The developing playstyle of the NBA

COMP 30780 Data Science in Practice

Thomas Thornton (18466574),

School of Computer Science

University College Dublin

07/05/2021

In recent years there has been constant debate amongst sports pundits and basketball fans alike, about the current playstyle of the NBA. Some claim that teams have become too focused on offense and that they have slackened on the defensive side of the ball [1]. Others claim that players have simply gotten so efficient at scoring that it's borderline impossible to stop them [2]. These types of arguments raised questions in my mind. Do teams score more than in the past? Do they score more because they have become more efficient or because the pace of the game has risen? What has led to this rise in scoring? Do teams allow star players to take control of the team more so than in the past? These kinds of questions are too often answered with anecdotal evidence. I wish to definitively answer these questions using statistical evidence for the sake of basketball fans, but also for players and coaches that may be interested in how they can better replicate the performance of the top players in the NBA on their own teams. In this paper, using game data from the last two decades, I will show how teams have started to score more points per game and more points per possession each season. I will show how the increased popularity and proficiency of the 3-point shot has helped lead to this increase in scoring. I will use player data from 2003 to 2019 to analyse how team's use their star players and how their efficiency has changed over the years. I will also look at the kind of performance you should expect out of an NBA MVP (most valuable player). This paper will contribute to the ongoing discussions of how the sport of basketball is developing and will provide an analytical answer to some popular arguments amongst fans.

Declaration. I, Thomas Thornton, 18466574, declare that this assignment is my own work and that I have correctly acknowledged the work of others. This assignment is in accordance with University and School guidance1 on good academic conduct in this regard.

[1] See https://www.cs.ucd.ie/sites/default/files/cs-plagiarism-policy_august2017.pdf

1. Introduction

The American National Basketball Association (NBA) is the most popular basketball league in the world, drawing millions of viewers worldwide [3]. The NBA is widely considered the highest level of basketball in the world. Players come from Europe, Asia, and Africa to play in the NBA and viewers follow them. Like with any other sport debate rages about who the best players are and, even, who the best player of all time is. Sports pundits are constantly debating about how teams from the past would have fared in the NBA today and vice versa [4]. I have personally found some of these arguments quite interesting. It is fascinating to watch how players across generations have changed their style of play as society, as a whole, has learned more about the game and what we think the limitations of what a basketball player, and what a basketball team, can do, have changed. The main issue with these arguments is how they are often fraught with anecdotal evidence. Pundits frequently call upon their memories of how games played out when making claims about the sport rather than showing any evidence to support these statements [5]. I want to take a much more analytical look at how the league has changed over the years rather than trusting the memory of people I do not know. A side effect of this study is that it will provide a better understanding for fans, players and coaches that wish to learn more about how the best players in the world have adjusted their playstyle in recent years. I expect many players and coaches would be very interested in learning how to better replicate the performance of the players and teams in the NBA.

In this paper I cover the increase in scoring from the 2003 season onwards. I will also discuss the increase in pace and the change in defensive and offensive efficiency across the league in the same time frame. I set out to discover how the popularity of different kinds of shots have changed over the past 10 years. Using Granltand-style shot charts [6] and heatmaps I will show how teams have become more focused on the 3-point shot in the past decade. I will also show how players in different positions have changed the kinds of shots they take. You will also see the shots of Stephen Curry compared to the rest of the league. I believe Curry is largely responsible for the popularity of the 3-point shot and I will show his shots in comparison with the rest of the league as well as the heatmap from his unanimous MVP season [21]. Later you will see my findings of the importance of a team's 3-point percentage. It can be seen that in recent years the 3-point shot has a larger effect on a team's winning chances than it did in 2003. Finally, I will show how the game-by-game stats of top players have not changed as dramatically as you might think. Other than the increase in 3-point shots players play in a similar fashion to how they did in the past. Top players tend to dominate control of the ball as much as they did in 2003. We will also take a look at what kind of performance can be expected of MVP winning players and rookies of the year and how they compare to the average player in the league.

The remainder of this paper is organized as follows. In the next section we will take a look at the motivations that drove this paper and the questions that I set out to answer. After that I will discuss how I set about gathering and cleaning the data to answer these questions. Following this I will present my findings, explaining the approach I took to answer each question and the results I obtained from this analysis. A discussion of the findings and the implications of these findings will follow each result. After the results analysis I will discuss the overall findings and implication of this paper along with how it could be improved and built upon in future works. The final section of this paper will, naturally, be the conclusion. A bibliography can be found near the end of this paper. The final section is an appendix containing extra graphs that did not fit in my results analysis.

2. Motivations & Objectives

2.1. Background & Motivations

The idea for this project stems from my love of basketball. I started watching basketball at the beginning of the 2019-2020 season. At the beginning of the first COVID lockdown I started playing basketball regularly as a hobby and to stay healthy. I then started watching many more games and started paying attention to statistics of players and teams. The release of the Netflix documentary 'The Last Dance' [7] also piqued my interest about how teams from past years would have fared against modern teams and vice versa. Anyone that has been a fan of basketball at any point in the last 10 years I'm sure has heard the debate of whether Michael Jordan or LeBron James is the greatest player of all time, or whether the 2017 Golden State Warriors would've beaten the 1996 Chicago Bulls [4]. Of course, these arguments are impossible to resolve but it got me interested in how players and teams have changed their playstyles. One of the main changes in playstyle in the past 10 years is the increased focus on the 3-point shot. The current era of basketball is frequently referred to as 'the 3-point era' [8] due to this change in playstyle. My aim for this project was not to decide which playstyle is more enjoyable to watch but simply to understand why this fundamental change in the game has occurred and what influenced it.

2.2. Research Questions

2.2.0. RQ1: Have NBA teams become more focused or better at offence? Do teams score more than in the past?

The main focus of this question is to discover if teams really score more than in the past. If teams do score more than in the past is this due to a higher offensive efficiency or simply due to an increase in the pace of the game?

a) What caused the dip in scoring in 2011?
b) What caused the sharp rise in offensive efficiency since 2014?

2.2.1. RQ2: Has the popularity of different kinds of shots changed? Which shot ranges are the most popular?

This question aims to look at the overall change in shot popularity across the league. Players in different positions naturally play quite differently to each other so we will also look at how the shot popularity for each position has changed.

a) Can Stephen Curry's influence be seen in the shot data?
b) How has the shot popularity in each position changed?

2.2.2. RQ3: Does the number of 3-point shot attempts affect a team's win percentage? Does the percentage of shots made matter more than the number of shot attempts?

This question will focus on the importance of the 3-point shot. This era of basketball is frequently referred to as the 3-point era so how important is the 3-point shot to a team's winning chances really?

a) Is the 3-point shot more important now than it was in the past?

2.2.3. RQ4: How has the game-by-game stats of top players changed over the last 15 years? How does this compare to new star players?

This question aims to take a look at whether players really have gotten better over time. What remains consistent amongst the top players over time?

a) What is valued in an MVP?
b) How do MVPs compare to rookies of the year and the rest of the league?

3. Data Wrangling

3.1. Data Acquisition

There are three main datasets that are part of this project. The first dataset I found on Kaggle. It was made by a user called Nathan Lauga [9], who gathered his data from the NBA stats website [10]. This data contains a file called 'games.csv' which contains the result of each game from 2003 to 2019 including the number of points each team scored and the number of shots they attempted and other similar statistics. The second important file in this dataset 'games_details.csv' contains the performance of each player in each game from 2003 to 2019, e.g., points, field goals, turnovers, etc.

The second dataset was also found on Kaggle but unfortunately the creator of this dataset has seemingly deleted it since I downloaded it. This dataset contains the game data of each game from 1996-2019. I use this dataset for research question 3 because it is in a better format to answer this question.

The third and largest dataset I created myself. It contains the data of every shot from 2009 to 2019 as well as some stray extra data from seasons before 2009. To gather this data I used the python library 'nba_api' [11]. I used this library in conjunction with the 'players.csv' file from my first dataset. I created a list of each player and the team they played for each season from 2009 onwards and made a list of the unique pairs of player and team id. I then passed this list to the nba_api to gather the shot data of each player. This process takes a very long time to complete which is why I have saved this dataset in a local file. Each row in this dataset contains the information of a single shot i.e., who took the shot and where they took the shot from as well as if the shot went in.

3.2. Data Cleaning & Preparation

3.2.1. Cleaning

The cleaning process for each of these datasets wasn't particularly complicated but it was a bit tedious and tricky to keep track of the file names due to the similar nature of the data in some of the files.

3.2.1.1. RQ1

For research question 1 I used the raw datasets 'games.csv', 'games_details.csv' and 'teams.csv'.
'games.csv' contains the data of how each team performed in each game from the 2003 to 2019 seasons.
'games_details.csv' contains the data of how each player performed in each game from the 2003 to 2019 seasons.
'team.csv' contains the name and team ID of each team that played in each season from 2003 to 2019.

'games.csv': This dataset was straightforward to clean. There were not many empty rows in the dataset, but I filtered out the rows that were empty by removing any row where the home team didn't

have more than 0 points. There has never been a game in NBA history where a team has scored 0 points, so I was confident that I was not filtering out any proper data.
'games_details.csv': This dataset was filtered in a similar fashion. This dataset contains each players performance for each game, but it also contains the players data for the games they missed, whether they were injured or just not played by the coach. Whenever a player didn't play the 'MIN' column, which tracks the number of minutes a player played in a game, was null. I filtered out any rows with a null 'MIN' column. I then grouped this data by game ID and team ID to get a file of more advanced game data. This file became '010_details.csv'.
'teams.csv': This dataset did not require cleaning.

Files produced:
'001_base_games.csv': Contains the cleaned data of each game from 2003 to 2019 seasons.
'010_details.csv': Contains more advanced data of each game from 2003 to 2019 seasons.

3.2.1.2. RQ2

For research question 2 I used the raw datasets 'shot_dataset.csv' and 'players.csv'.
'shot_dataset.csv' contains the data of every shot from 2009 to 2019.
'players.csv' contains the player ID and team ID of the team they played for in each season from 2009 to 2019.

'shot_dataset.csv': I created this dataset myself using the python library 'nba_api'. To do this I created a list of the players and the teams they played for from the 'player.csv' dataset and passed these to the API. 'nba_api' reads from the NBA stats website [10]. When given a player ID and the team ID that they played for the API returns every shot that player took while playing for that team in a json format. I then converted this json to a pandas data frame.
I then added a season column by writing a function which read the GAME_ID column and returned the appropriate season. The game ID is an 8-digit integer where the 2nd and 3rd digit indicate the season i.e., 21100002 is a game from 2011.
If a player played before 2009 but also played in or after 2009 then the shots they took before 2009 were also stored in this dataset. After I created the season column I filtered out all shots before the 2009 season.
Finally, I removed all the columns that didn't provide any information. For example, the column 'GRID_TYPE' contained the string 'Shot Chat Detail' in each row. This was simply to reduce the size of the file as it is a very large dataset.

Files produced:
'002_shot_dataset.csv': Contains every shot from 2009 to 2019 seasons.

3.2.1.3. RQ3

For research question 3 I used the raw dataset 'NBA'.
'NBA.csv' contains the game data of every game from 1996-2018 season.

'NBA.csv': I used this dataset instead of the 'games.csv' dataset because it was structured better for this particular research question. The only cleaning this file needed was to change the season column. Each season was written as the last 2 digits of the year e.g., 2002 was written as 02. I wrote a function to correct this and applied it to the data frame.

Files produced:
'003_clean_3point.csv': Contains performance of each team from 1996 to 2018 seasons.

3.2.1.4. RQ4

For research question 4 I used the raw datasets ‘games.csv’ and ‘games_details.csv’.
Refer to section 3.2.1.1 for information on these datasets.

For this question I combined these two datasets in to one to get the performance of each player per game and, in the same row, have the performance of their team. This will be necessary later when calculating advanced statistics for individual players. Refer to section 3.2.2.4 for more information on this.

Files produced:
‘004_player_game_dataset.csv’: Contains the performance of each player in each game from 2003 to 2019 seasons.

3.2.2. Preparation

3.2.2.1 RQ1

For research question 1 I wanted to create multiple new columns, ‘POSSESSIONS’, ‘OER’ (offensive efficiency rating), ‘DER’ (defensive efficiency rating), ‘NER’ (net efficiency rating) and ‘CONFERENCE’. To do this I grouped my ‘games_details’ dataset by ‘team_id’ and ‘game_id’ and got the mean. This created a new dataset of each game with some extra data that the base game data didn’t have (Offensive rebounds etc.). I then joined this dataset with the base game dataset to add home and away team points. Using this data, I calculated the ‘POSSESSIONS’ column. There is a widely used formula to calculate possessions which can be found on the website NBAstuffer [12]. I then calculated the offensive efficiency rating (OER) and defensive efficiency rating (DER). The OER and DER are the number of points scored and number of points allowed per 100 possessions, these are also widely used statistics [19]. The net efficiency rating (NER) can be calculated by taking the DER from the OER. To add the conference data, I made a list of each team and the conference they play in then applied it to the dataset. I saved this dataset as ‘110_team_efficiency’. I then grouped this dataset by team and season and applied a function to correct the columns containing percentages to create a new dataset ‘111_seasonal_efficiency’.

Files produced:
‘110_team_efficiency.csv’: Advanced data on team performance in each game 2003-2019, including possessions, OER and DER.
‘111_seasonal_efficiency.csv’: Advanced data of how each team performed in each season 2003-2019, including possessions, OER and DER.

3.2.2.2 RQ2

There wasn’t much data preparation involved in research question 2. From the ‘games_details’ (for more information on this dataset check Section 3.2.1.1.) dataset I made a list of every player and the position they played i.e., guard, forward or center [20]. I then applied this list to the shot data frame. There was one issue with this, my players dataset only contained the position data for players that started the game, meaning I don’t have the position of any player that came off the bench. This is something I would like to correct in any future iterations of this study.

Files produced:
‘210_position_shot_dataset’: This file contains the shot data of every shot from 2009-2019 including a new column containing the position the player played.

3.2.2.3 RQ3

Research question 3 also didn’t have much data preparation involved. For this question I merged the ‘003_clean_3point’ dataset with itself to get the opposing team’s data on the same row as the first team. I then made 3 new columns ‘comp_3PM’, ‘comp_3PA’ and ‘comp_3P%’. These columns are

the comparative 3-point shots made, comparative 3-point shots attempted, and comparative 3-point shot percentage respectively. This means if the first team attempted 10 3-point shots and the second team attempted 12 3-point shots then the first team would have a comp_3PA of '-2'.

Files produced:
'310_3point_opp.csv': This contains the game data of each team from 1996-2018 as well a comparison to the opposing team's 3-point performance.

3.2.2.4 RQ4

In research question 4 I merged '004_player_game_data.csv' with '010_details.csv'. This provided me with sufficient data to calculate a variety of advanced player statistics. I calculated 3 new columns: 'USG%' (usage percentage [13]), 'TS%' (true shooting percentage [14]) and 'EFF' (efficiency [15]). Usage percentage tracks the percentage of plays a player is involved in while on the floor. True shooting percentage is the percentage of shots a player makes but it takes in to account the extra difficulty level and extra benefit of scoring a 3-point shot. Efficiency or EFF is a statistic that gives a numeric value to the performance of a player. It adds points for when a player does something good e.g., score or assist a teammate, and it removes points when a player does something bad e.g., miss a shot or turnover the ball. This statistic is arguably favored to players that play an offensive style of basketball, but defensive performance is very difficult to measure in basketball. This is a limitation that I would like to overcome in future iterations of this study.

Files produced:
'410_advanced_player_data.csv': This contains the performance of every player in every game from 2003 to 2019 including statistics that I calculated. (EFF, TS%, USG%)

4. Data Analysis & Results

Any figures labeled with an A e.g., Figure A.1, refers to a graph in the appendix, section 9.1.

4.1. RQ1

Have NBA teams become more focused or better at offence? Do teams score more than in the past?

4.1.1. Datasets

For this question I used the processed files '001_base_games.csv' which contains the basic data of each game. '111_seasonal_efficiency.csv' which contains team OER and DER and '110_team_efficiency.csv' which is used for pace.

4.1.2. Approach

Do teams score more than in the past?
For this question I created a line graph of the points scored by the home team and away team for each season.

Have NBA teams become more focused or better at offence?
For this question I used a parallel coordinates plot to show how offensive and defensive efficiency ratings have changed.

a) What caused the dip in scoring in 2011?
All this took was some googling.

b) What caused the sharp rise in offensive efficiency since 2014?
From my prior knowledge of basketball, I believe the cause of the rise in offensive efficiency was from the Golden State Warriors. To show this I plotted the Golden State Warriors OER against the league average from 2011 to 2018.

4.1.3. Results

Do teams score more than in the past?
And
a) What caused the dip in scoring in 2011?
Figure 1 shows the rise in the average points scored per game by home and away teams from 2003 to 2019. The NBA lockout, which will be discussed in section 4.1.4. is labelled on the 2011 season. The Golden State Warriors dynasty team is highlighted from 2014 to 2018. This team won 3 championships in 5 seasons and will be discussed more throughout this project.

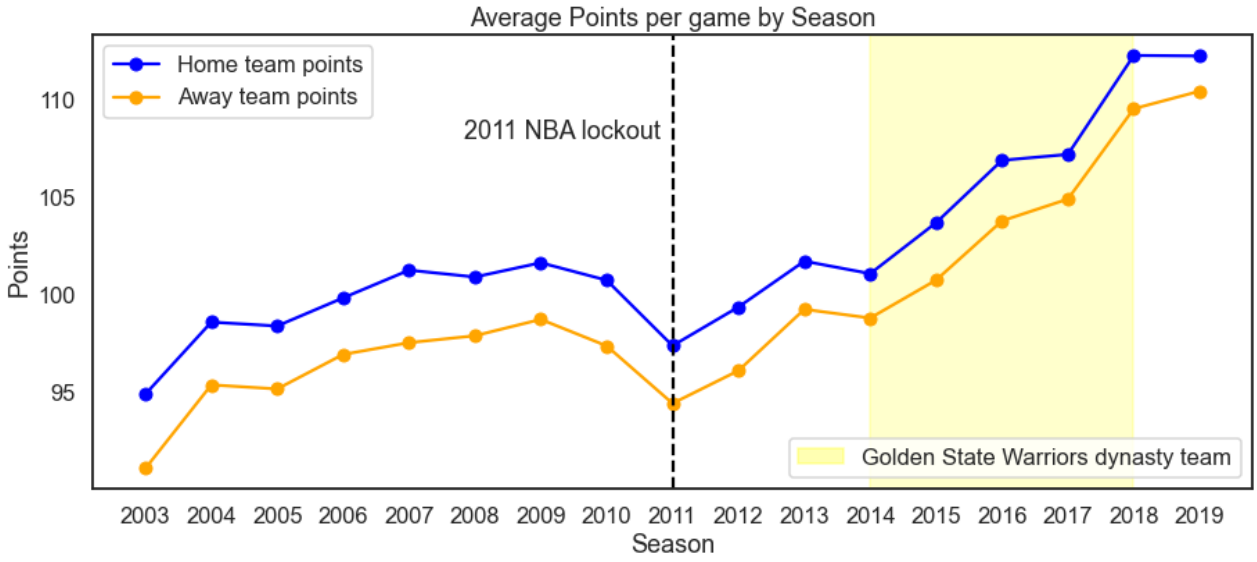

Figure 1. A graph of points per season of home and away teams.

Have NBA teams become more focused or better at offence?
Figures 2 and 3 show the rise in offensive efficiency and the decline in defensive efficiency respectively. The NBA is divided in to 2 conferences, East and West, with 15 teams in each conference. I have divided the offensive and defensive efficiency of each conference. The faded lines in the background show the change in efficiency of individual teams in each conference.

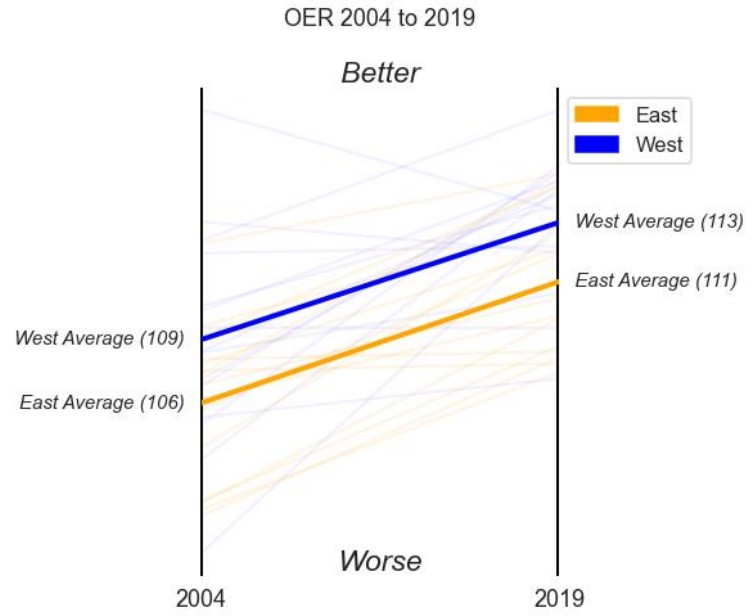

Figure 2. Offensive efficiency 2004 to 2019 by conference

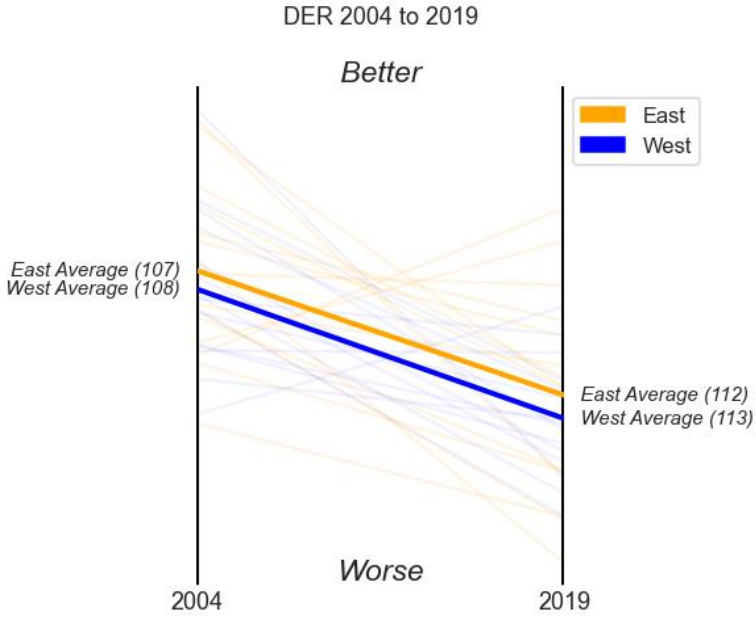

Figure 3. Defensive efficiency 2004 to 2019 by conference

b) What caused the sharp rise in offensive efficiency since 2014?

Figures 4 and 5 below show the offensive efficiency and defensive efficiency of the Golden State Warriors in comparison to the league average. Figure 4 shows the sudden rise in offensive efficiency from the 2014 season onwards. I believe the reason for the rise in offensive efficiency across the league is largely due to teams copying the playstyle of the Warriors but I will discuss that more in depth in section 4.1.4. and 4.2.4. Figure 5 shows that the Warriors dynasty initially had a very strong defence but came down to around league average while still remaining title contenders.

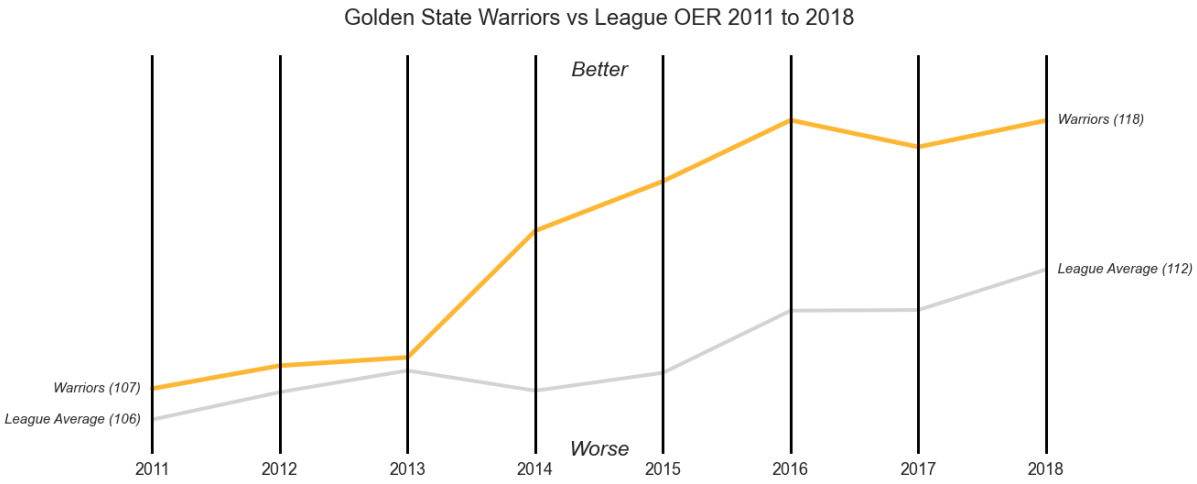

Figure 4. Offensive efficiency of the Golden State Warriors compared to the league average 2011-2018

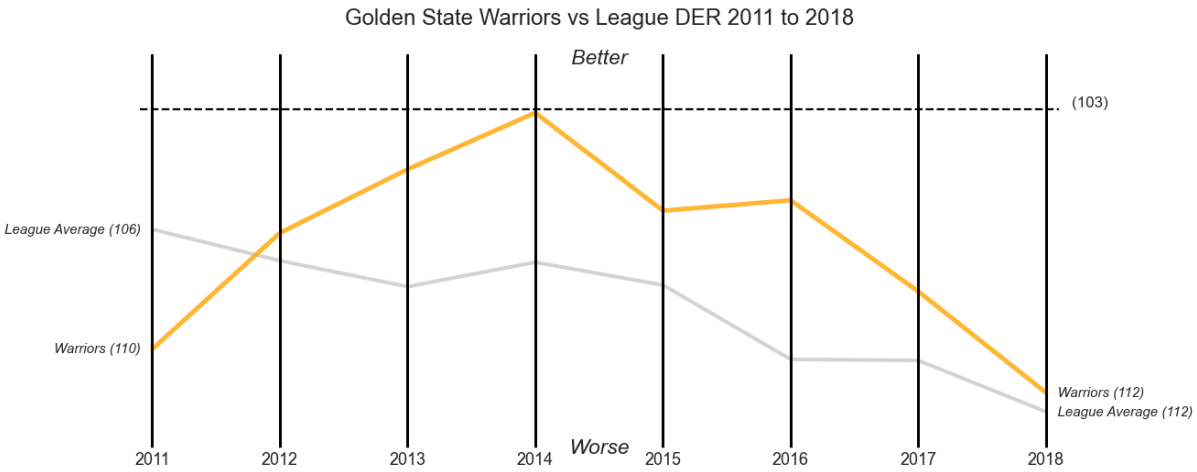

Figure 5. Defensive efficiency of the Golden State Warriors compared to the league average 2011-2018

4.1.4. Discussion

As you can see from Figure 1 there has been a consistent rise in the number of points scored by home and away teams since the 2003 season. The one exception to this is the 2011 season. As you can see on the graph, I have annotated the 2011 NBA lockout [16]. This lockout was caused by the expiration of a contract that decided what percentage of revenue the league would get and what percentage the players would get. While the league and the players were negotiating a new contract players were prohibited from using team facilities such as gyms and courts. They were also prohibited from practicing with teammates. When the 2011 season finally began players missed out on months of preseason training and the season was 20 games shorter than a regular season. This is the clear cause for the drop in quality in the 2011 season.

The next part of Figure 1 I wish to draw your attention to is the yellow highlighted section labeled ‘The Golden State Warriors dynasty team’. This team won 3 championships in 5 years and also holds the record for most wins in an NBA regular season [17]. I will come back to the Warriors at the end of this section.

To find out if teams have become more focused or better at offence, we cannot simply look at the increase in points per game. This is because the pace of the game has increased as you can see in

Figure A.1 in section 9.1. This graph shows that the pace of the game has also been increasing each season. This means that teams play faster and attempt more shots per game. A faster pace naturally leads to an increase in scoring.

To get a better idea of how much better teams have really gotten at offence we need to look at their offensive efficiency. A team's offensive efficiency rating is simply the number of points they are expected to score per 100 possessions. Conversely a team's defensive efficiency rating is the number of points they are expected to allow per 100 possessions. In Figure 2 and Figure 3 above you can see that across the league teams have become more efficient on offense which has naturally led to a drop in defensive efficiency. By the fact that offensive efficiency is adjusted for pace we now know that teams both play faster and score more per possession than they have in the past.

As I mentioned earlier, I believe the reason there was such a sharp rise in offensive efficiency from 2014 is due to teams trying to replicate the new style of basketball that the Golden State Warriors began playing in 2014. In Figure 4 and Figure 5 above you can see the OER and DER of the Golden State Warriors from 2011 to 2018 compared to the league average.

The Golden State Warriors won the NBA championship in the 2014 season, 2016 season and the 2017 season. They were the runners up in the 2015 season and 2018 season [18]. While it is clear the Golden State Warriors dynasty team in 2014 initially had an excellent DER it started to come back to around the league average over the following years. However, the OER was consistently far beyond the league average. In the 2017-18 season (2017 on the graph) when the Warriors had a DER close to the league average, they still went on to win the NBA championship. I believe other teams around the league saw how the Warriors were playing and started to emulate this style of play. In Figure 6 in Section 4.2.3 I will discuss how the Warriors began focusing more on the 3-point shot. I believe the growing popularity of the 3-point shot is what led to the rise in OER.

Clearly teams are scoring more than they did in the past. They are playing faster but also scoring more points per possession than any other period in the last 18 years.

About team's defenses: These results clearly prove that teams have gotten more efficient at offense, but I don't believe teams have lost focus on defense. Due to the fact that teams are scoring more of their points as 3-pointers it has put strain on the defense of teams. Instead of only covering the court inside the 3-point line they now must guard players at the 3-point line and even beyond it. This opens a lot of space for attackers to move around inside the 3-point line. I believe teams haven't gotten worse at defense; it has simply become more difficult to stop teams from scoring. This is a similar conclusion that Justin Verrier from The Ringer [2] came to.

4.2. RQ2

Has the popularity of different kinds of shots changed? Which shot ranges are the most popular?

4.2.0. References

The code to produce the shot charts and heatmaps in this section were written by GitHub user 'eyalshafran'. I have made slight tweaks to the code to fit the data that I have gathered but it was almost entirely written by eyalshafran.

Name: eyalshafran,
Date: 06/11/2017,
Title: NBAapi/plot.py,
URL: https://github.com/eyalshafran/NBAapi/blob/master/NBAapi/plot.py

4.2.1. Datasets

For this research question I used the processed file '210_position_shot_dataset' which contains data on the location of every shot, whether it went in, who took it, and, in most cases, what position they were playing.

4.2.2. Approach

Has the popularity of different kinds of shots changed?
To answer this question, I plotted a Grantland-style chart [6] to visualize how the popularity and proficiency of shots have changed from 2009 to 2019. The graphs from the next segment of this research question also provide some answers to this question. I also plotted the percentage of shots made by the Golden State Warriors being 3-point shots in comparison with the rest of the league.

Which shot ranges are the most popular?
To answer this question, I produced heatmaps of all shots from each season from 2014 to 2019. I have filtered all shots from the restricted area in each heatmap. This is because over every season the restricted area is by far the most popular shot location and it took all detail away from the rest of the graph. I filtered the restricted area out so we could get a better idea of how every other shot location has changed in popularity

a) Can Stephen Curry's influence be seen in the shot data?
In order to answer this question, I first visualized Stephen Curry's shot data for the 2015-16 season in which he won unanimous MVP by using the Grantland chart. I then plotted the heatmap for Stephen Curry this season and the league's shot data for each season from this season onwards.

b) How has the shot popularity in each position changed?
For this question I plotted a comparison of the shots by guards and forwards for 2019 compared to 2009. I also plotted the heatmap for each position in 2009 and 2019.

4.2.3. Results

If you are unfamiliar with a basketball court you can see a basic diagram in the appendix. Figure A.2.

Has the popularity of different kinds of shots changed?

Let's first look at the increase in popularity of the 3-point shot in Figure 6 below. This graph shows us that the 3-point shot has grown in popularity since 2011. The Golden State Warriors realised before the rest of the league how effective the 3-point shot can be.

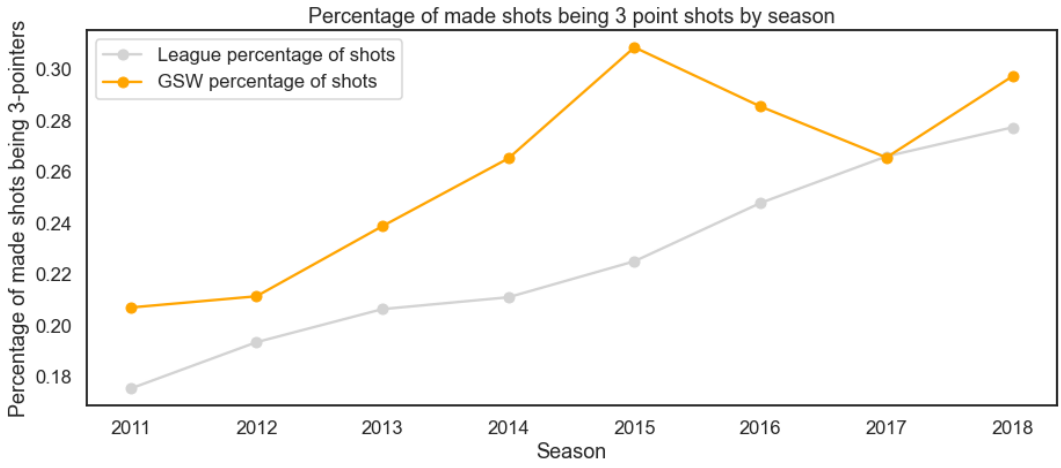

Figure 6. Percentage of made shots being 3-pointers, Golden State Warriors vs League Average

Figure 7 below shows how the popularity of certain shot locations have changed in the last 10 seasons. On this graph the larger the hexagon, the more shots were taken in that location. This compares shots from 2019 to 2009. A large hexagon means more shots were taken from that location in 2019 than in 2009. If there is no hexagon it means there were fewer shots from this position in 2019. If the hexagon is close to red it means the shots from this region were scored at a higher percentage in 2019 than in 2009. Closer to blue means they were shot at a lower percentage than 2009. It appears that 3-point shots have grown significantly in popularity. Shots were also mostly made at a higher percentage in 2019 than they were in 2009.

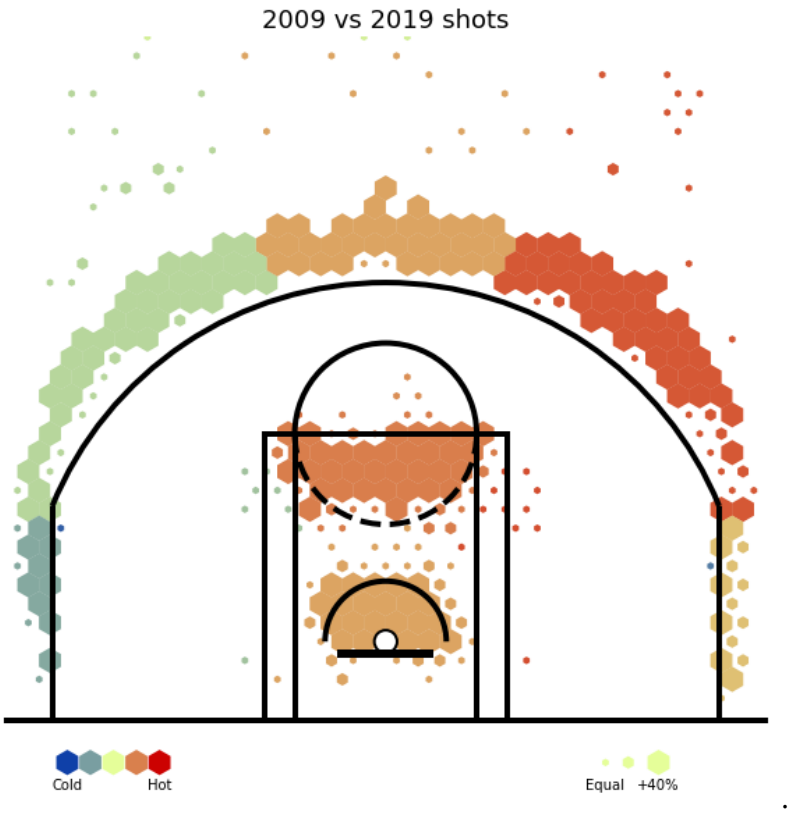

Figure 7. Shot comparison 2009 vs 2019

Which shot ranges are the most popular?

In Figure 8 and Figure 9 below you can see a side-by-side comparison of two heatmaps of the shots in 2009 and 2019 respectively. There is a very obvious difference in the popularity of the 2-point shot and 3-point shot in these graphs. There are very few shots from midrange in 2019 and a much higher concentration of 3-point shots. Long range 2-point shots are almost non-existent in Figure 9.

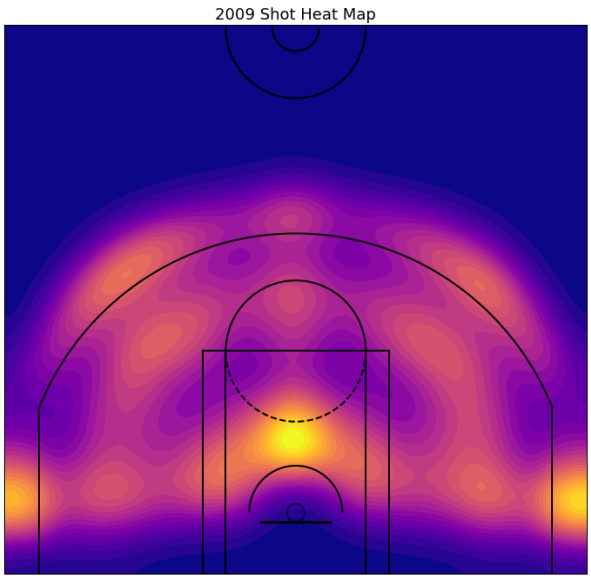

Figure 8. Shot heatmap 2009

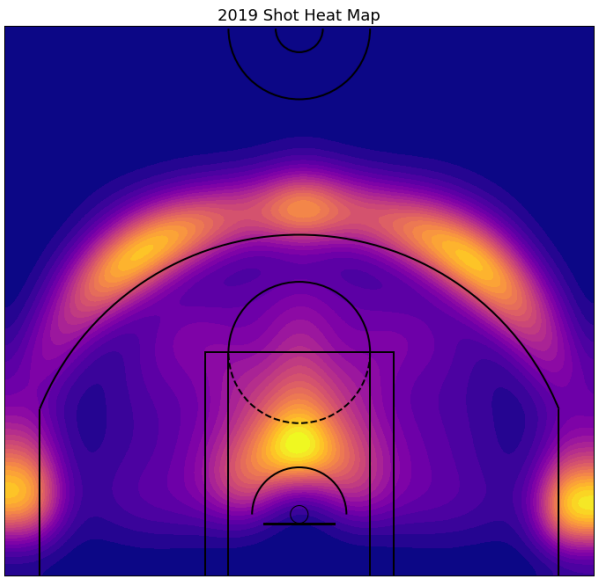

Figure 9. Shot heatmap 2019

a) Can Stephen Curry's influence be seen in the shot data?

Figure 10 below shows the shot chart of Stephen Curry in his unanimous MVP season compared to the rest of the league. It is clear that Curry shot far more 3-point shots than the rest of the league and scored at a higher percentage from almost all locations. If you wish to see the heatmaps of the league's shots from the 2015 season onward look for Figures A.3 to A.6. Figure 11 shows the heatmap of Stephen Curry for the 2015-16 season. Each league heatmap after this year begins to look more like Figure 11.

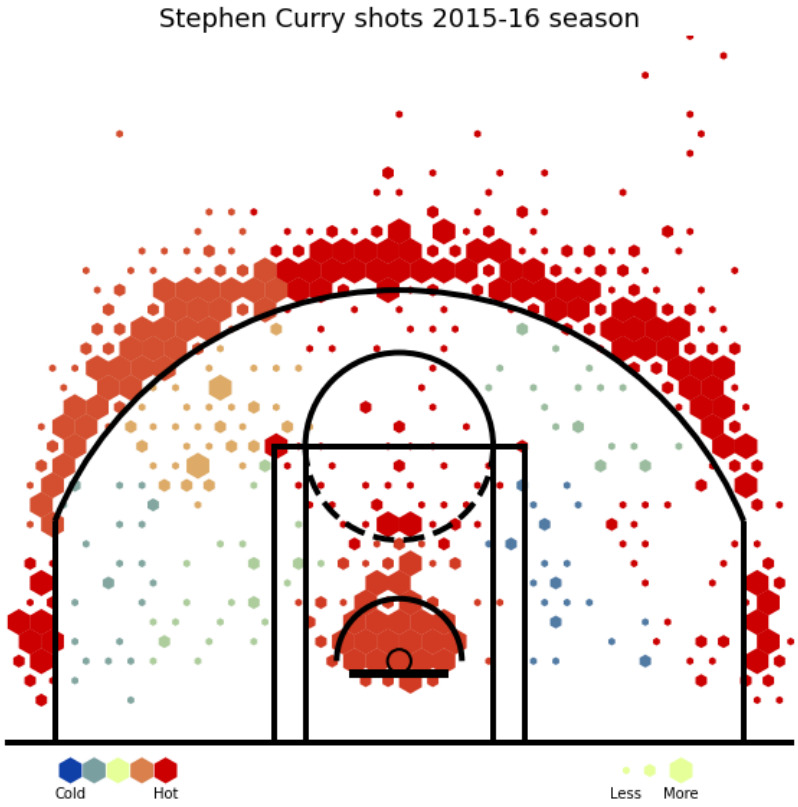

Figure 10. Stephen Curry shots 2015-2016 season compared to the rest of the league

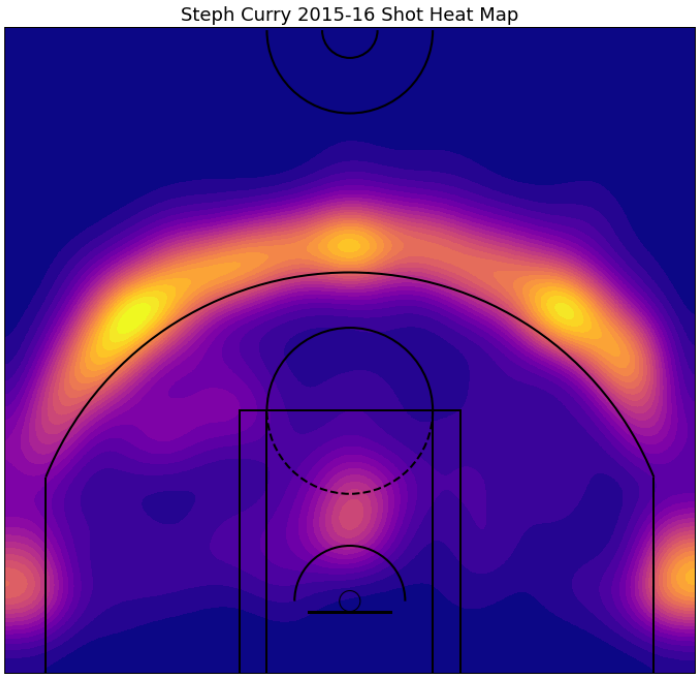

Figure 11. Stephen Curry heatmap 2015-16

b) How has the shot popularity in each position changed?

Figures 12 and 13 below show the shots of guards and forwards, respectively, from 2019 compared to 2009. The heatmap correspondents of these graphs can be found in the appendix, Figures A.7 and A.8 for guards, Figures A.9 and A.10 for forwards. I also have heatmaps for the centers here, Figures A.11 and A.12. We can see that the 3-point range has become more popular in each position.

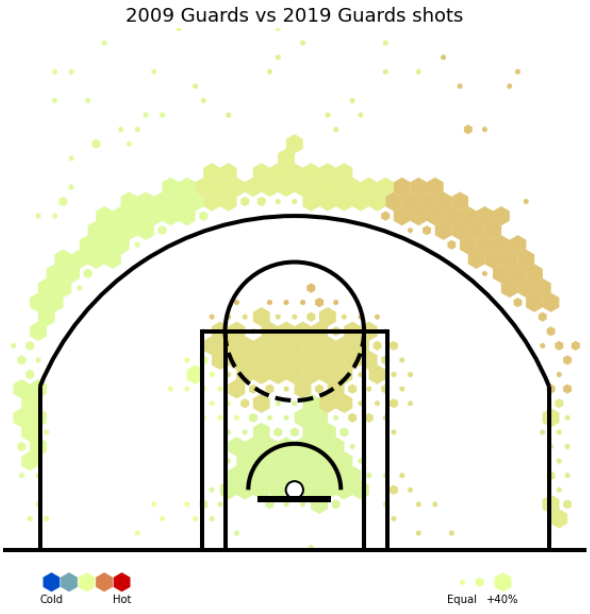

Figure 12. Guards shot chart 2019 compared to Guards 2009.

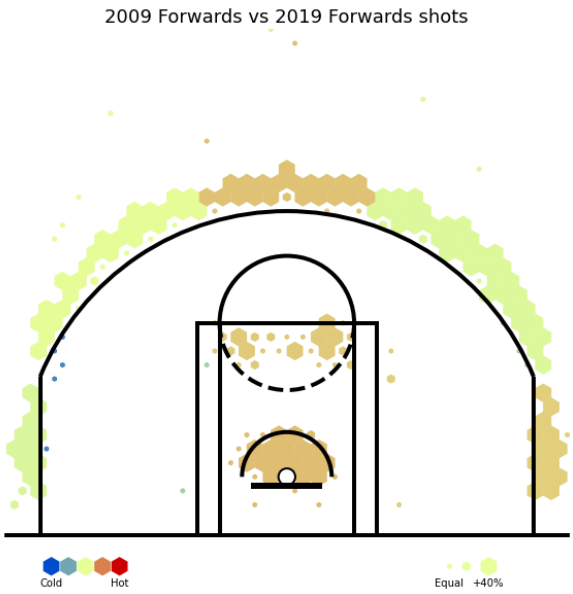

Figure 13. Forwards shot chart 2019 compared to Forwards 2009

4.2.4. Discussion

Has the popularity of different kinds of shots changed?

It is very clear from all of these graphs that there has been a shift in the popularity of shots. Since 2015 the 3-point shot has significantly grown in popularity and naturally the 2-point shot has fallen out of popularity a bit. As I mentioned in research question 1, I believe the Golden State Warriors are responsible for the rise in OER and I believe their use of the 3-point shot is a key part in that. In Figure 6 we see that the Warriors started scoring a very high percentage of their shots as 3-point shots.

We can see the after effect of this shift in popularity in Figure 7. The 3-point shot is clearly far more popular than it was in 2009. I was not surprised by this but I was surprised that the percentage for some of these shots has dropped. After some consideration I believe I have figured out the reason for this. In basketball there is a position called the shooting guard which is, traditionally, the best shooter on the team. The shooting guard mostly plays at around the left side of the court, where the shots are at their lowest percentage in Figure 7. Shooting guard used to be a very specialized position that was highly focused on being a good shooter. They would mostly take shots when undefended to ensure a good success rate. In the NBA today shooting guards often take many contested shots and this is what I believe caused the drop in percentage of shots made from this area. I would like to test this hypothesis in future studies but I currently do not have the data of the position of defensive players for each shot so this is not yet possible for me.

Which shot ranges are the most popular?

These graphs show us similar results to those in the first part of this question. The difference in shot popularity from 2009 to 2019 is extremely clear. 2-point shots were extremely popular and 3-point shots were shot at a similar frequency to the 2-point shot. In 2019 3-point shots are shot much more frequently. You can also see that long range 2-point shots are almost non-existent in 2019. This is because players realised that by taking long range 2-point shots they were essentially taking a 3-point shot but being rewarded with 1 less point.

These results somewhat confirmed my hypothesis. Something I was very surprised by was the number of shots in the restricted area. As a basketball fan I'm aware that there are many shots taken from this location as layups or dunks, but I was surprised by how many. In 2009 the only place that was lit was the restricted area due to the frequency of shots from this location. I expected 2019 to show some highlights around the 3-point arc but it produced an almost identical graph to the 2009 graph but with very slight highlights at some points on the 3-point arc.

a) Can Stephen Curry's influence be seen in the shot data?

I have discussed the Warriors a few times now and the favour that they show to the 3-point shot. The player that led the way on this team was Stephen Curry. Stephen Curry is famous for his 3-point shooting and he just broke the record for the most 3-point shots scored in a month [22] as of writing this. Figure 10 shows his unanimous MVP season. The MVP is decided by the votes of 100 sports reporters. In the 2015-16 season Stephen Curry got 100 1st place votes for NBA. This is the only time in history this has happened. His proficiency at scoring was a key part in this and we can see that he heavily favoured the 3-point shot. If you follow the Figures A.3 to A.6 you can see how the league followed suit of Stephen Curry in subsequent seasons. Each season resembles Figure 11 more and more. I believe Stephen Curry's influence on the NBA can be seen from these graphs.

b) How has the shot popularity in each position changed?

Figures 12 and 13 confirm what must be expected at this point. We can see that guards and forwards have both followed the trend of the league. Both positions shooting more shots from beyond the 3-point arc than in 2009. What I was surprised by was the drastic change in shots for forwards which

you can see in Figure A.9 and A.10. I expected forwards to shoot slightly more shots from the 3-point range than they did in 2009 but I didn't expect a change this drastic. I expected the guards to be responsible for the massive shift to the 3-point line because guards are, traditionally, better shooters than forwards. This goes to show just how important the 3-point shot has become because even forwards have drastically changed their focus.

4.3. RQ3

Does the number of 3-point shot attempts affect a team's win percentage? Does the percentage of shots made matter more than the number of shot attempts?

4.3.1. Datasets

For this research question I used the processed file '310_3point_opp'. (Section 3.2.2.3.)

4.3.2. Approach

Part a. (Is the 3-point shot more important now than it was in the past?) is answered as part of both of these questions.

Does the number of 3-point shot attempts affect a team's win percentage?
My idea for this question was initially to plot the number of 3-point shots attempted against the teams winning margin. I then realised that if a team attempted 30 3-point shots and lost that it could be possible their opponent shot 40 3-point shots. To solve this, I created new columns which compared the team's 3-point stats to their opponents 3-point stats and I plotted these columns instead.
I plotted a graph with the comparative 3-point shots attempted on the x-axis and comparative 3-point shot percentage on the y-axis. I colour coded wins and losses as green and red and size was related to the margin the team won by

Does the percentage of shots made matter more than the number of shot attempts?
For this question I created a small multiples graph which shares an x-axis. Winning margin is plotted on the x-axis. The first graph plots 3-point percentage on the y-axis, the 2nd graph plots 3-points attempted and the 3rd graph plots 3-points made. In this way you can compare which 3-point statistic is most important.

4.3.3. Results

Does the number of 3-point shot attempts affect a team's win percentage?
Does the percentage of shots made matter more than the number of shot attempts?

Figure 14 below shows each game of the 2018 season with each team's comparative 3-point percentage on the y-axis and comparative 3-point shots attempted on the x-axis. There is a clear correlation between wins and 3-point percentage but there is no distinct correlation between 3-points attempted and wins.

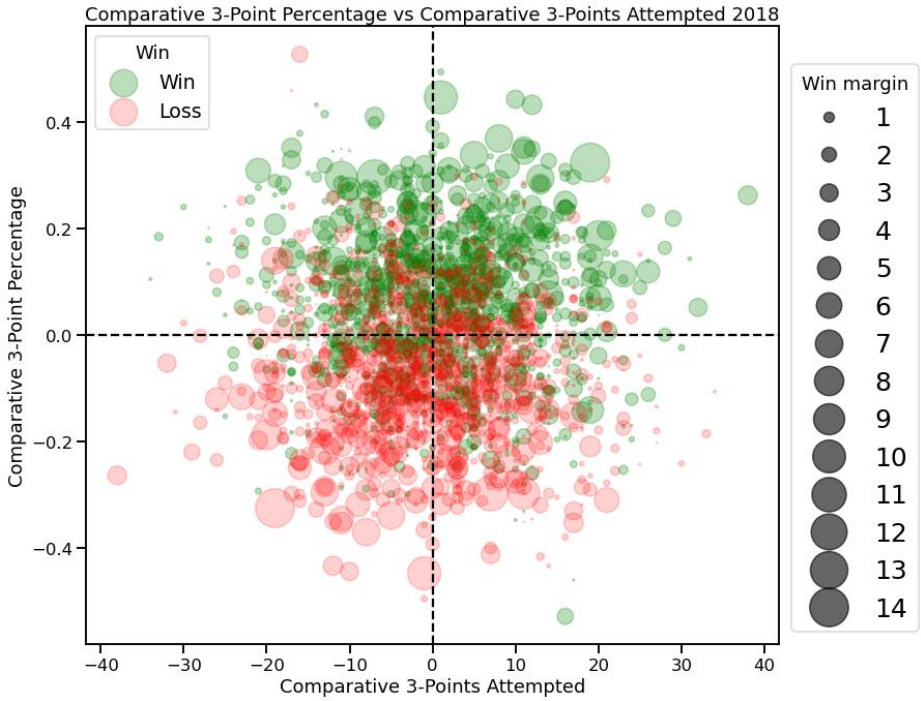

Figure 14. 3-point percentage compared to 3-points attempted with winning margin as size and wins and losses are green and red respectively.

Figures 15 and 16 below compares 3-point percentage to 3-point shots made. These statistics are again in comparison to the percentage and made shots of the team's opponents. Figure 15 shows every game of the 2003-04 season. Figure 16 shows the same plot but for the 2018-19 season. We can see that if you make more shots than your opponent and score them at a higher percentage than your opponent there is a much better chance of your team winning.

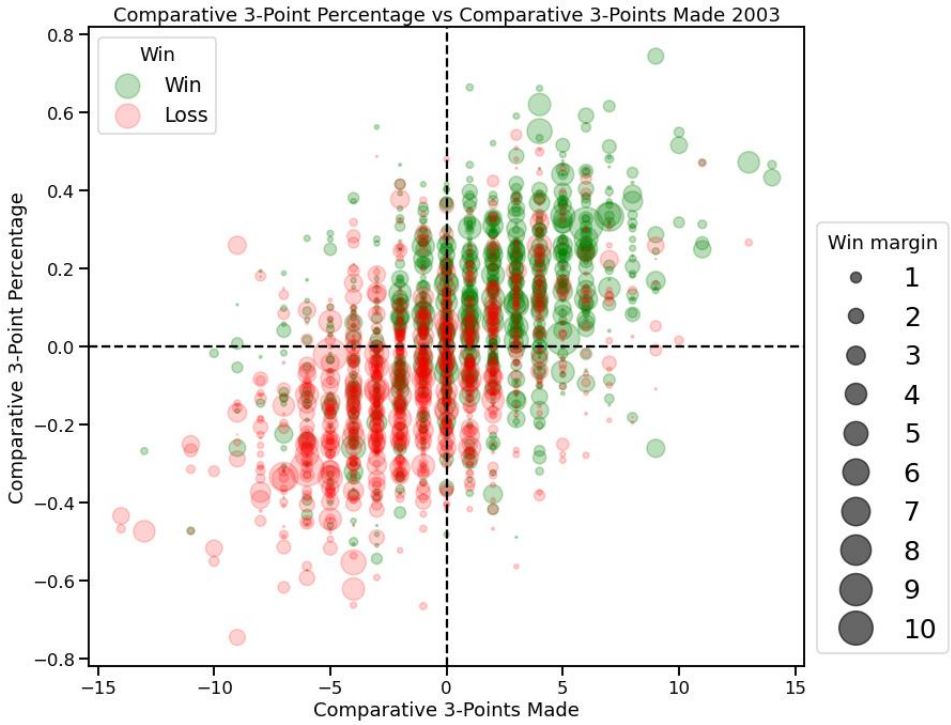

Figure 15. 3-point percentage vs 3-points made compared to opponents for 2003-04 season. Size indicates winning margin. Green indicates win and red indicates loss.

In Figure 16 below you can see that in the top right quadrant your team is significantly more likely to win. We can also see that dots further up and to the right have a higher winning margin than anywhere else on the graph. Clearly there is a stronger corelation of being in the top right quadrant and a team winning in the 2018 graph (Figure 16) than in the 2003 graph (Figure 15).

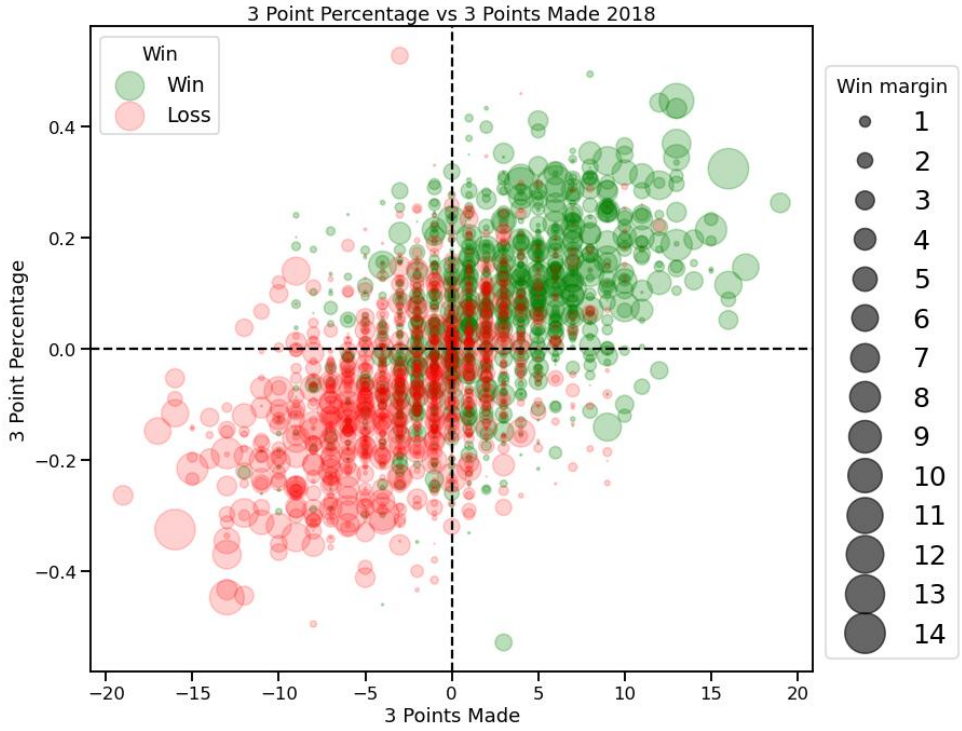

Figure 16. 3-point percentage vs 3-points made compared to opponents for 2018-19 season. Size indicates winning margin. Green indicates win and red indicates loss.

Figures 17 and 18 below show a comparison of the 3-point percentage, 3-point attempts and 3-points made. Winning margin is plotted on the x-axis. The line of linear regression is overlayed on each graph with the correlation coefficient next to it. We can see that shots made and shot percentage both show a positive correlation with winning margin. Shots attempted has virtually no correlation with winning margin.

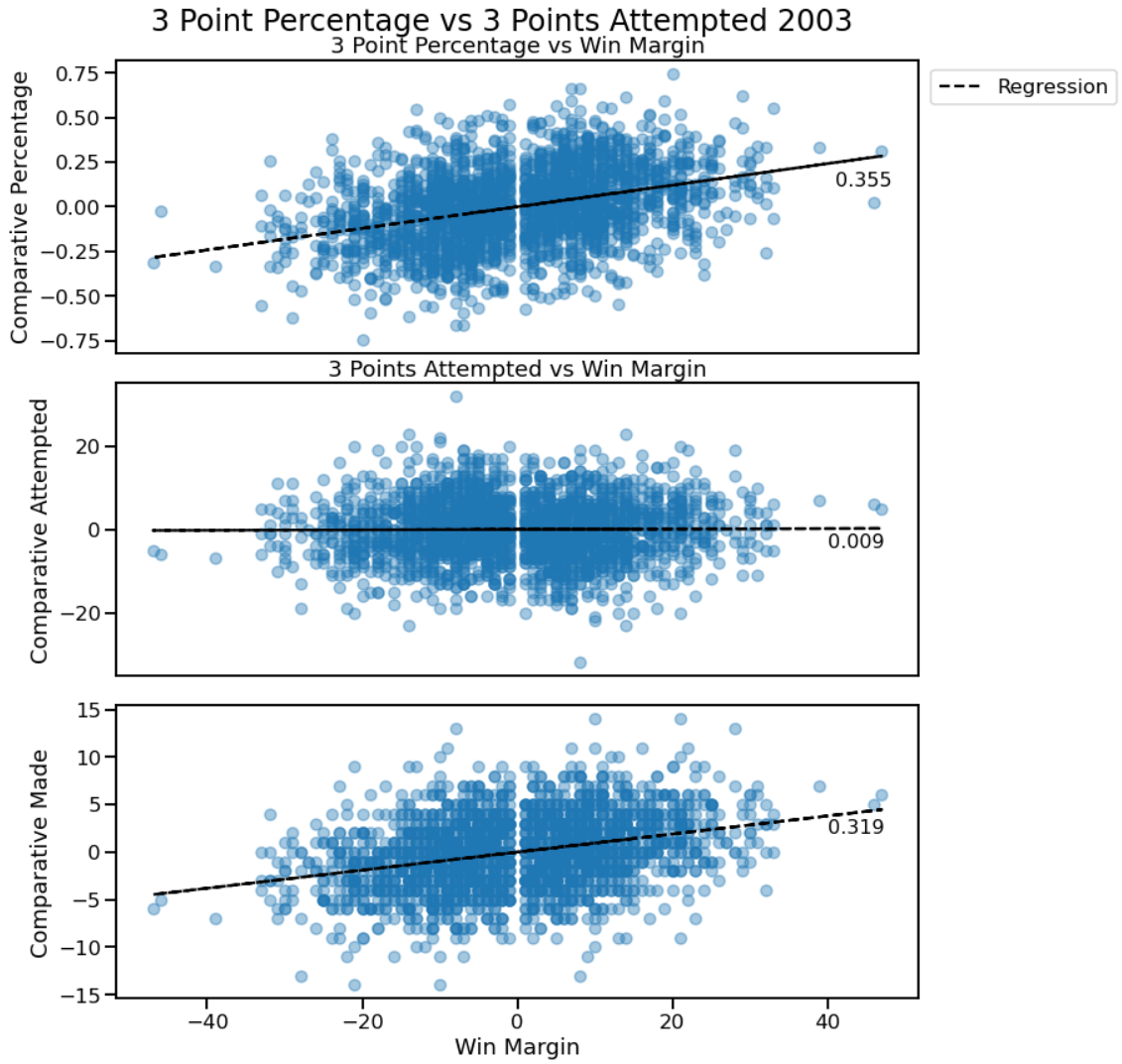

Figure 17. small multiple of 3-point percentage, 3-points attempted, and 3-points made 2003. Winning margin is on the x axis. Linear regression and correlation coefficient is overlayed on each graph.

Figure 18 below shows the 2018 equivalent of Figure 17. We can see there is a stronger correlation in the 2018 season between 3-point percentage and 3-points made and winning margin then there was in the 2003 season. There is a slight increase in the correlation of attempts to winning margin from 2003 to 2018.

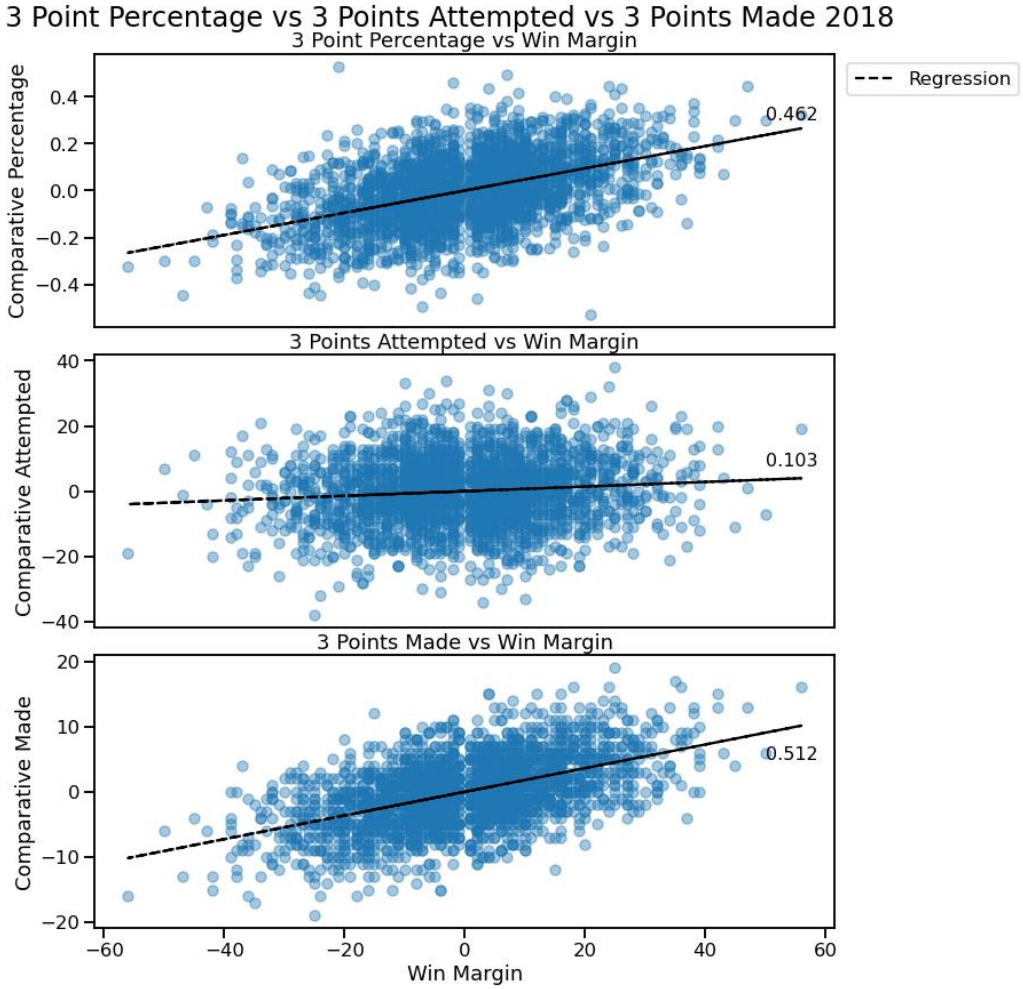

Figure 18. small multiple of 3-point percentage, 3-points attempted, and 3-points made 2018. Winning margin is on the x axis. Linear regression and correlation coefficient is overlayed on each graph.

4.3.4. Discussion

I found that the number of 3-point attempts does not inherently have much of an effect on a team's winning chances. In Figure 14 there is a very slight difference in the effect of having a positive or negative number of 3-point attempts. It's clear that 3-point shot percentage has a much stronger correlation with winning. After seeing this I wanted to find out if making more shots at a higher percentage had more of an impact. This is the idea behind Figures 15 and 16. Instead of looking at the number of shots attempted we are looking at the number of shots scored. This showed a very strong correlation which tells us if you outscore your opponent from 3 and do it at a higher efficiency you are significantly more likely to win. I was surprised to see the bottom right and top left quadrant seem to have little correlation to wins. This tells us that just scoring more 3-point shots than your opponent isn't enough if you attempt a lot of shots at a low success rate and similar for a high completion percentage with a low volume of shots.

I wanted to get a more accurate understanding of how each category of 3-point statistics affect a team's winning chances. Figures 17 and 18 give us a very clear idea of how important each category of the 3-point shot is. I was surprised to see that in 2018, in Figure 18, that shots made were more important than shot percentage. This is in contrast to in Figure 17 where 3-point percentage was slightly more important. I also expected to see 3-point attempts to have a much higher correlation with winning in 2018 than in 2003. Although there was a slight increase it was much less than I expected.

Overall, we can see that the 3-point shot has grown in importance since the 2003 season. In each 3-point category (shots made, shot percentage and shots attempted) there is a stronger correlation between outperforming your opponent and winning in 2018 compared to 2003. From question 2 we found that teams have started attempting much more 3-point shots than they did in the past so it stands to reason that how they perform in this area has become much more important.

4.4. RQ4

How has the game-by-game stats of top players changed over the last 15 years? How does this compare to new star players?

4.4.1. Datasets

For this research question I used the processed file '410_advanced_player_data.csv'. For more information on this dataset check section 3.2.2.4.

4.4.2. Approach

How has the game-by-game stats of top players changed over the last 15 years? How does this compare to new star players?
My approach to these questions was to plot the statistics of the top 30 players each season. I then looked at usage percentage and found that there has not really been much of a change in how players play. The main difference I found in game stats is simply the number of 3-point shots they attempt which we have already covered extensively. However, I did find some interesting patterns in how the Most Valuable Player is chosen.

a) What is valued in an MVP?
In order to answer this question, I plotted the efficiency of the MVP each season to get a general idea of how MVPs were chosen. I then produced sub graphs looking at different statistics that might have been influential in the MVP voting.

b) How do MVPs compare to rookies of the year and the rest of the league?
For this question I plotted the efficiency of the MVP each year compared to the efficiency of the Rookie Of The Year (ROTY). I also plotted the league average efficiency and the most efficient player each season for reference.

4.4.3. Results

How has the game-by-game stats of top players changed over the last 15 years? How does this compare to new star players?

As I mentioned in the approach segment there was very little change in how the top players performed. Figure 19 below shows the increase in number of 3-point shots. This is the most notable change in playstyle.

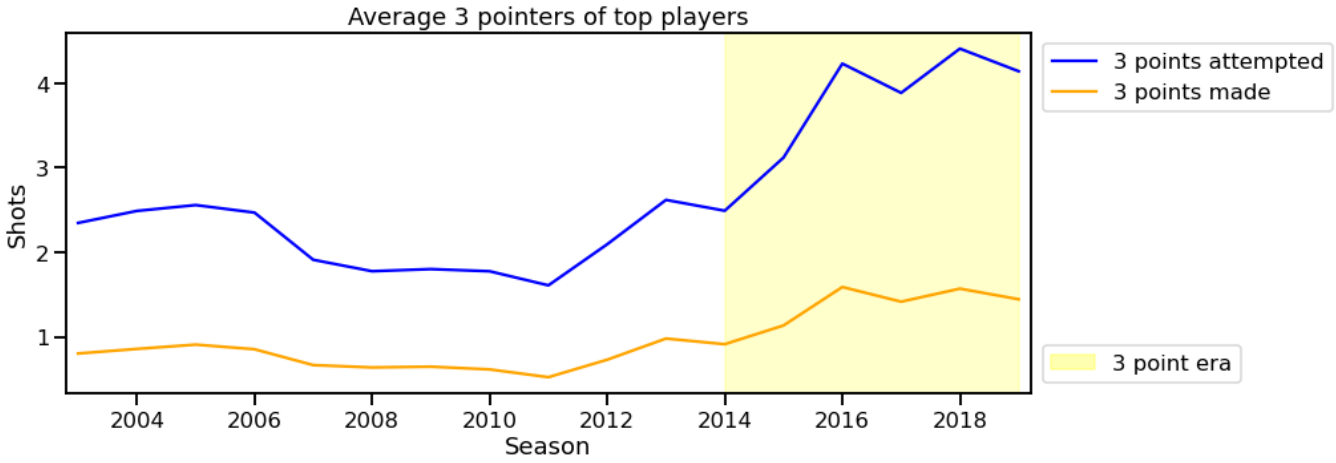

Figure 19. Average 3-points attempted and 3-points made by top players from 2003 to 2019.

a) What is valued in an MVP?

Figure 20 below shows the efficiency of each MVP winner in each season from 2003 to 2019. We can see MVPs generally have a high efficiency, at least 20. The gap between the efficiency of LeBron James and Derick Rose in the 2011 season has been highlighted. Derrick Rose won MVP this season. The efficiency of Steve Nash is also highlighted which I will discuss in Figure 21 and Figure 22. The league average is plotted for reference of an average player.

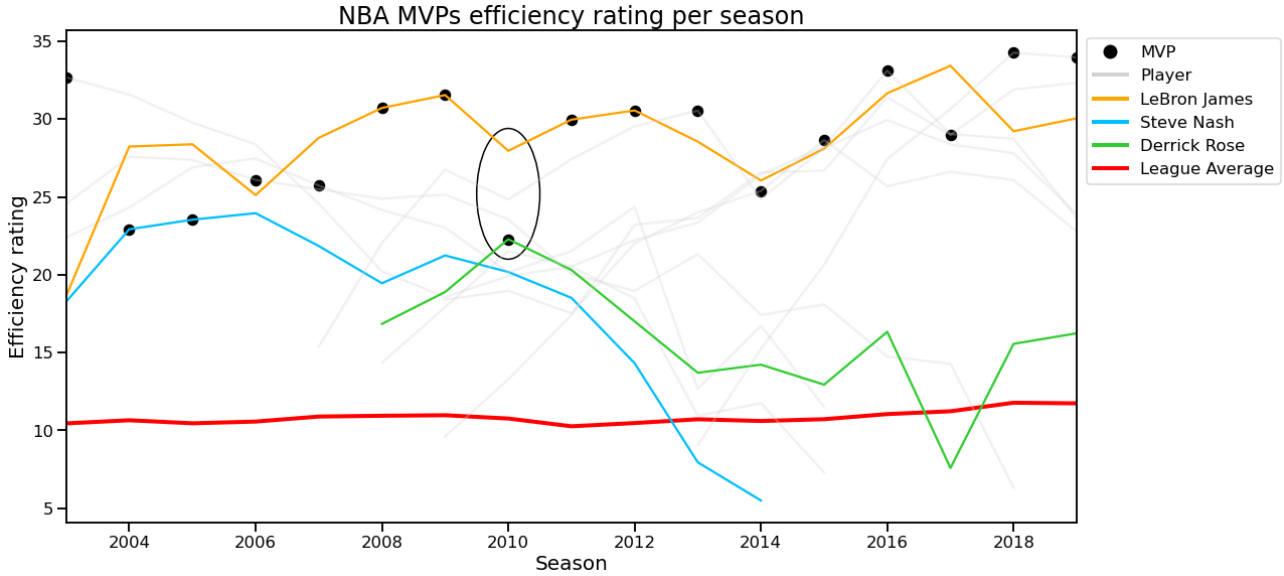

Figure 20. Efficiency of MVPs for each season 2003 to 2019 with LeBron James, Derrick Rose, Steve Nash and the league average are also highlighted.

Figure 21 below shows the number of points each MVP scored per game per season. Steve Nash is significantly lower than the other MVPs.

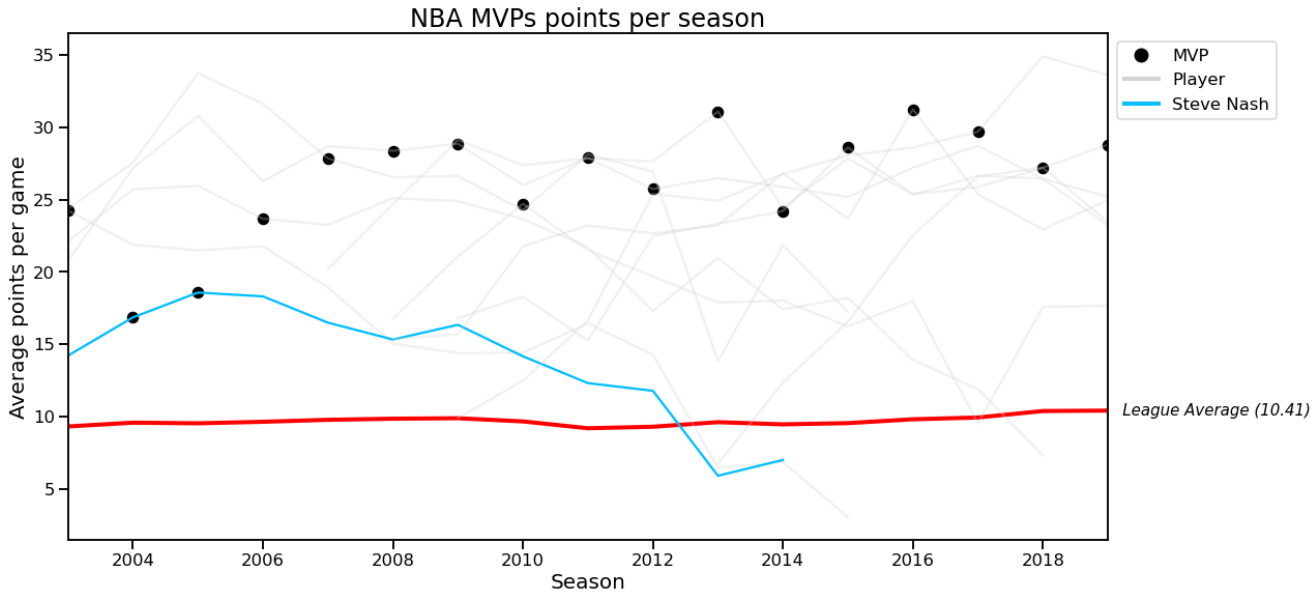

Figure 21. Average points per game by MVPs per season. Steve Nash is highlighted

Figure 22 below shows the average assists per game of MVPs per season. This graph shows where the majority of Steve Nash's efficiency comes from. Steve Nash is far ahead of every other MVP in terms of assists.

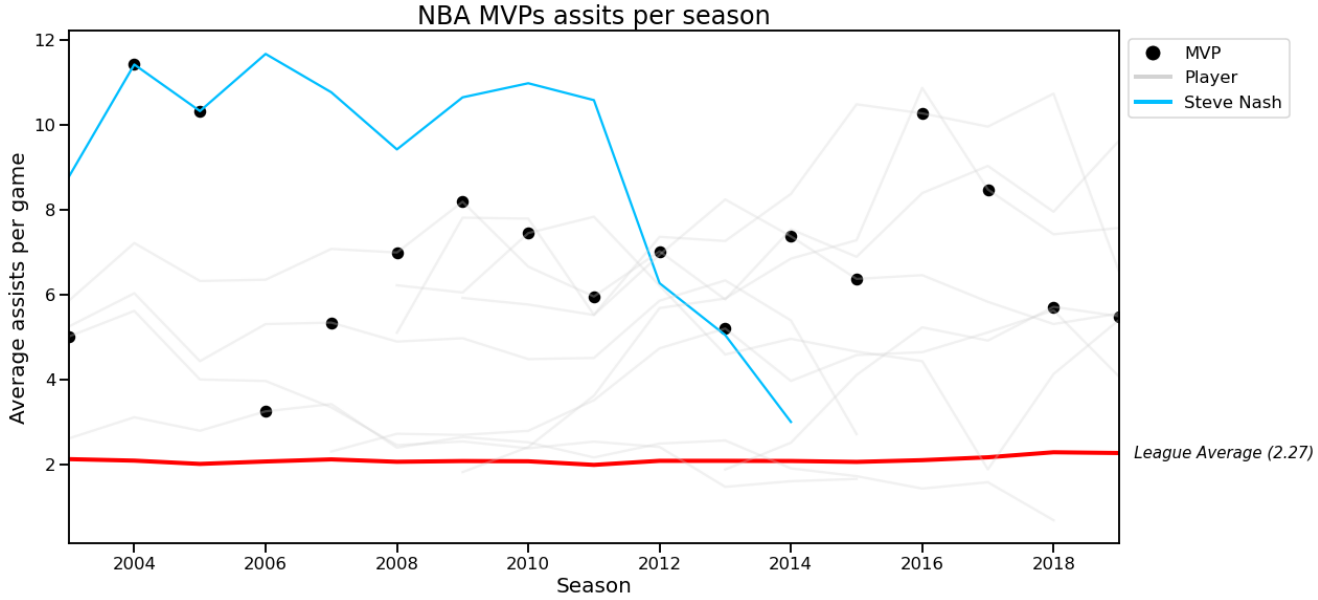

Figure 22. Average assists per game by MVPs per season. Steve Nash is highlighted.

b) How do MVPs compare to rookies of the year and the rest of the league?

Figure 23 below compares the efficiency of the most efficient player in the league, the MVP, the ROTY and the league average. This graph shows us that the most efficient player is often voted as MVP. We can also see that rookies of the year perform much better than the average player in the league.

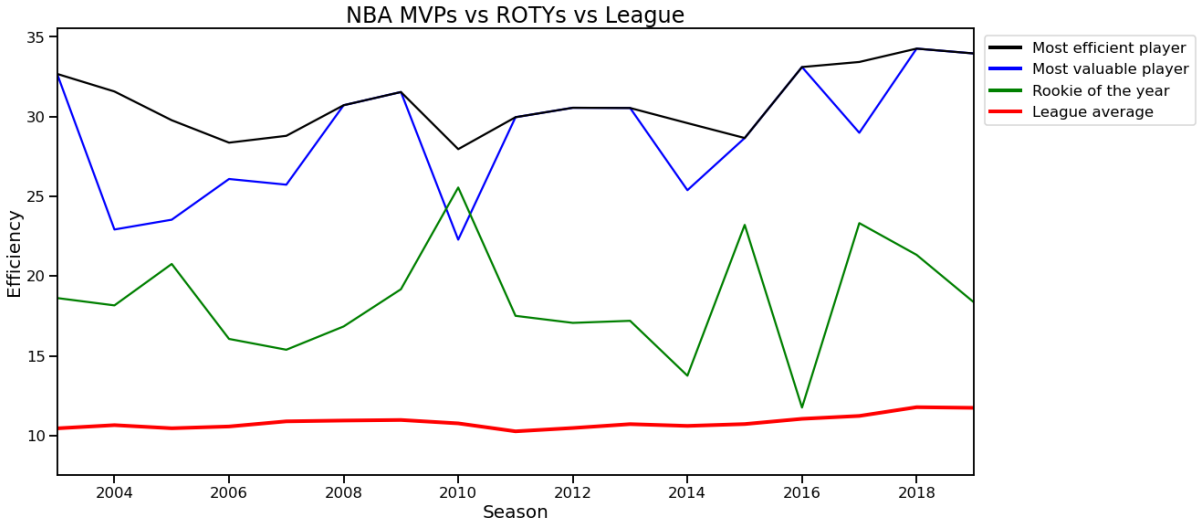

Figure 23. Efficiency of the most efficient player, the MVP, the ROTY and the league average.

4.4.4. Discussion

As I have mentioned a couple of times there was very little change in the game-by-game stats of top players. You can see in Figure 19 that there has been a rise in the number of 3-point shots. This correlates to a slight rise in the efficiency of top players that you can see in Figure A.13. Something I was expecting and was surprised that I didn't find was a change in usage percentage. There was a sentiment in my mind that teams have become more focused on allowing star players to control the ball and have relied less on teamwork in recent seasons. This assumption was clearly proven false from Figure A.14 which shows that top players have generally always controlled a lot of possessions, but it hasn't changed much over time.

While analysing the efficiency of MVPs I did find some interesting patterns. In Figure 20 you can see the efficiency performance of each MVP. I have highlighted the gap between LeBron James and Derrick Rose for the 2010 season. I wasn't watching basketball at this time but I have heard many people claim that LeBron deserved the award in this season [23]. I was surprised to see that the data strongly supports this argument. There are many reasons people give for Derrick Rose winning MVP over LeBron, I think the two most likely reasons are that Derrick Rose's team won more games [24] in the season and LeBron had arguably better teammates than Derrick Rose.

I want to discus Steve Nash's MVP seasons because they are somewhat unique amongst the rest. In Figure 21 you can see that Nash scored at a much lower rate than every other MVP. I mentioned earlier that efficiency is a statistic that rewards offensive play more than defensive play, but I believe it also provides too little weight to team play. Figure 22 shows how Nash was far beyond every other MVP in terms of assists. From Figure 20 you can see that Nash's MVP seasons have the lowest efficiency of any MVP other than Derrick Rose despite his extreme contribution to his team in the form of assists. The only other MVP to come close to Nash in assists was Russell Westbrook in the 2016-17 season [25] when he averaged 10 assists per game. However, you can see in Figure 21 that in 2016 Westbrook also scored a very high number of points. I don't believe there has been another MVP quite like Steve Nash, where their contribution was primarily assisting rather than scoring.

Finally. Figure 23 shows how MVPs compare to ROTY. The rookie of the year is given to the player with the best debut season. I was surprised by how high the efficiency of these rookies was. Many rookies are only around 19 or so when they win ROTY. The winner in 2003 (LeBron James) and 2018 (Luka Doncic) were both 19 when they won the award. It is also worth noting that Blake Griffin had a higher efficiency that Derrick Rose in the 2010-11 season. It is surprising that rookies can perform at such a high level when it is often the first time playing in a professional league for these players.

5. Discussion

5.1. Ethical Considerations

An ethical consideration with this project is coaches and teams using my study to measure the performance of a player and cutting this player from the team. This study was solely carried out to understand more about how teams in the NBA have changed playstyle throughout the last 2 decades, not to weed out underperforming players.

5.2. Reproducibility

My results are easily reproduced from my collection of notebooks and data that I used. A requirements.txt file is supplied with my notebooks.

5.3. Limitations

The main limitations of this study are surrounding defensive analysis. It is very difficult to track defensive performance of basketball players. Information on the position of each player on the court is needed and other data such as if a defensive player allowed an attacking player to run past them. This data is very hard to gather and even harder to analyze.

Another limitation is measuring how heavily contested a shot is. I would like to have the data of how close the closest defender was to the player when they took a shot so I could compare if players take more contested shots than they did in the past and if they score them at a higher percentage than in the past.

6. Conclusions & Future Work

In this study I set out to analyze how the playstyle of the NBA has changed over the past 20 years.

Through this study I learned that the pace of the game has increased over the years and teams have become more efficient on offense in the same time period. This has led to a large increase in scoring. The majority of the growth in offensive efficiency has come from the proficiency of players at 3-point shooting. They are taking more shots from the 3-point range than ever before and they are scoring them at a higher percentage than before. Because of the increased focus on 3-point shooting by players, the 3-point shot has become pivotal to a team's success. I also found that, other than the number of 3-point shots they shoot, top players have performed at a relatively consistent level over the past 20 years. But in a league that has become more focused on scoring there seems to be less respect given to players with a high assist output compared to players with a high scoring level.

As I mentioned in the limitations section, I would like to do future work in this area but I would take a much closer focus on how defensive play has changed. We saw in section 4.1 that the defensive efficiency rating of teams has dropped but this does not necessarily mean that players have gotten worse at defense, it could simply be because players have gotten rapidly better at offense. I would like to analyze some more advanced statistics such as tracking if players let attackers run past them with the ball or if a player on defense managed to gather all the rebounds that were feasibly available to them.

In summation I feel I have answered the questions I set out to answer but throughout my study some more questions have arisen that I would like to look at in future work. work.

7. Responsibilities

I was solely responsible for all work on this project.

8. Bibliography

[1] ‘What happened to defense in the NBA?’ - Bleacher Report, https://bleacherreport.com/articles/302807-what-happened-to-defense-in-the-nba

[2] ‘You Can’t Stop NBA Offenses—and Now, You Can’t Even Hope to Contain Them’ – The Ringer, https://www.theringer.com/nba/2018/10/30/18038802/nba-defense-offensive-boom

[3] NBA viewership, https://www.statista.com/statistics/289993/nba-number-of-tv-viewers-usa/

[4] ESPN debate best basketball team in history, https://www.youtube.com/watch?v=-1vdT0jY_xM&ab_channel=ESPN

[5] Specific quote from [1], https://bleacherreport.com/articles/302807-what-happened-to-defense-in-the-nba#:~:text=when%20i%20was%20younger

[6] Grantland shot charts, https://grantland.com/the-triangle/the-nba-draft-illustrated-shot-charts-for-the-lottery-picks/

[7] The Last Dance, https://en.wikipedia.org/wiki/The_Last_Dance_(miniseries)

[8] The 3 point era, https://steemit.com/english/@danielct/the-three-point-era-nba

[9] First dataset, https://www.kaggle.com/nathanlauga/nba-games

[10] NBA stats website, https://www.nba.com/stats/

[11] nba_api, https://pypi.org/project/nba-api/

[12] possessions formula, https://www.nbastuffer.com/analytics101/possession/

[13] usage percentage, https://www.basketball-reference.com/about/glossary.html#:~:text=usg%25%20

[14] true shooting percentage, https://www.basketball-reference.com/about/glossary.html#:~:text=ts%25%20-%20true

[15] efficiency, https://en.wikipedia.org/wiki/Efficiency_(basketball)

[16] 2011 NBA lockout, https://en.wikipedia.org/wiki/2011_NBA_lockout

[17] GSW record breaking season, https://www.nba.com/warriors/news-blogs/breaking-down-73-win-season-20200616

[18] List of NBA champions, https://en.wikipedia.org/wiki/List_of_NBA_champions

[19] Offensive efficiency rating, https://www.nbastuffer.com/analytics101/offensive-efficiency/

[20] Basketball positions, https://jr.nba.com/basketball-positions/

[21] Stephen Curry unanimous MVP, https://sportsworld.nbcsports.com/steph-curry-unanimous-mvp/

[22] Stephen Curry most 3-pointers in a month, https://bleacherreport.com/articles/2941223-video-warriors-stephen-curry-sets-nba-record-for-most-3-pointers-in-a-month

[23] LeBron deserved MVP 2010-11, https://www.opencourt-basketball.com/lebron-james-former-teammate-says-derrick-rose-shouldve-never-won-mvp-in-2011/

[24] 2010-11 NBA standings, https://www.basketball-reference.com/leagues/NBA_2011_standings.html

[25] Russell Westbrook MVP season, https://theundefeated.com/features/russell-westbrook-takes-home-nba-mvp-award/

9. Appendix

9.1. Graphs

Fig A.1

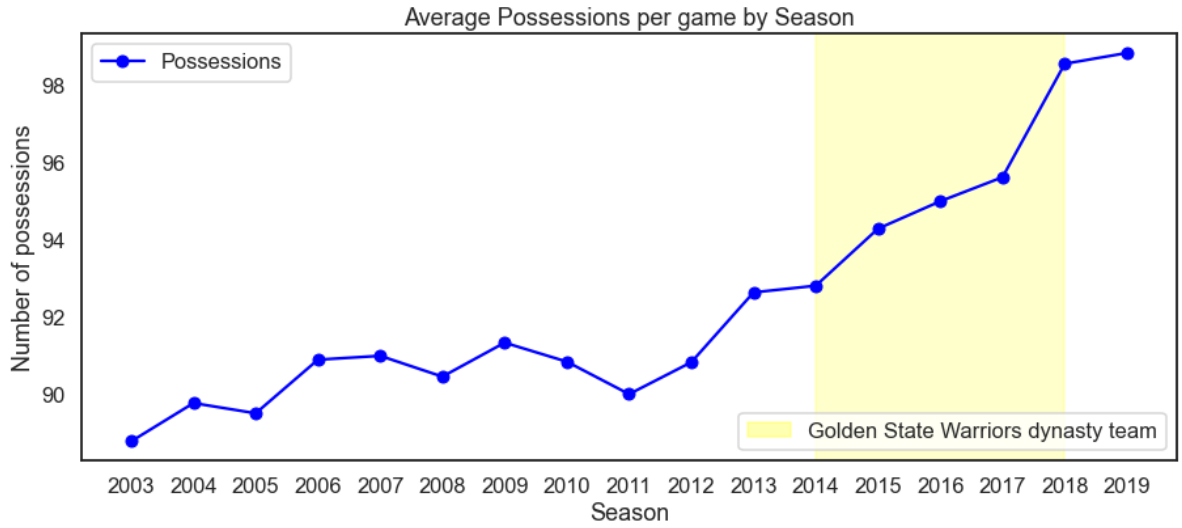

Figure A.1, Average possessions per game by season

Fig A.2

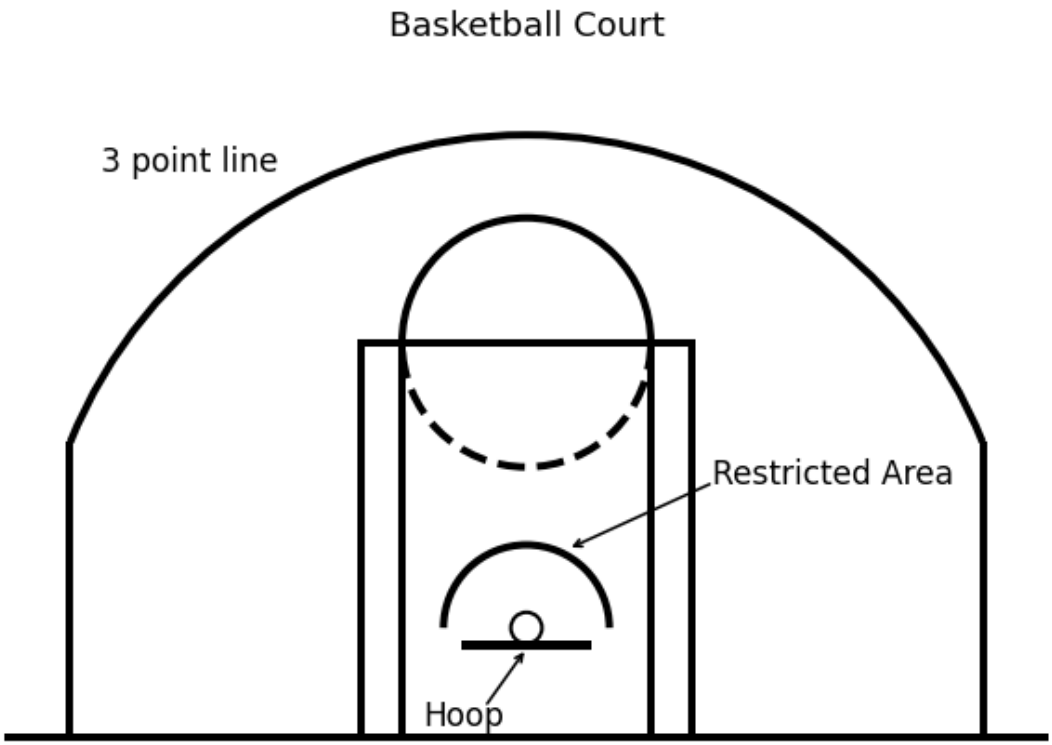

Figure A.2, half of a basketball court

A.3 to A.6

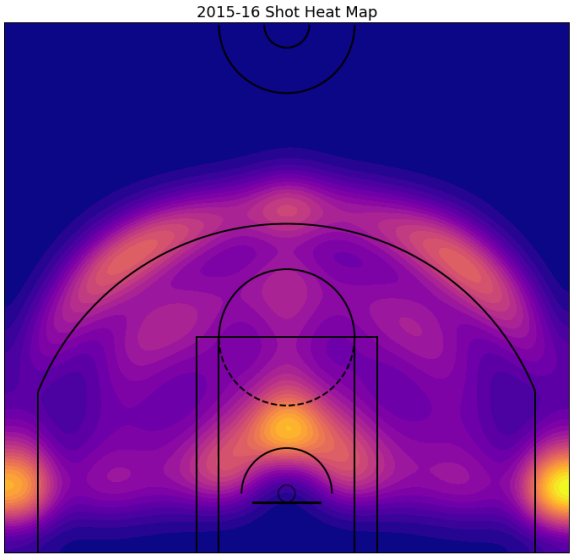

Figure A.3 shot heatmap 2015-16

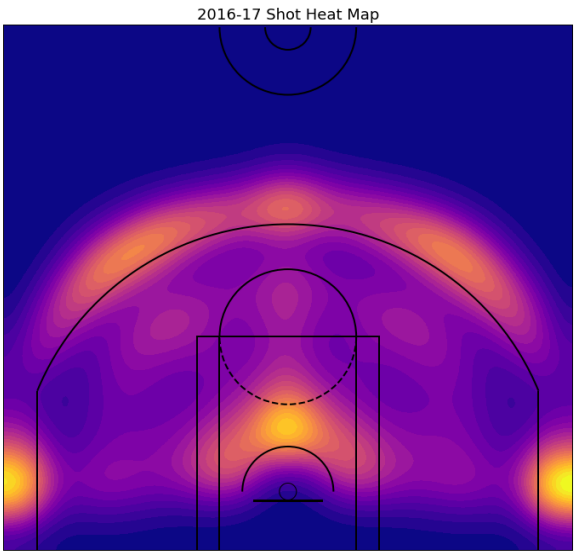

Figure A.4 shot heatmap 2016-17

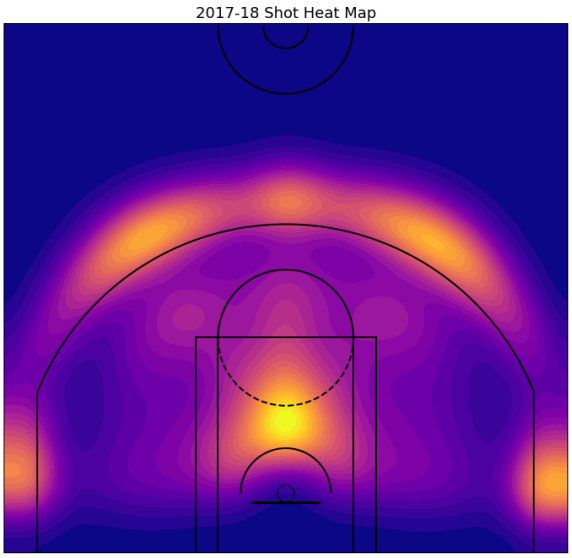

Figure A.5 shot heatmap 2017-18

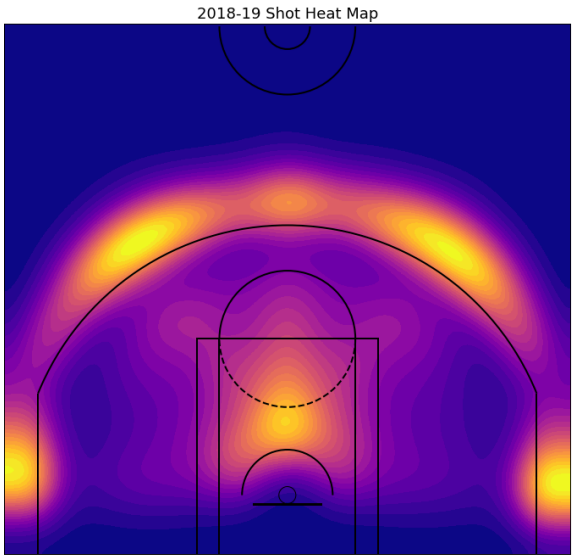

Figure A.6 shot heatmap 2018-19

A.7 to A.12

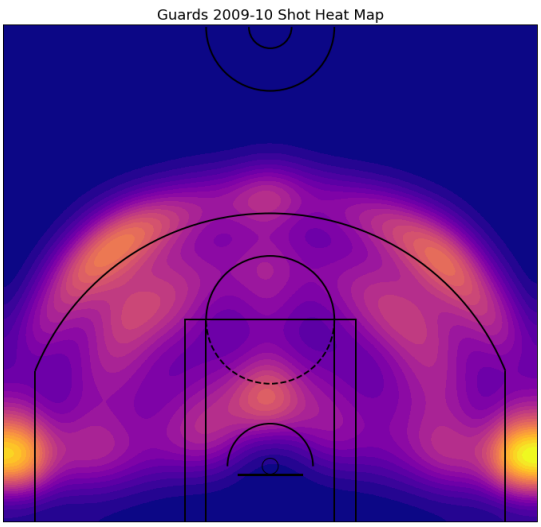

Figure A.7 Guards shot heatmap 2009-10

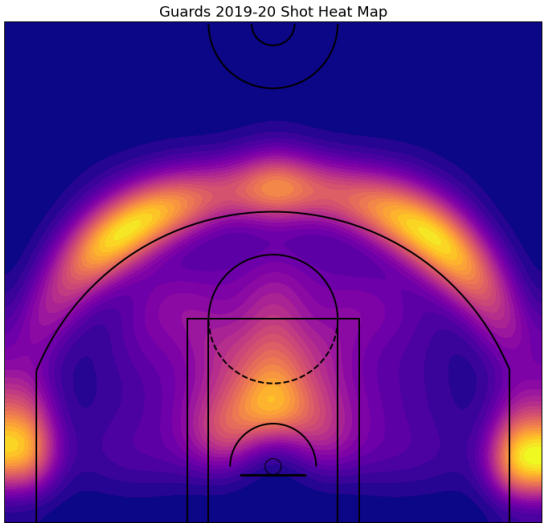

Figure A.8 Guards shot heatmap 2019-20

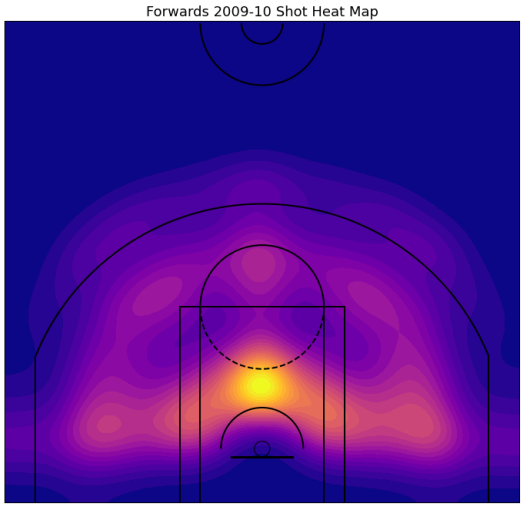

Figure A.9 Forwards shot heatmap 2009-10

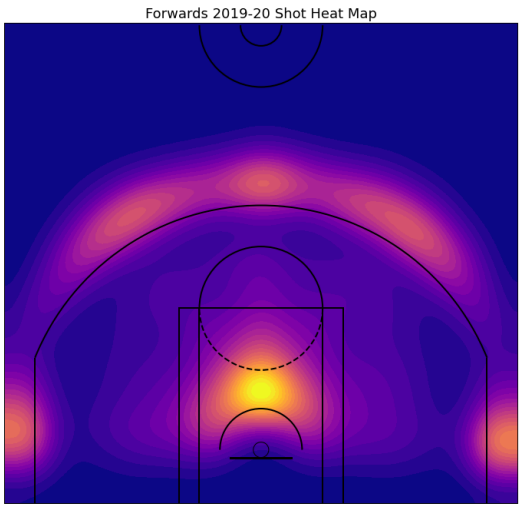

Figure A.10 Forwards shot heatmap 2019-20

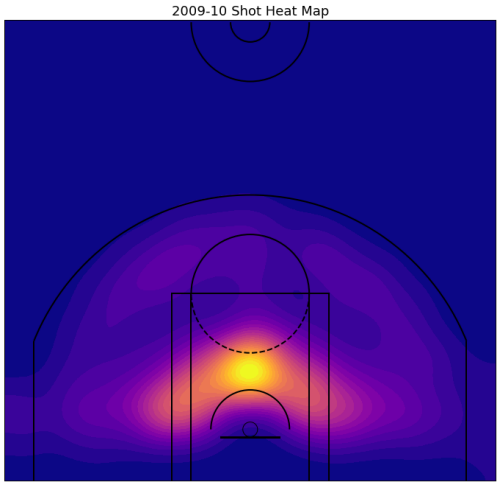

Figure A.11 Centers shot heatmap 2009-10

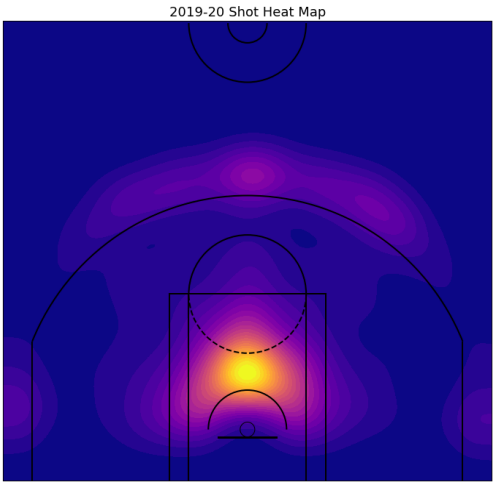

Figure A.12 Centers shot heatmap 2019-20

A.13

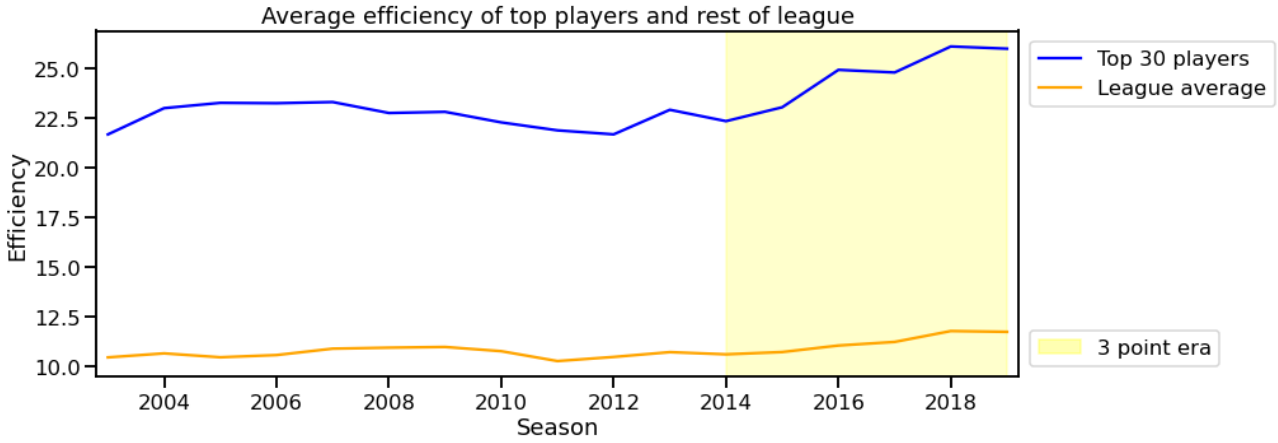

Figure A.13. Efficiency of top players compare to league average per season.

A.14

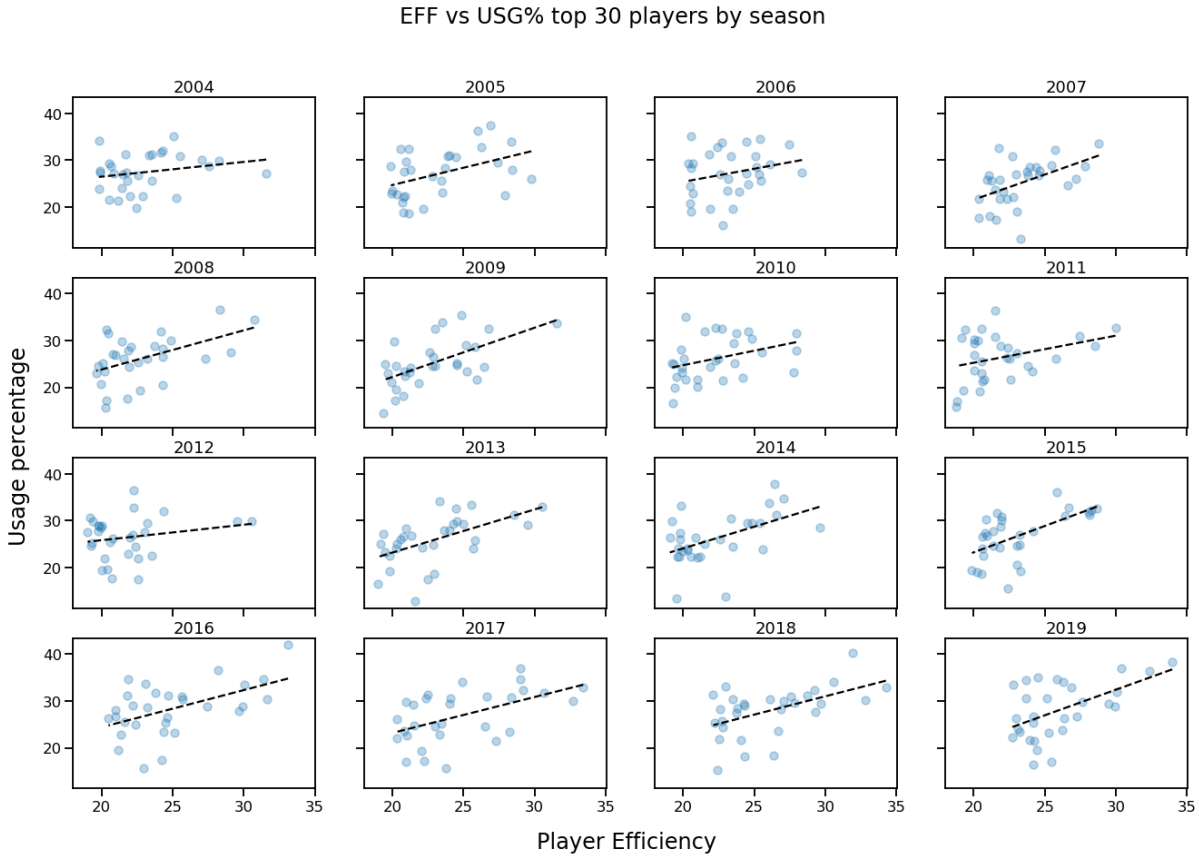

Figure A.14. Efficiency vs usage percentage of top 30 players per season.