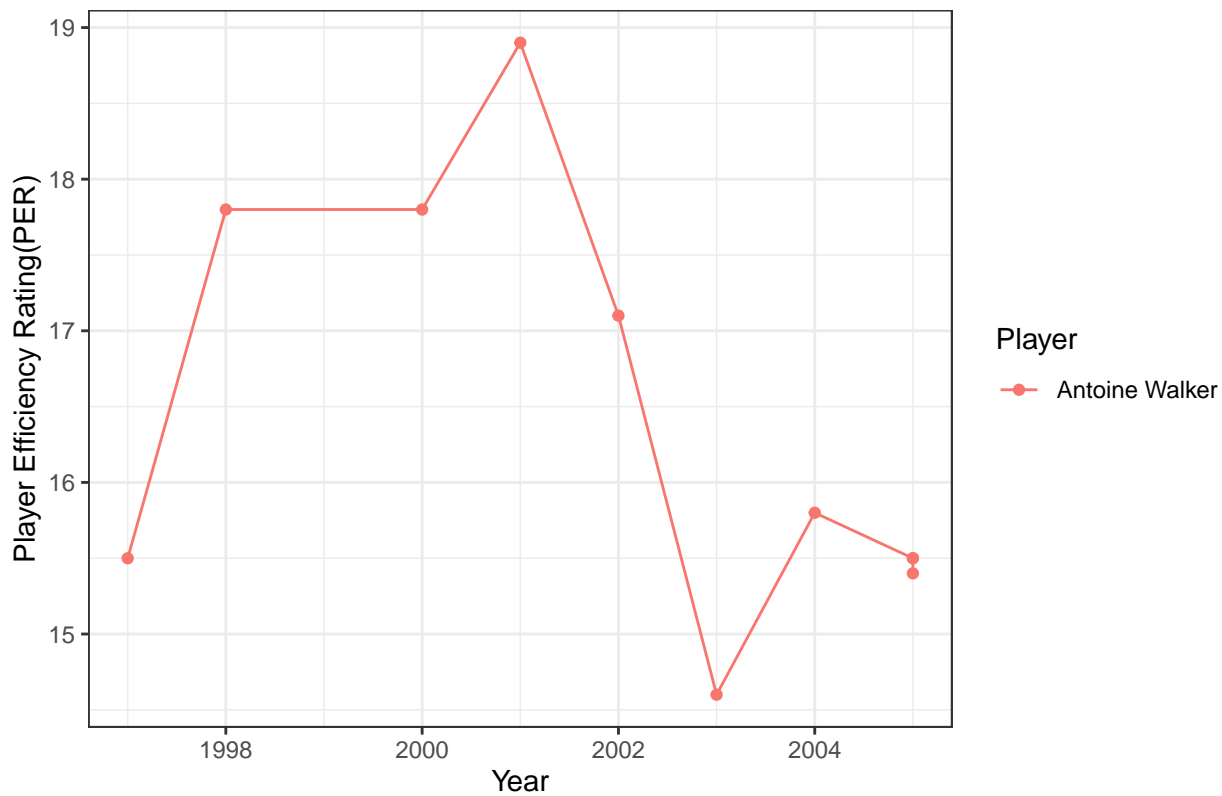
```
Antoine_Walker_BLK_per_game<-Antoine_Walker%>%
  ggplot(aes(x=Year,y=blk.game,color=Player))+
  geom_point()+
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Antoine Walker's Average Blocks per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Blocks per game(mBlk)")
print(Antoine_Walker_BLK_per_game)
```

Antoine Walker's Average Blocks per game from 1995 to 2005



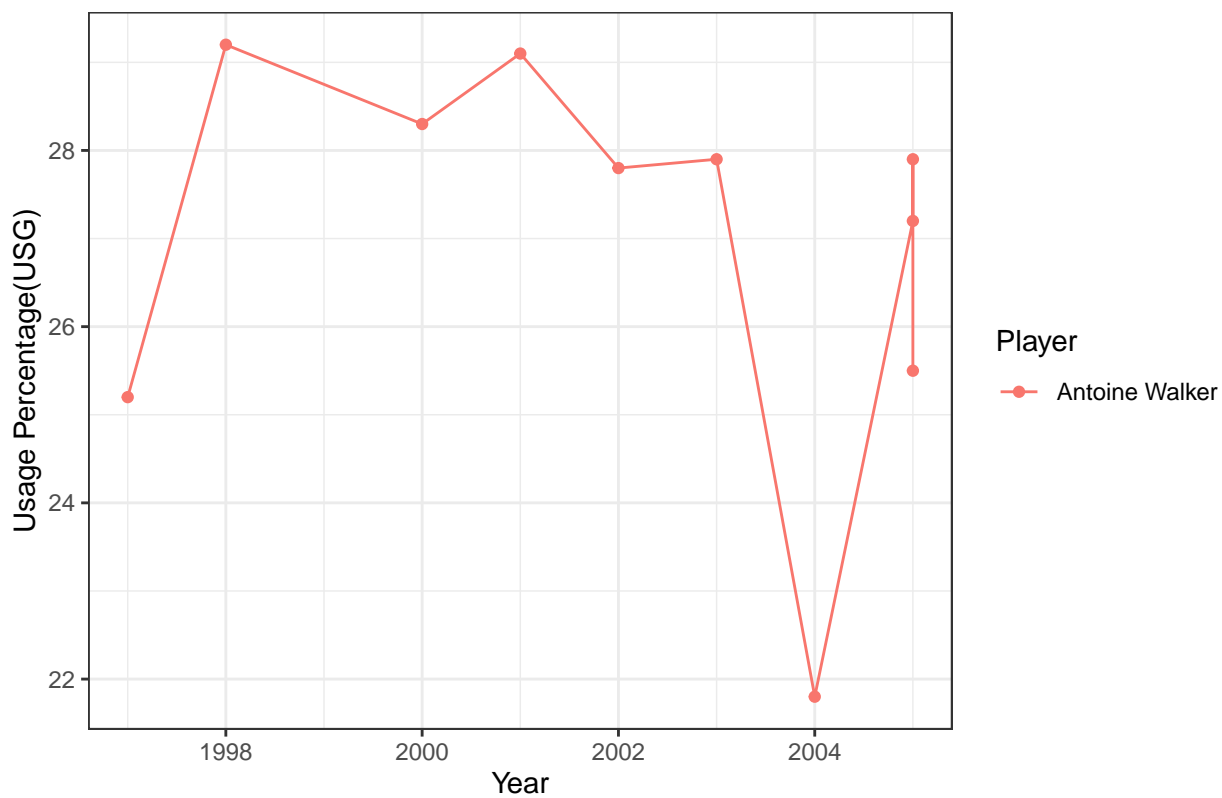
```
Antoine_Walker_PER<-Antoine_Walker%>%
  ggplot(aes(x=Year,y=PER,color=Player))+
  geom_point()+
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Antoine Walker's Player Efficiency Rating(PER) from 1995 to 2005") +
  xlab("Year") +
  ylab("Player Efficiency Rating(PER)")
print(Antoine_Walker_PER)
```

Antoine Walker's Player Efficiency Rating(PER) from 1995 to 2005



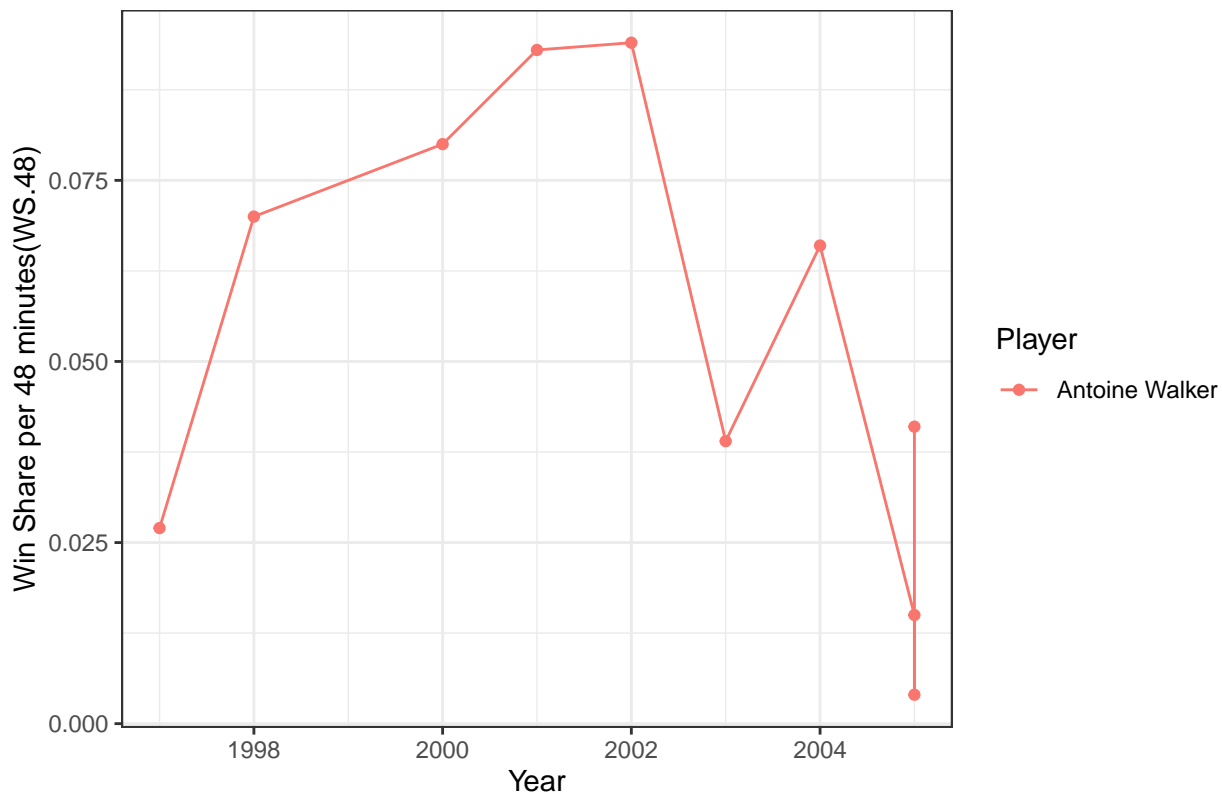
```
Antoine_Walker_USG<-Antoine_Walker%>%
  ggplot(aes(x=Year,y=USG.,color=Player))+
  geom_point()+
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Antoine Walker's Usage Percentage(USG) from 1995 to 2005") +
  xlab("Year") +
  ylab("Usage Percentage(USG)")
print(Antoine_Walker_USG)
```

Antoine Walker's Usage Percentage(USG) from 1995 to 2005



```
Antoine_Walker_WS_48<-Antoine_Walker%>%
  ggplot(aes(x=Year,y=WS.48,color=Player))+
  geom_point()+
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Antoine Walker's Win Shares per 48 minutes(WS.48) from 1995 to 2005") +
  xlab("Year") +
  ylab("Win Share per 48 minutes(WS.48)")
print(Antoine_Walker_WS_48)
```

Antoine Walker's Win Shares per 48 minutes(W.S.48) from 1995 to 2005



For 3PA,AST, and MP, we see a drastic increase in his stats from the 2001-2002 season, when the two rules were implemented. Unfortunately, such an increasing pattern was not long-term. All of the stats analyzed, except for BLK, started to decrease around 2003. Some of these decreasing patterns are drastic, while others are more gradual. Such a phenomenon might be associated with the fact that Walker was traded to the Dallas Mavericks at the beginning of 2003-2004 season, where he then had limited minutes on the floor due to the number of outstanding players on his new team. Thus, we cannot say that the rule change has association with the temporary improvement in his stats.

Overall, from what the graph has presented, we are inconclusive about the impact of rule change on Antoine Walker, as his improvement was very short-lived. This also makes us reflect upon the drawbacks of our filtering system, which we will mention later in the conclusion section.

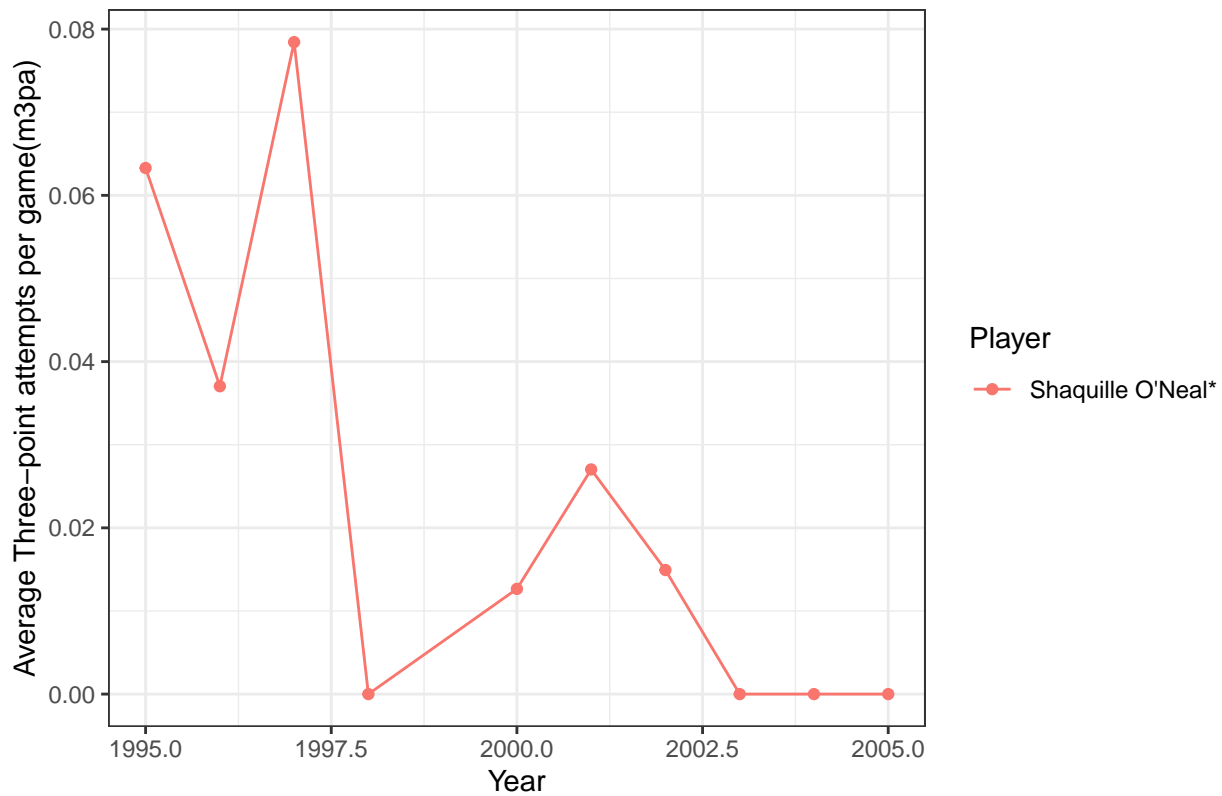
Shaquille O'Neal: Since there was such a narrative that zone defense was implemented to limit Shaq's ability to outperform and overwhelm his opponents, we conduct a case study to analyze whether the rule changes actually had an impact on his box score stats and performance.

In the following code, we investigate 6 different variables in total, including MP(minutes played per game), USG(usage percentage), PER(player efficiency rating), AST(assist per game), BLK(block per game), 3PA(3-point attempts per game), and WS.48(win share per game).

```
Shaq <- nba_raw %>%
  filter(Player=="Shaquille O'Neal*" & Year >= 1995 & Year <= 2005)
```

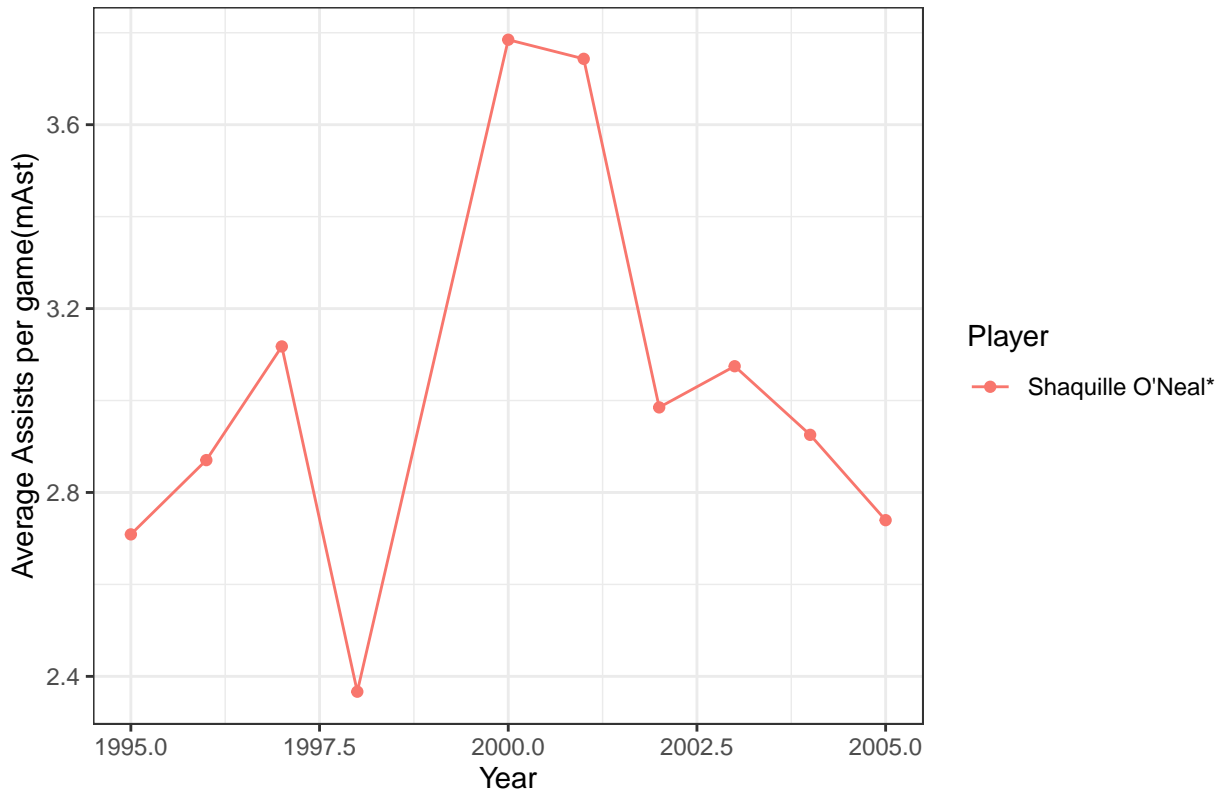
```
Shaq_3PA_per_game <- Shaq %>%
  ggplot(aes(x=Year,y=pa3.game,color=Player)) +
  geom_point() +
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Shaq's Average Three-point attempts per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Three-point attempts per game(m3pa)")
print(Shaq_3PA_per_game)
```

Shaq's Average Three-point attempts per game from 1995 to 2005



```
Shaq_AST_per_game <- Shaq %>%
  ggplot(aes(x=Year,y=ast.game,color=Player)) +
  geom_point() +
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Shaq's Average Assists per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Assists per game(mAst)")
print(Shaq_AST_per_game)
```

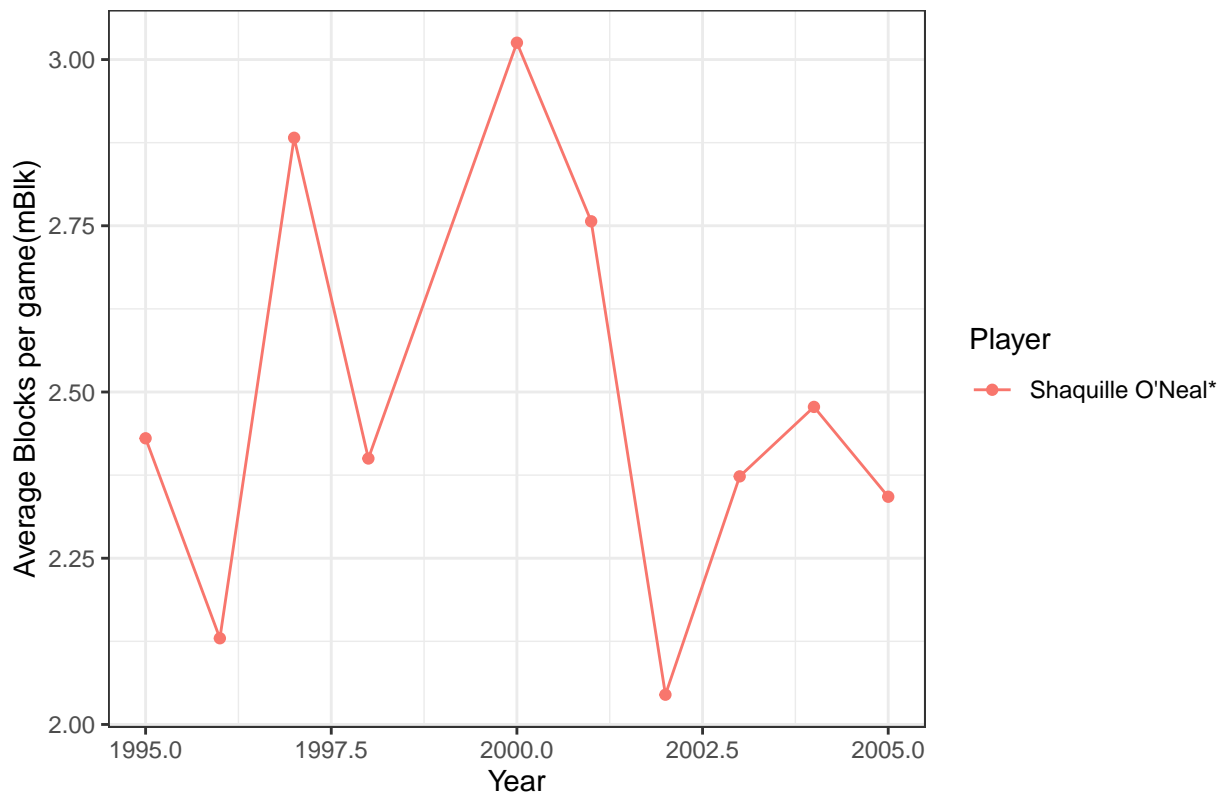

Shaq's Average Assists per game from 1995 to 2005



For both AST and 3PA, the trend-lines slightly fluctuate. For three-point attempts specifically, Shaq never attempted more than one three point ball per game in any season before and after the rule change. Indeed, after the rule change, he shot fewer threes than before. Considering that Shaq was named MVP of the NBA Finals from 2000 to 2002, winning three consecutive titles, we can reasonably speculate that the rule change did not prompt Shaq to change his playing style, neither it was shooting more threes nor making more assists, as he still dominated by scoring in the low-post.

```
Shaq_BLK_per_game <- Shaq %>%
  ggplot(aes(x=Year,y=blk.game,color=Player)) +
  geom_point() +
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Shaq's Average Blocks per game from 1995 to 2005") +
  xlab("Year") +
  ylab("Average Blocks per game(mBlk)")
print(Shaq_BLK_per_game)
```

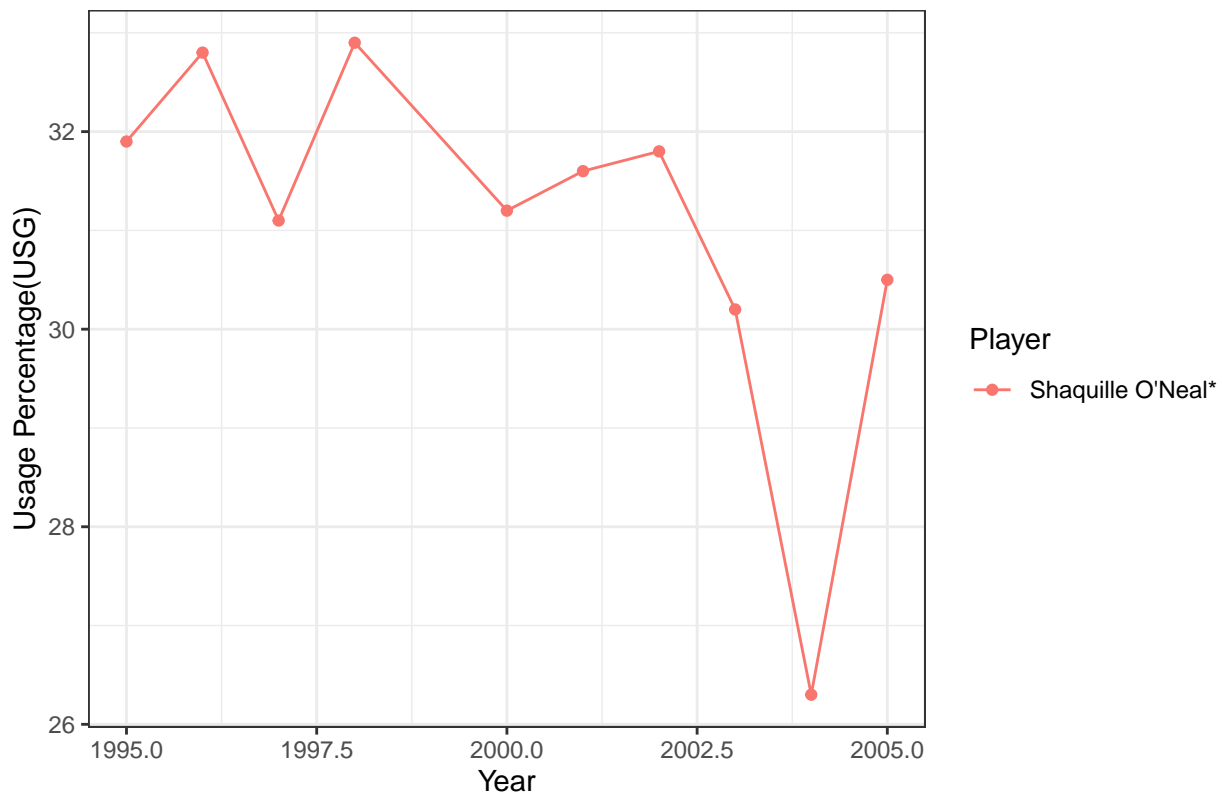
Shaq's Average Blocks per game from 1995 to 2005



For blocks per game, though we did not see a consistent pattern after the rule change, we can see that Shaq's average blocks per game have decreased after the rule change and never rebounded back to its peak thereafter. Considering that Shaq was at his prime from year 2000 to 2005 since he won three championships in a row, the factor of aging is negligible. Thus, we can say that in terms of Shaq, there might be an association between blocks and the rule change.

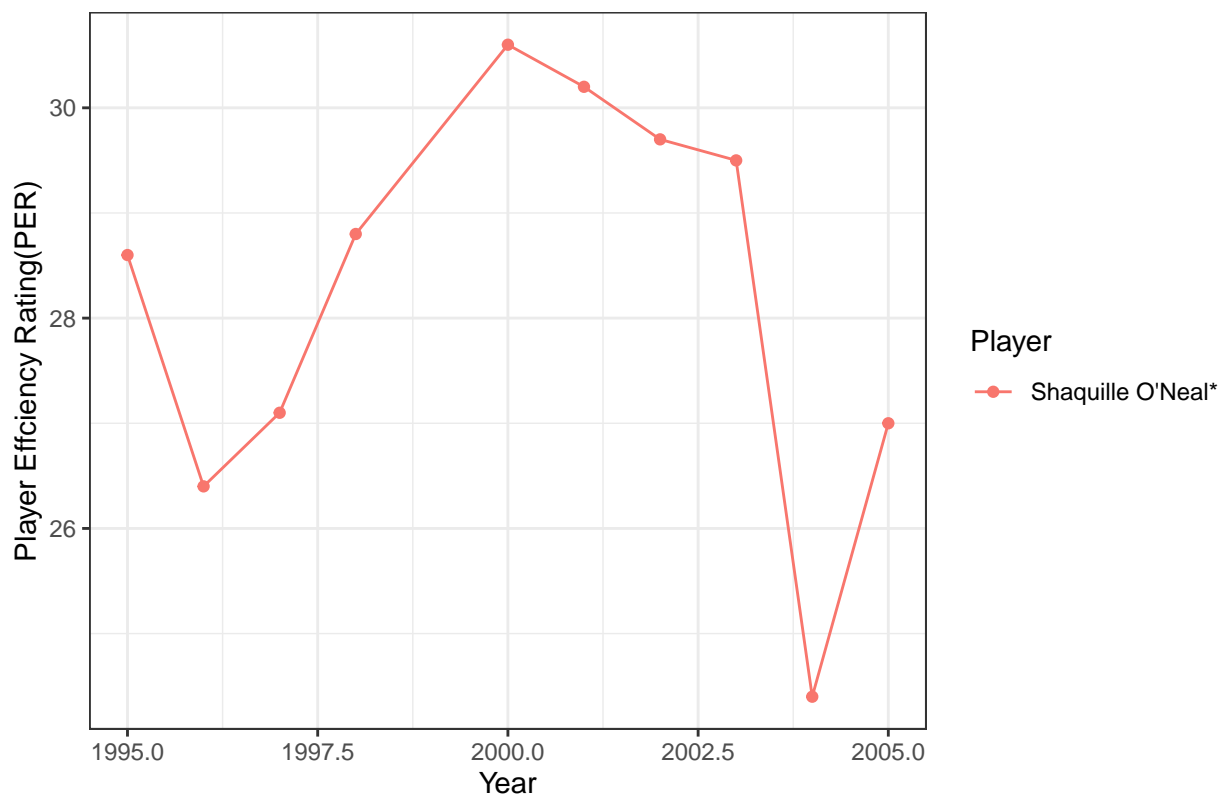
```
Shaq_USG <- Shaq %>%
  ggplot(aes(x=Year,y=USG.,color=Player)) +
  geom_point() +
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Shaq's Usage Percentage(USG) from 1995 to 2005") +
  xlab("Year") +
  ylab("Usage Percentage(USG)")
print(Shaq_USG)
```

Shaq's Usage Percentage(USG) from 1995 to 2005



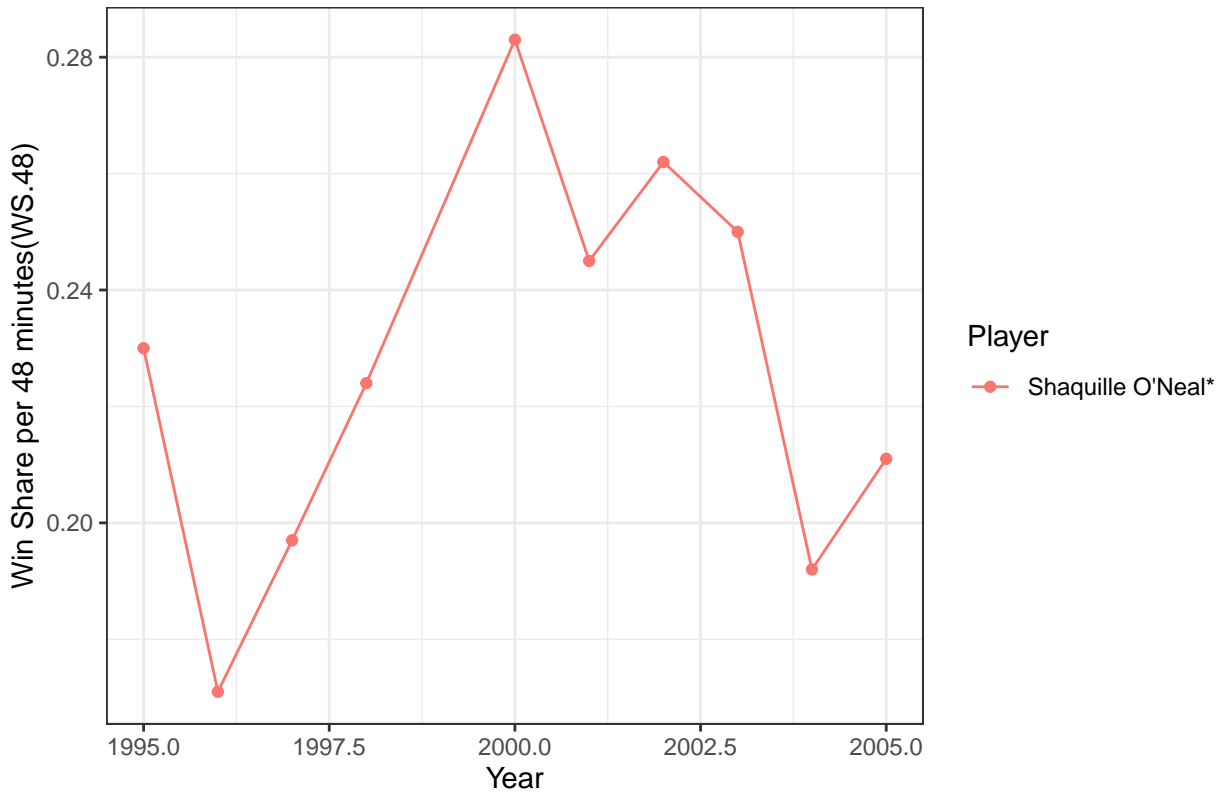
```
Shaq_PER <- Shaq %>%
  ggplot(aes(x=Year,y=PER,color=Player)) +
  geom_point() +
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Shaq's Player Efficiency Rating(PER) from 1995 to 2005") +
  xlab("Year") +
  ylab("Player Efficiency Rating(PER)")
print(Shaq_PER)
```

Shaq's Player Efficiency Rating(PER) from 1995 to 2005



```
Shaq_WS_48<-Shaq %>%
  ggplot(aes(x=Year,y=WS.48,color=Player)) +
  geom_point() +
  geom_line(stat = "identity", position = "identity") +
  theme_bw()+
  ggtitle("Shaq's Win Shares per 48 minutes(WS.48) from 1995 to 2005") +
  xlab("Year") +
  ylab("Win Share per 48 minutes(WS.48)")
print(Shaq_WS_48)
```

Shaq's Win Shares per 48 minutes(W.S.48) from 1995 to 2005



For PER, WS per 48 minutes, and USG, we see a similar declining pattern for Shaq after the year of rule change, 2001. Though such decrease is not drastic, considering that Shaq was in his prime years from 2001 to 2005, such a declining pattern could be associated with the rule change. The implementation of zone defense may have incentivized the coach to run more game tactics that are designed for long-range shooting or different execution of scoring, thus reducing Shaq's USG. For instance, the Laker coach Phil Jackson developed the Triangle offence to create more movement of the ball and versatility of scoring. Shaq might also have fewer opportunities attacking in the post or have lower efficiency in scoring as zone defense allowed other teammates to help, making it more challenging for him to score when double-teamed. These factors might have contributed to decline in WS.48 and PER.

Interestingly, in all these four graphs, we can see the drop from 2002 to 2003. Despite the rule change, this can be also associated with a toe injury at the beginning of 2002-2003 season, which made him miss 12 games and had a negative impact on his game performance thereafter.

Overall, the changes are not drastic. Shaq still outperformed most players in the NBA league, which potentially indicates that the rule change did not have a significant impact on his game.

8. Conclusion and Improvement

Conclusion

After analyzing the whole league and several specific players, we find out the most possible effect of the rule change is that the average 3-point attempts for players tends to increase. This corresponds to our initial expectation of this variable: we expect more 3-point attempts since zone defense lowers players' willingness to cut through defense. This increasing trend can be found not only in the plots where we analyze the league as a whole, but also in the part we analyze top players. Specifically, for power forwards, the effect on 3-Point attempt is the most obvious, as shown in the t-test in our modeling part. For other variables, for example assist and block, which we expected to change, they are actually not really affected by the rule change in 2001.

As the main driving force of the rule change in 2001, Shaquille O'Neal's game statistics was not drastically affected by the rule change. He still outperforms other players in the NBA league. However, some interesting associations were

suggested. First, Shaq did not intend to change his playing style by shooting more threes or making more assists, as he still found himself dominating in the paint. Second, Shaq's performance with regards to WS, 48, PER, USG, and BLK did decrease during his prime years from 2001 to 2005. Arguing from a perspective supporting our assumptions, the rule change might have prompted his team to design more versatile game tactics deviating from low-post plays. Secondly, the rule change may have reduced his offensive efficiency and defensive presence. On the other hand, this drop might be partially associated with his toe injury.

Reflection & Improvement

There are several improvements can be made regarding our project.

First of all, in terms of case study of individual players, there is one drawback on the function we filter the outliers. The outliers we filtered out may not actually qualify to be "outliers." Just like Kobe and Tracy, they are two "outliers" only because they just started their careers in 1997/1998. After entering NBA, their performance became better and better. This is also the reason why they were not outliers before rule change. However, this type of players is not what we intend to find. We want to find out those players who were really changed by the rules. Hence, we can include draft year in our filtering function so that we can remove those players who just start their careers several years before rule change. In addition, we can also write functions that count the total times of being outlier for a player in order to remove players who only had temporary good performance, for example, one year.