

# MA415 Final Project: NBA 3 Point Shooting Analysis

*Ziran Min*

*5/7/2018*

## Introduction

Over the past 4 years in NBA, one team, Golden State Warriors, or one person, Stephen Curry, has changed the world of basketball by 3 point shooting. Because of Stephen Curry's insane 3 point shooting skills and Golden State Warriors' dominating 3 point shooting plays, it seems more and more teams in NBA start to shoot more threes in games. Meanwhile, it seems players become less preferable to attack the rim, since there are less dunks but more threes in the All Stars Games in recent years, which is the trend that many fans are complaining about.

Do players shoot further and further from the basket? Do NBA teams shoot more and more threes? Is the whole league changing shooting habit in general? Can shooting more threes bring more wins to a team?

In this project, I will use the game shooting data of recent 15 regular seasons (2003-04 to 2017-2018) from NBA official stats website stats.nba.com, to analyze the trend and development of 3 point shooting in the league and the relationship between 3 point shooting and game victory.

## Data

The NBA official website provides API links for people to access detailed data of every single player's every single shot attempted during every single game. The API link of every player's shooting data in a season contains an unique ID number of that player. So I also need to get every player's ID number to gain the shooting data. Luckily, NBA also has API link that has the basic information of every historical and current player which contains player ID number.

Therefore, I firstly get the data of every player's ID number and their active NBA seasons from "player basic info" API. (Both APIs I mentioned above have JSON data structures). Then by looping through the shooting data API of every active player in every season, I get the full shooting data of 15 seasons and download those 15 csv files for later use.

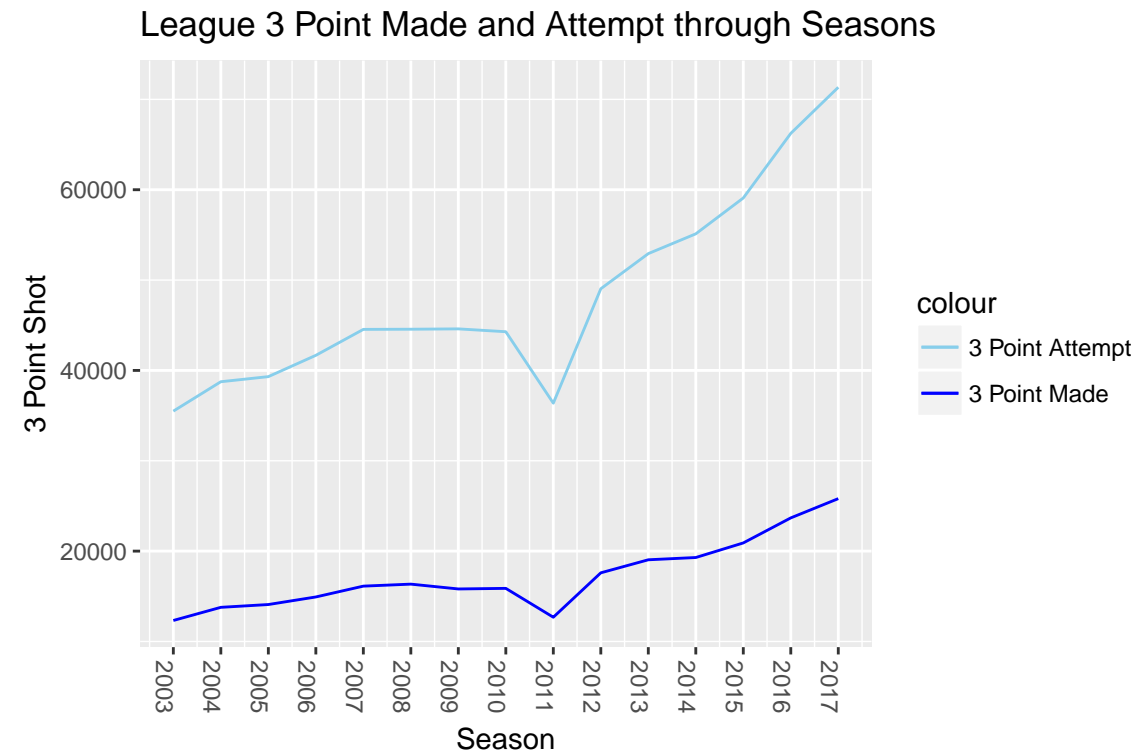
Furthermore, the NBA API doesn't give me the win/lose result of every game. In order to get feature about team win percent, I use package ballr which is a package that extract basketball data from www.basketball-reference.com. I find game result by using ballr and then calculate every team's win percent in every regular season.

To analyze single player's shooting habit, I can directly use the 15 files I create. But the my main goal is focus on team level and league level, so I manipulate the 15 big data sets and get a concise and small dataset called "ideal\_whole\_data.csv" for further analysis of shooting on team level and league level.

## Three Point Shooting in the League through 15 Seasons

After cleaning data and create useful features, I start to build several ggplot graphs to see the development of 3 point shooting in the league.

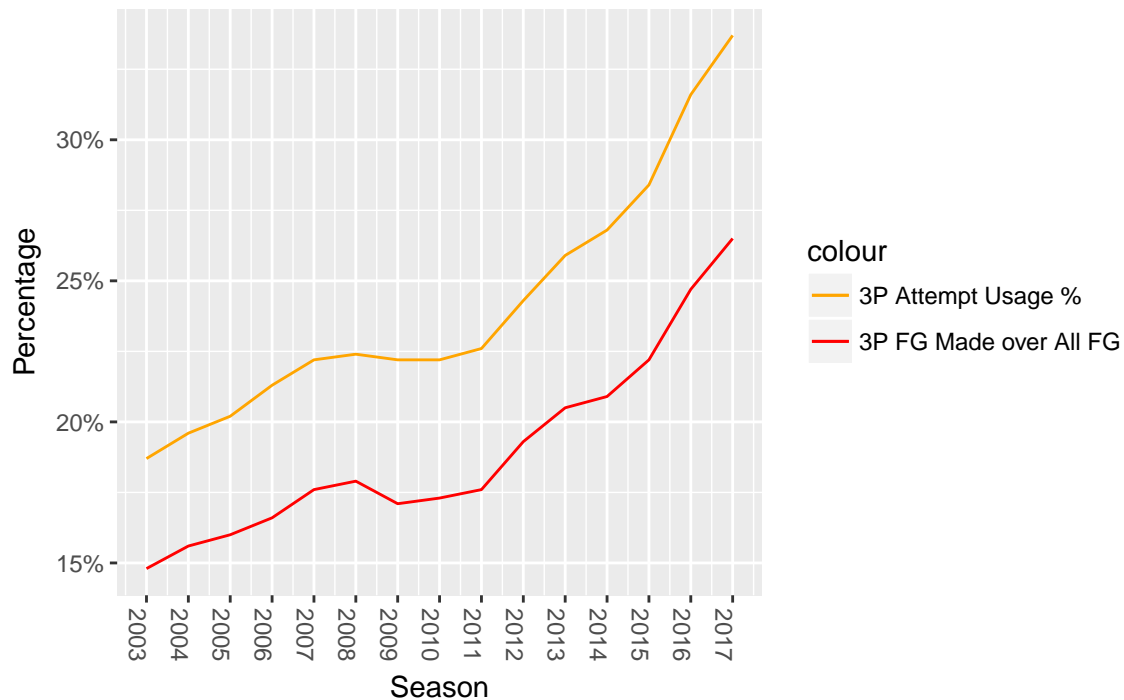
The first thing I want to check is whether teams shoot more and more threes. So I create following graph to see how the league total numbers of 3 point shot made and 3 point shot attempted changes through years.



It indeed seems both numbers of 3 point shot made and 3 point shot attempted is rising through years, but there is an obvious valley in season 2011-2012. The reason is that season had lockout which means each team only played 66 games (instead of 82 games as usual) in regular season. Since teams played less games, less 3 point shots then makes sense.

To solve the problem of teams playing less games in one special season, I caculated the percentage of 3 point attempt over all field goal attempt and the percentage of 3 point made over all field goal made. Then I get the following better lookng graph.

## League 3 Point Made and Attempt Percentage through Seasons



Now, we can clearly see that teams make more and more threes since 2011-2012 season. This indeed shows a rising popularity of 3 point shooting in the league in the past 6-7 years. However, which makes me surprised in the above graph is that between 2007 and 2011 teams didn't increase their favor in 3 point shooting and even became less preferable to shoot threes. As a die-hard NBA fan, I think the reason is that most league top players during that time were frontcourt players like Kevin Garnett, Tim Duncan, and LeBron James, or backcourt players like Kobe Bryant and Dwyane Wade. Most of them were good at and famous for midrange shooting and attacking in paint, so they lead the "fashion of 2 point shooting" during that time.

## Shooting Usage Percentage in Different Distances through Seasons

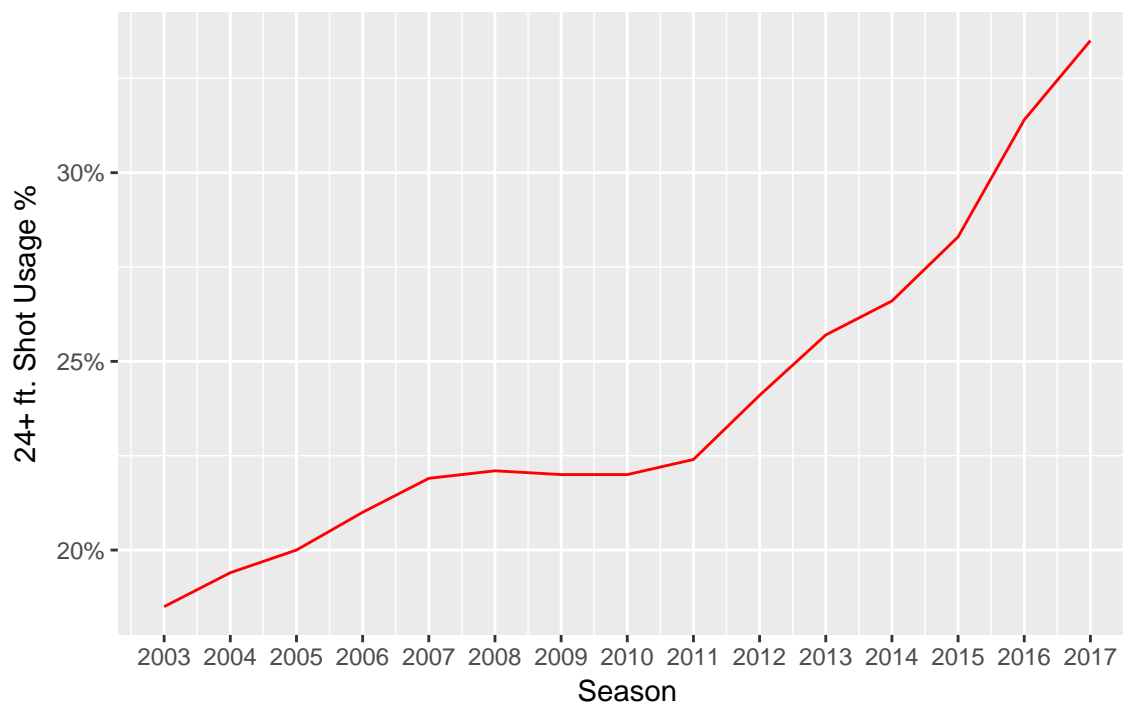
We have seen the rising popularity of 3 point shooting in recent years. How about other ways of offense? Because there are so many ways of offense counted as 2 point field goal, we can't know how the popularities of midrange shooting and paint attacking change by just looking the data of 2 point field goal.

In the dataset I create, it also records the distance to the basket of each shot. The distance is classified as 24+ ft., 16-14 ft., 8-16 ft., less than 8 ft., and back court. Here the number of back court is too few and some teams even had zero back court shot attempt in a season, so I ignore the shot in last distance category. Another note is that the length of NBA three point line range from 22 ft. to 23.75 ft., so the number of shot in 24+ group is not exactly the total number of 3 point shot, but slightly less than it.

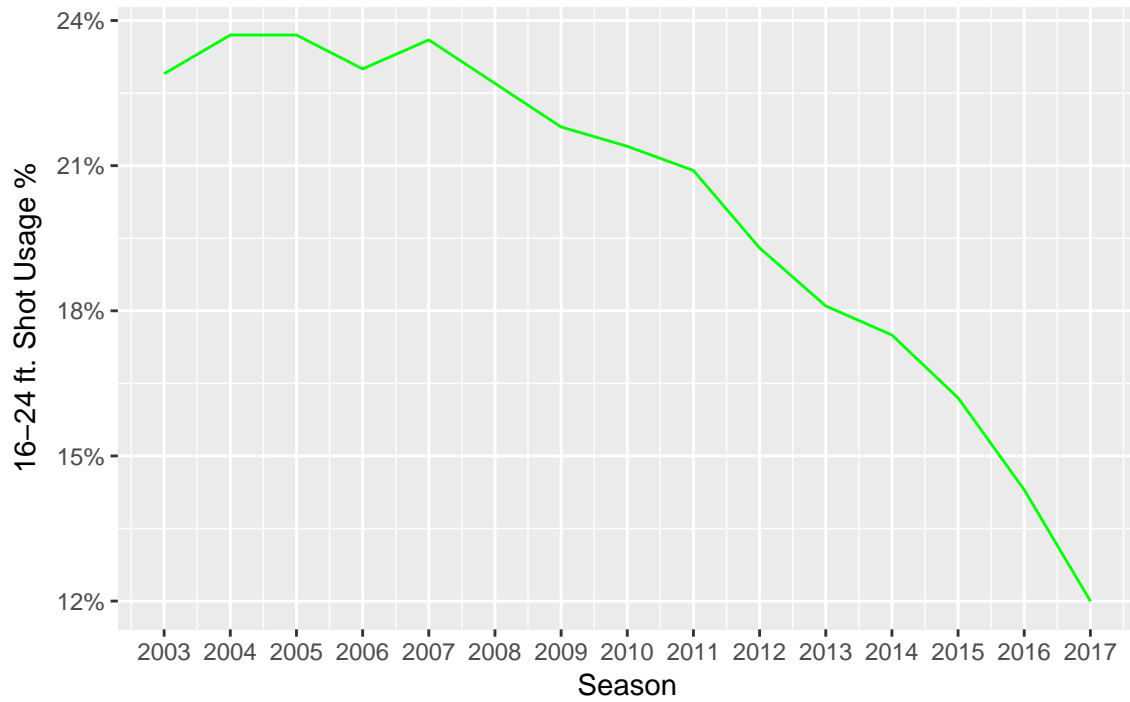
In my dataset, I calculate the league shooting usage percentgaes in different distance by dividing the total field goal attempt of one kind of shot by the total field goal attempt of all kinds of shot. Now I assume that in each season  $(24+ \text{ ft usage})\% + (16-24 \text{ ft usage})\% + (8-24 \text{ ft usage})\% + (\text{less than } 8 \text{ ft usage})\% = 1$  and regard 24+ ft shot as three point shot, 16-24 ft shot as long midrange shot, and less than 8 ft shot as paint area (restricted area) shot.

To visualize the usage of four kinds of shot through years, I firstly compute plots for them sperately.

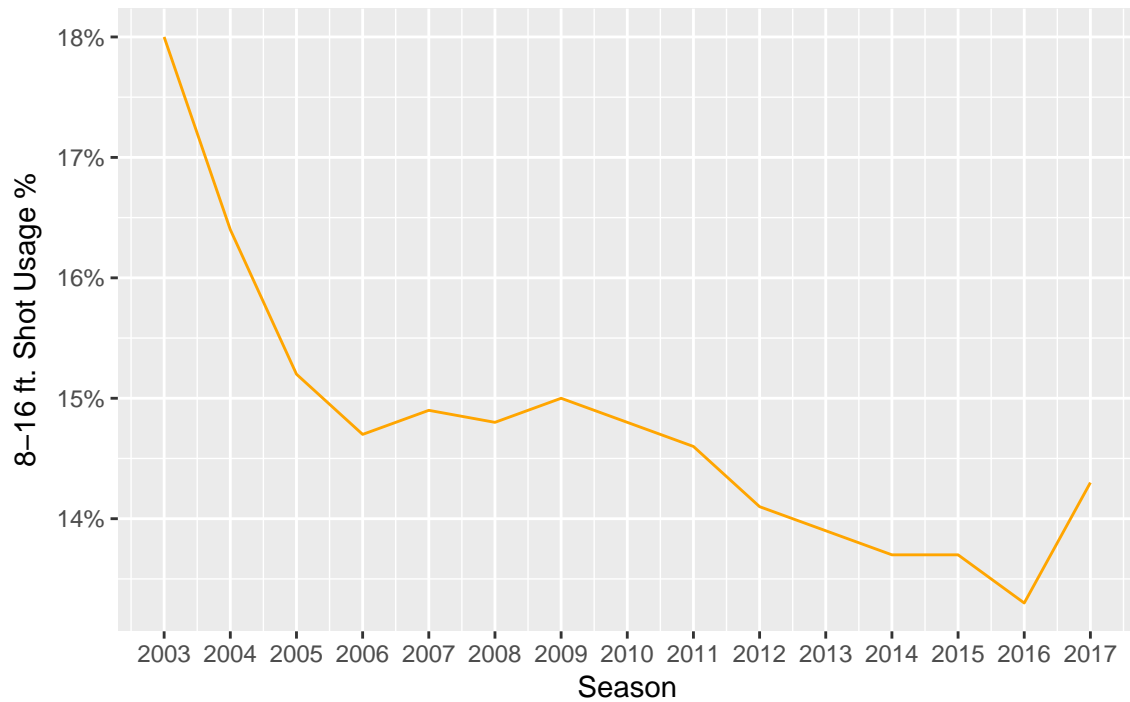
### League 24+ ft. Shot Usage % through Seasons

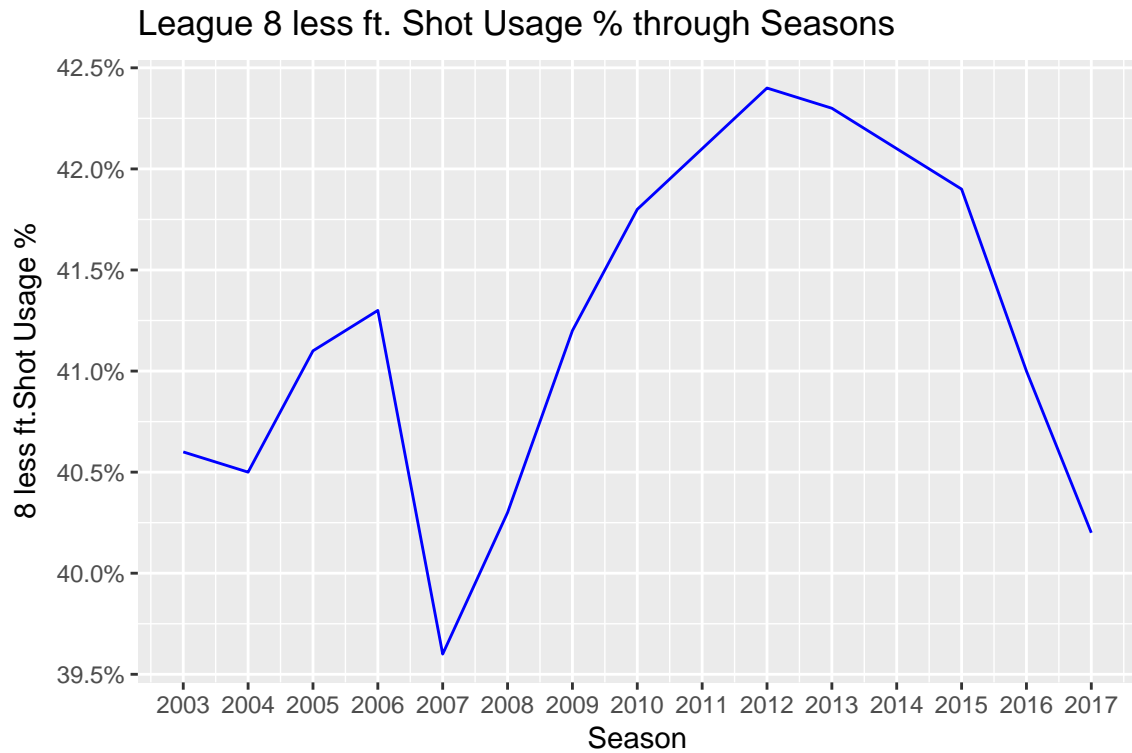


League 16–24 ft. Shot Usage % through Seasons

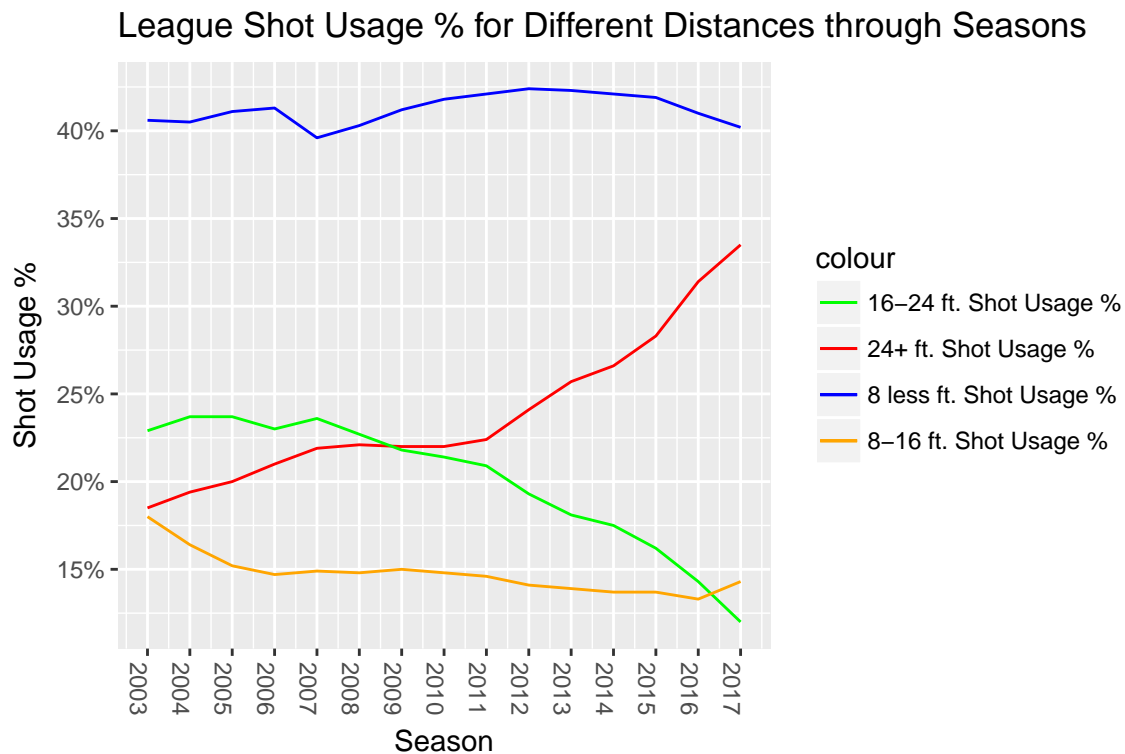


League 8–16 ft. Shot Usage % through Seasons





In the last graph of less than 8 ft shot usage, it surprisingly shows “huge fluctuation”. Later I guess the reason might be the scale of y axis is different in the four graphs above, so they could show different scale of fluctuation. To solve this, I put them into one graph as the following.



Now we can see the blue line (less than 8 ft shot usage) doesn't fluctuate a lot. Compared to other three lines, it seems the most stable one. This is reasonable, because attacking to the rim is always the major offensive method for players. Layup, dunk, cut, and post move are always the most efficient offensive method with high field goal percent. So even 3 point shooting becomes more and more popular, teams and players don't lose interest of paint attack.

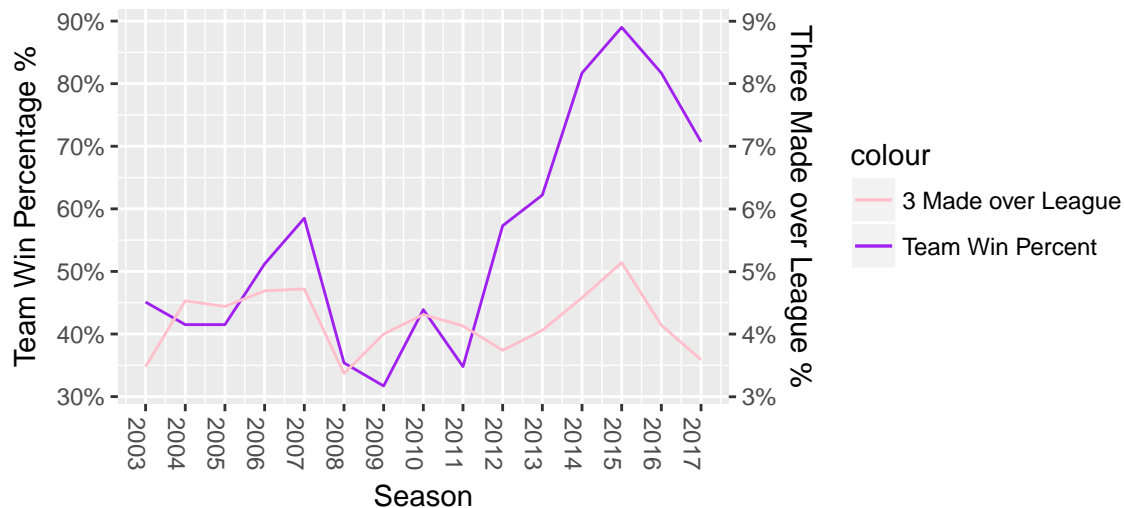
On other hand, we can see the green line keeps decreasing since 2007-2008 season. This means long midrange jump shot is becoming less and less popular today. One explanation is that, teams and players know that is hard to gain a high field goal percentage in midrange shot and stepping a few feet further could possibly make one more point, so the "price-performance ratio" of midrange jump shot is not high in their mind.

## Three Point Shooting vs Team Win Percent

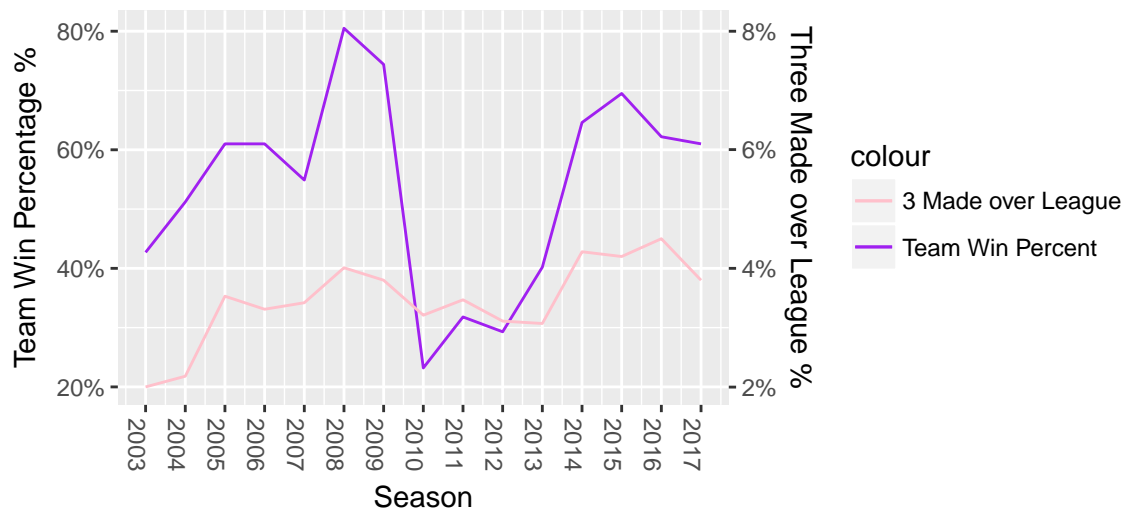
In order to see how 3 point shooting can effect a team's performance. I create a new feature called "Three Made over League" which is the percentage of onw team's total number of 3 point made over the total number of 3 point made by all teams in one season. Then I can build graph to see both the change of a team's "Three Made over League" through seasons and the change of its regular season win percent through seasons.

The following two graphs show the changes of the Golden State Warriors' and Cleveland Cavaliers' "Three Made over League" and win percent.

GSW's Team Win Percent and Three Made over League Percent thi

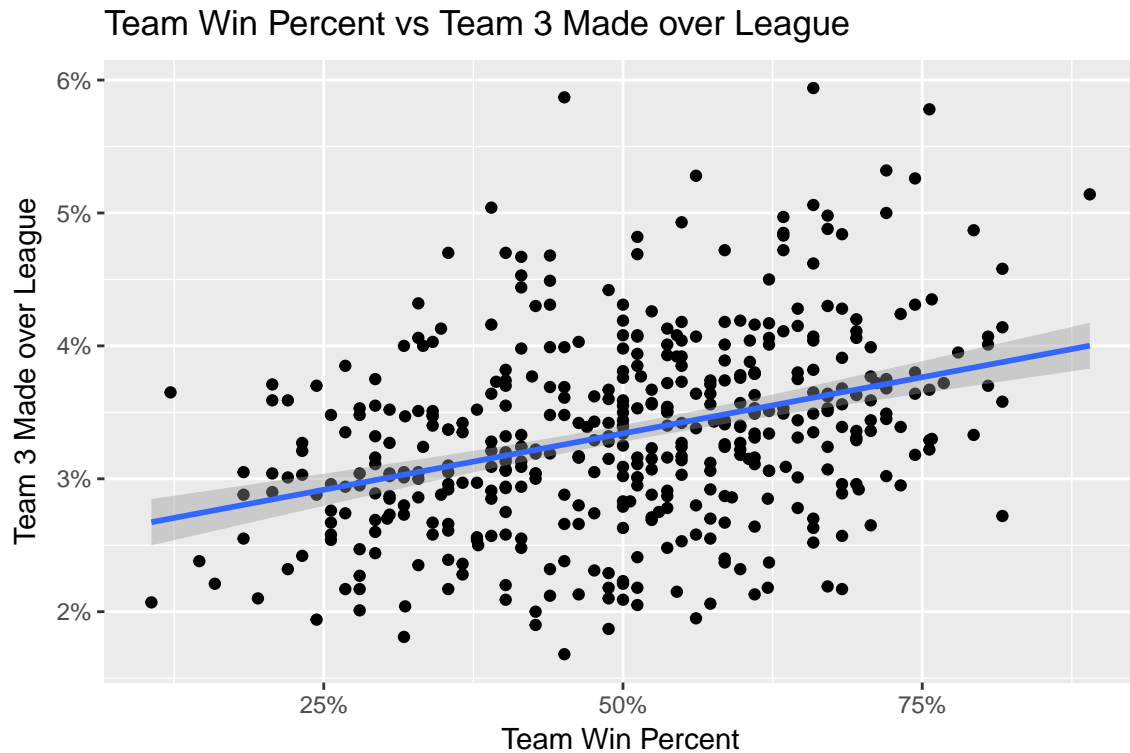


CLE's Team Win Percent and Three Made over League Percent thr



From the above graphs of two teams, we can see that "3 Made over League" and "Team Win Percent" usually go up or down in the same trend. This may imply that "3 Made over League" and "Team Win Percent" have a positive correlation. To further prove and show this hypothesis, I build a scatter plot of every team's "3 Made over League" against "Team Win Percent" in recent 15 years.





By building a best fit line, we do see a positive slope. To further understand the correlation between Team Win Percent and Team 3 Point Made over League, I do the Pearson correlation test.

```
cor.test(league_table_with_three_over_league$Season_Win_Per, league_table_with_three_over_league$three_made_over_league)

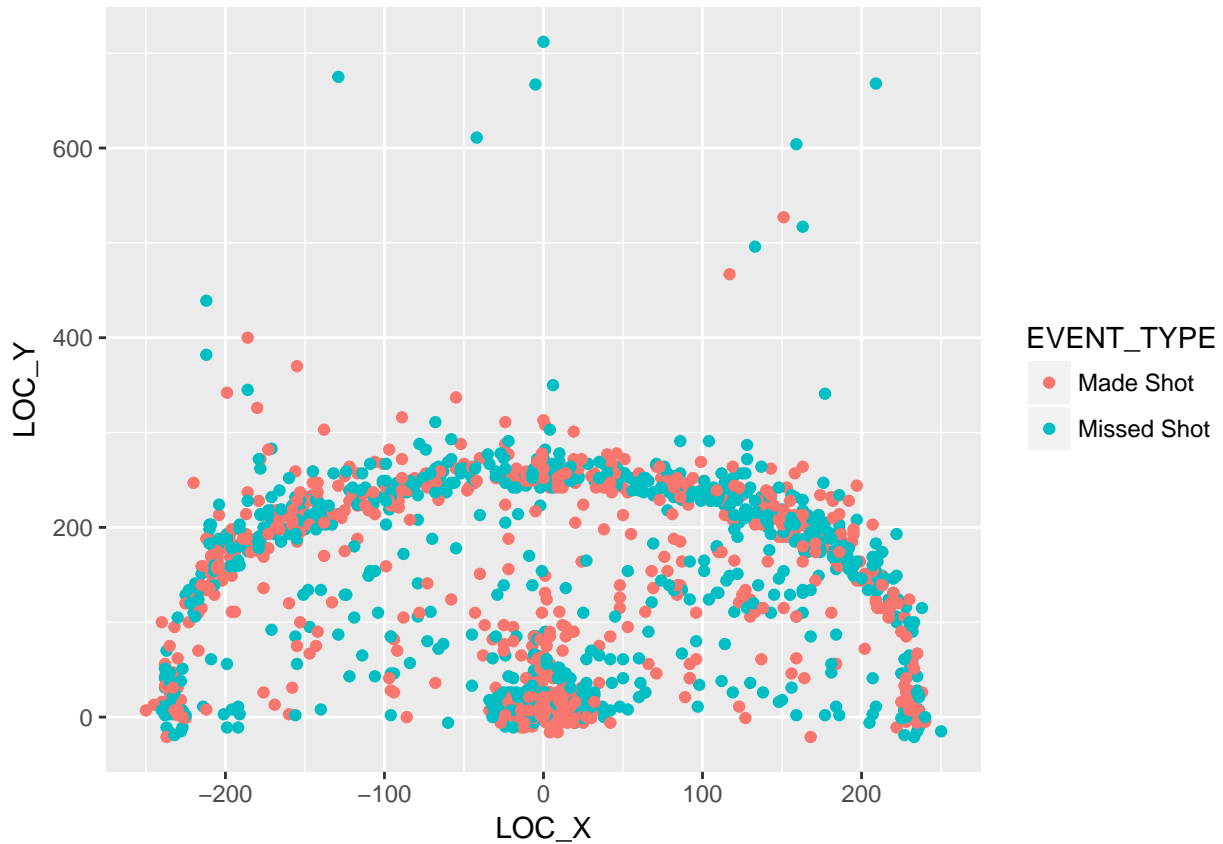
##
## Pearson's product-moment correlation
##
## data: league_table_with_three_over_league$Season_Win_Per and league_table_with_three_over_league$three_made_over_league
## t = 8.0577, df = 447, p-value = 7.152e-15
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2725694 0.4343545
## sample estimates:
##      cor
## 0.3561279
```

The test gives us a correlation coefficient of 0.3561279, so I could conclude that Team Win Percent and Team 3 Point Made over League do have moderate positive correlation. Therefore, I do think the more 3 point shot a team makes, the more game it can win in a season.

## Side Project & Future Works: Shooting Chart

Because the original shooting data I get from NBA API have features that record the coordinates of a shot a palyer made on the court, I can use ggplot function to build a shooting plot chart of a player of team in a season. One example is as the following shooting chart of Stephen Curry in his second MVP season (2015-2016 season).

This is a rough chart of Curry's shot plot.



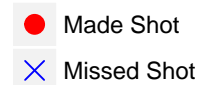
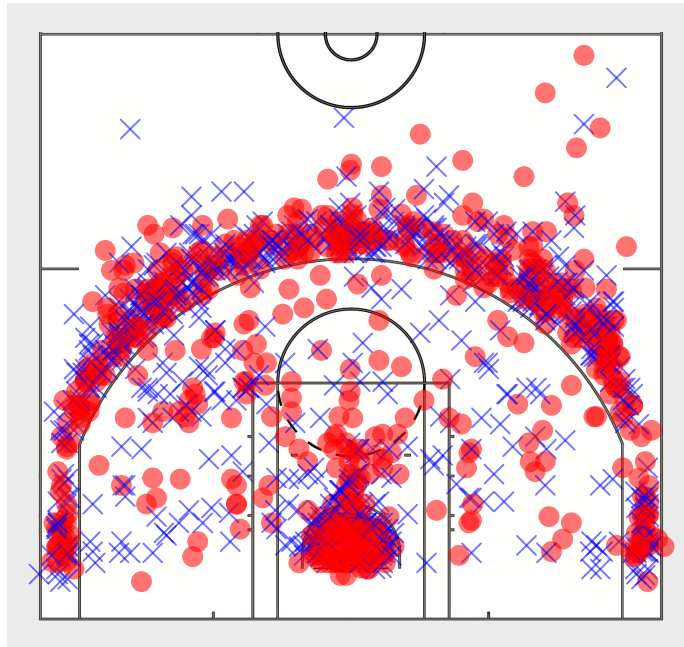
After beautifying process, I can compute the following shot chart.

```
## NULL
```

## Shot Chart

**Stephen Curry**

**2015–2016 Season**



I feel designing shot chart is very interesting and intriguing, so I also want to make a shot chart for a whole team in one season. However because a team makes too many shots in a season, points almost cover everywhere in the chart. Then we can hardly see any shooting pattern of a team from scatter plot. Inspired by <http://toddschneider.com/posts/ballr-interactive-nba-shot-charts-with-r-and-shiny/>, I will further study how to build Hexagonal Maps and Heat Maps because they can better show the field goal percentage and hot zone of a player or a team.

## Conclusion

By importing JSON structure data from NBA official APIs, cleaning online downloaded data, and creating features, I create my ideal data that records the shot information of every player and every team in recent 15 NBA regular seasons. Through several ggplot graphs and correlation test, I conclude that:

1. 3 point shooting indeed is becoming more and more popular in the whole league in recent years;
2. On the other hand, long midrange shooting is becoming less and less popular;
3. But as the most efficient way of scoring (this may need further proof in the future), 2 point shooting in paint area always has the highest usage percent in the league;
4. Team Win Percent and Team 3 Point Made over Total League 3 Point Made Percent have positive correlation, so making more 3 point shots can somehow bring a team more wins.

However, there are more things I could improve and do more works in the future:

1. High correlation doesn't mean absolute causality, so I need to find more useful features and better statistical modeling method to study the relationship between 3 point shooting and a team performance.
2. There are so many data and features produced in a basketball game. I don't consider some important features in my project, such as free throws, rebounds, turnovers, home game or away game, and defensive ratings from both a team self and its opponent.
3. Building player's and team's shot charts might be a better way of using the real time NBA shooting data. It could bring me more information than normal line charts and scatter plots.

As a NBA and basketball fan, I enjoy doing this project a lot. I will definitely keep doing data analysis about basketball in the incoming summer and future.