NBA Player’s Performance in Regular Seasons vs Playoffs

Haiting Huang, Matthew Brennan, Ermin Pinjic, Haorui Zhang

Contribution

Haiting Huang: Linear regression model, Analysis revision
Matthew Brennan: Linear Regression model, PCA, Consistency Analysis, Abstract
Ermin Pinjic: Correlation Matrix, OBPM vs USG, Discussion, Background
Haorui Zhang: OBPM vs WS/48, Analysis, Format

Abstract

A common assumption among NBA fans is that certain superstar players raise their levels of performance during playoffs, in which the winning team is declared the NBA champion. It became of interest to us to analyze NBA performance data to see if this idea was of merit. Our aims were to discover a way to predict which players would raise their performance in the playoffs based on their regular season statistics, and to identify which players most consistently raised their performance in the playoffs. Ultimately, we found none of our observed statistics were good predictors of increased performance in the playoffs. We were able to identify four players (Kevin Durant, Kawhi Leonard, Jamal Murray, John Wall) who consistently were the best players in terms of difference between playoff and regular season performance.

Introduction

Background

The NBA, the one place where the most talented basketball players from all over the world play together -- or against each other, in order to accomplish one goal; to win their team a championship. As players continue through the regular season, their game statistics are recorded and averaged out after the 82 games are finished. If a team is able to get into the playoffs, their statistics will continue to be recorded during the playoffs. Gathering data from a reliable sports statistics website, there are several catch-all variables which can be used to describe overall player performance, including win shares per 48 minutes and offensive box plus/minus. It became increasingly of interest to us to discover whether certain players’ statistics would change during playoff basketball, where the stakes of winning are much higher.

Aims

1. Our first aim was to see if there was a way to predict a player’s difference in performance in the playoffs vs. the regular season. We created scatterplots to chart certain statistics against our difference-in-performance variables. Then, using a selected group of players, we utilized a linear regression analysis to try to explain our difference-in-performance WS/48 variable. Ultimately, we found that no regular season statistics were able to significantly predict difference in the win shares per 48 minutes statistic between the playoffs and regular season.
2. Our second aim was to identify the players who most consistently appeared to perform better in the playoffs. Using the same group of players, we created scatterplots to graph their difference-in-performance WS/48, then grouped players’ seasons and identified players with the

highest mean value across seasons and lowest variation. This analysis determined Kawhi Leonard, Kevin Durant, Jamal Murray, and John Wall to be the best players at consistently performing better in the playoffs, though this difference in performance was moreso a consistent continuation of regular season performance into the postseason, while all other players seemed to perform worse.

Materials and Methods

Datasets

Our data contains information on NBA player statistics during the regular season and playoffs of one NBA season, with data from 2017-2021. Our data came from *basketball-reference.com*, which is part of a larger group of websites called Sports Reference which collects data for sports such as basketball, football, baseball, etc. The data on *basketball-reference.com* is provided by SportRadar, which is the official statistics provider of the NBA. THeir data is presumably collected by watching and recording the events of NBA games.

The population of our data is all players from 2017-2021 who played in both the regular season and playoffs that season. There is no scope of inference for this data (it is not generalizable) because it is an exhaustive list of all NBA players who played in both the regular season and playoffs for these years, not a sample. Each observational unit is one NBA player.

The most important variables used as part of our analysis are listed below:

Variable Descriptions

Name	Variable description	Type	Units of measurement
Player	Name of player and code from initial website	Character	n/a
Age	Age of player during season	Numeric	Age
Tm	Team the player played for	Character	n/a
MP_regseason	Minutes played per game (regular season)	Numeric	Minutes per game
FG%_regseason	Percentage of Field Goals Made (regular season)	Numeric	Field goals made/Field goals attempted_
USG%_regseason	Usage Percentage (regular season)	Numeric	Percentage of team's plays that the player "ended" with a Field Goal Attempt, Free Throw Attempt, or Turnover
WS/48_regseason	Win Shares per 48 minutes (regular season)	Numeric	read https://www.basketball-reference.com/about/ws.html (https://www.basketball-reference.com/about/ws.html)
OBPM_regseason	Offensive Box Plus-Minus (regular season)	Numeric	read https://www.basketball-reference.com/about/bpm2.html (https://www.basketball-reference.com/about/bpm2.html)
MP_playoffs	Minutes played per game (playoffs)	Numeric	Minutes per game
FG%_playoffs	Percentage of Field Goals Made (playoffs)	Numeric	Field goals made/Field goals attempted_
USG%_playoffs	Usage Percentage (playoffs)	Numeric	Percentage of team's plays that the player "ended" with a Field Goal Attempt, Free Throw Attempt, or Turnover
WS/48_playoffs	Win Shares per 48 minutes (playoffs)	Numeric	read https://www.basketball-reference.com/about/ws.html (https://www.basketball-reference.com/about/ws.html)
OBPM_playoffs	Offensive Box Plus-Minus (playoffs)	Numeric	read https://www.basketball-reference.com/about/bpm2.html (https://www.basketball-reference.com/about/bpm2.html)

Additionally, we included numerous more variables which were only used in our linear regression model, such as Points per Game, Assists per Game, Blocks per Game, etc.

Some example rows of our dataset are shown below:

	Player	Age	Tm	MP_regseason	FG%_regseason	USG%_regseason	WS/48_regseason	OBPM_regseason	MP_playoffs	FG%_pl
0	Bogdan Bogdanović\bogdabo01	28	ATL	29.7	0.473	21.3	0.142	3.3	33.2	
1	Clint Capela\capelca01	26	ATL	30.1	0.594	19.9	0.207	2.7	31.6	
2	John Collins\collijo01	23	ATL	29.3	0.556	22.2	0.174	2.5	32.0	
3	Kris Dunn\dunnkr01	26	ATL	11.3	0.083	16.2	-0.202	-13.7	6.6	
4	Bruno Fernando\fernabr01	22	ATL	6.8	0.409	14.4	-0.019	-6.1	2.0	

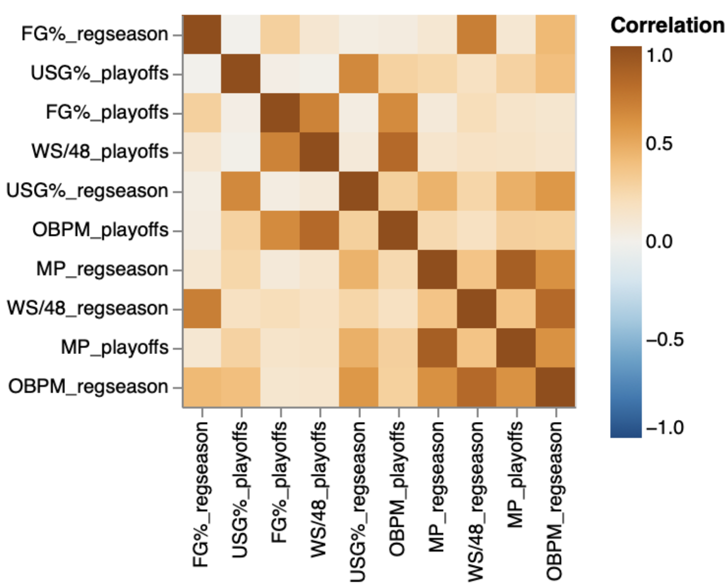
Methods

We firstly wanted to identify the best players in the NBA in these seasons so that our analysis would focus on the players with the highest value and most impact. We used a Principal Component Analysis (PCA) to create one combined numeric value to describe the players, and selected the top 50 from this analysis. Three players were arbitrarily removed since their inclusion in the group was due to extremely small sample size. Next, we calculated variables that were the difference in playoff and regular season performance using our catch-all metrics WS/48 and OBPM. We made exploratory scatterplots to check if certain variables had discernable relationships with these difference-in-performance variables. Finally, we made a linear regression model using regular season statistics to see definitively if any of the statistics could significantly explain our difference-in-performance WS/48 variable.

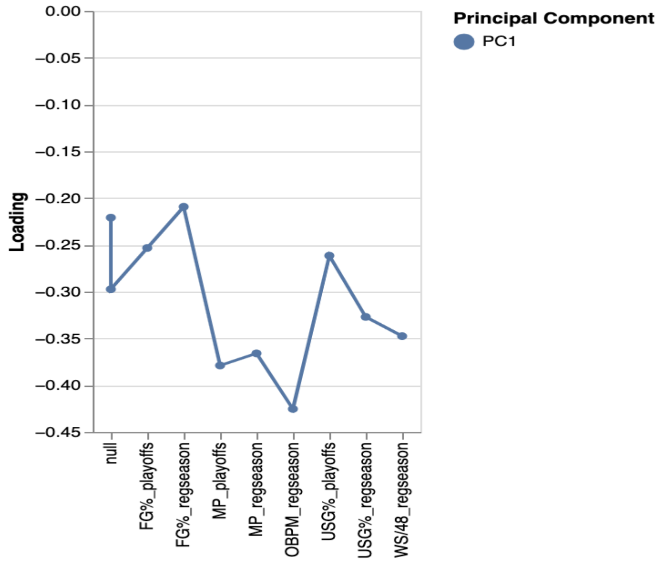
To determine the players who most consistently appeared to perform better in the playoffs, we took our selected group of players and charted their difference-in-performance WS/48, colored by season. Noticing some players had vastly different performance differences across seasons, while others posted a similar value each season, we decided to group each player's seasons and graph their average difference-in-performance WS/48 against the standard deviation, with the intention of discerning the players who consistently performed better in the playoffs as those with the highest mean and lowest standard deviation.

Results

Principal Component Analysis

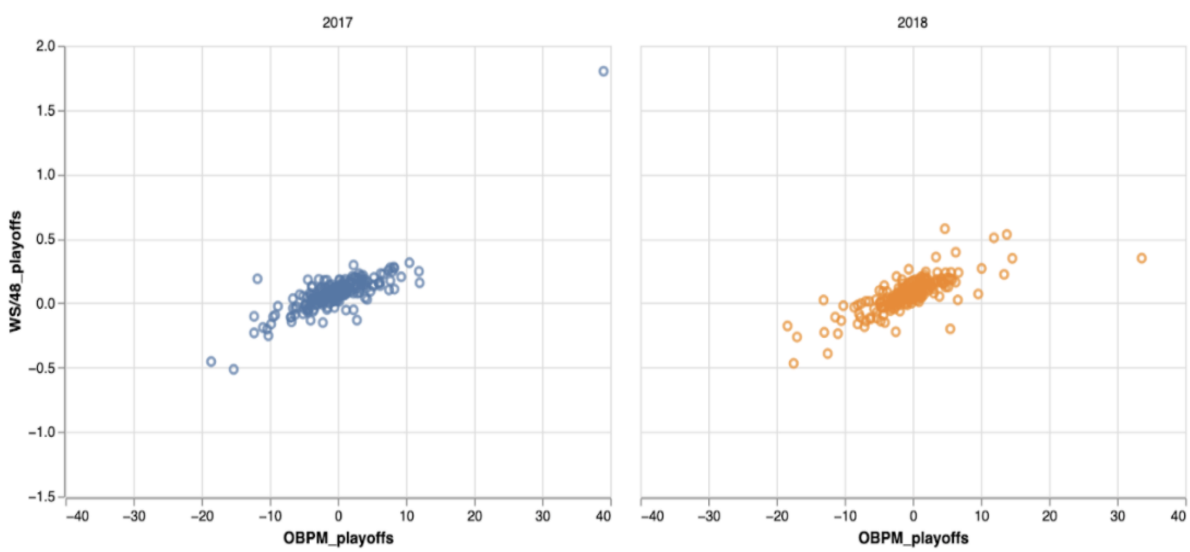


A correlation matrix of the variables used in our PCA. After inspecting closely, it is worthy to note that playoff variables have a very weak correlation to regular season variables, excluding USG % and MP. This could be because the regular season stats account for a larger span of time and games, compared to the playoffs where the max amount of games a player can partake in is 30; accounting for the play-in games and if a team were to play a seven game series for every matchup. The reason these two variables have a stronger correlation compared to the others, is because a player who plays more minutes is bound to be viewed as a top player on their team. Therefore, they are also finishing the majority of the team's possessions.

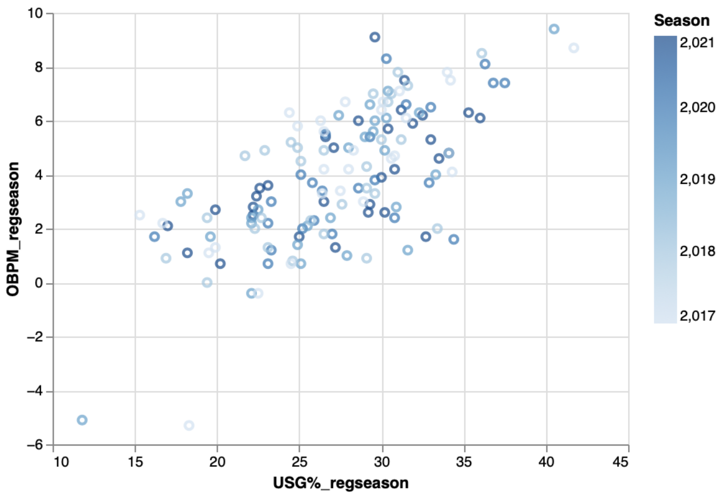


The loading values of our principal component 1, which was the only component that explained a significant proportion of the variance. This component grouped players with high WS/48, USG%, OBPM, and Minutes Played. The top 50 players with the lowest PC1 score were selected for our later regression analysis, with 3 players arbitrarily removed for low sample size.

Exploratory Analysis



OBPM in the playoffs charted against WS/48 in the playoffs. This clearly positive relationship confirmed our expectation that these catch-all metrics were very similar in their evaluation of player performance, and thus could be used almost interchangeably for our analysis.

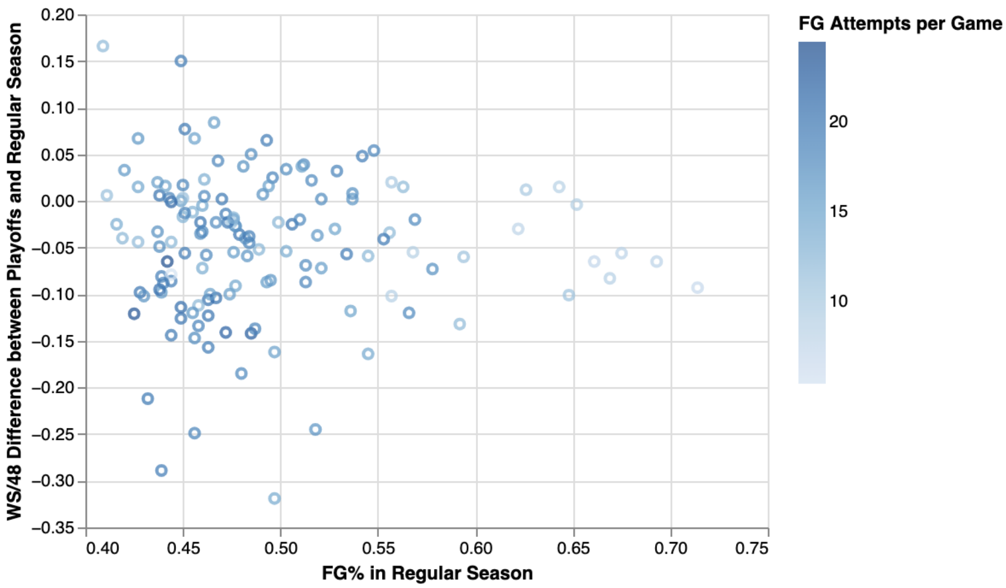


The plot above shows usage % in the regular season against OBPM in the regular season. The strong correlation between the two variables led us to plot their relationship. The graph shows that players who end their team's possessions with a shot attempt or turnover the most often are typically also the players with the higher OBPM rating. Intuitively, the players who the team wants to take a shot or have the ball in their hands are likely the team's best players. Players with a high usage % and low OBPM are likely turnover-prone.

Linear Regression Model

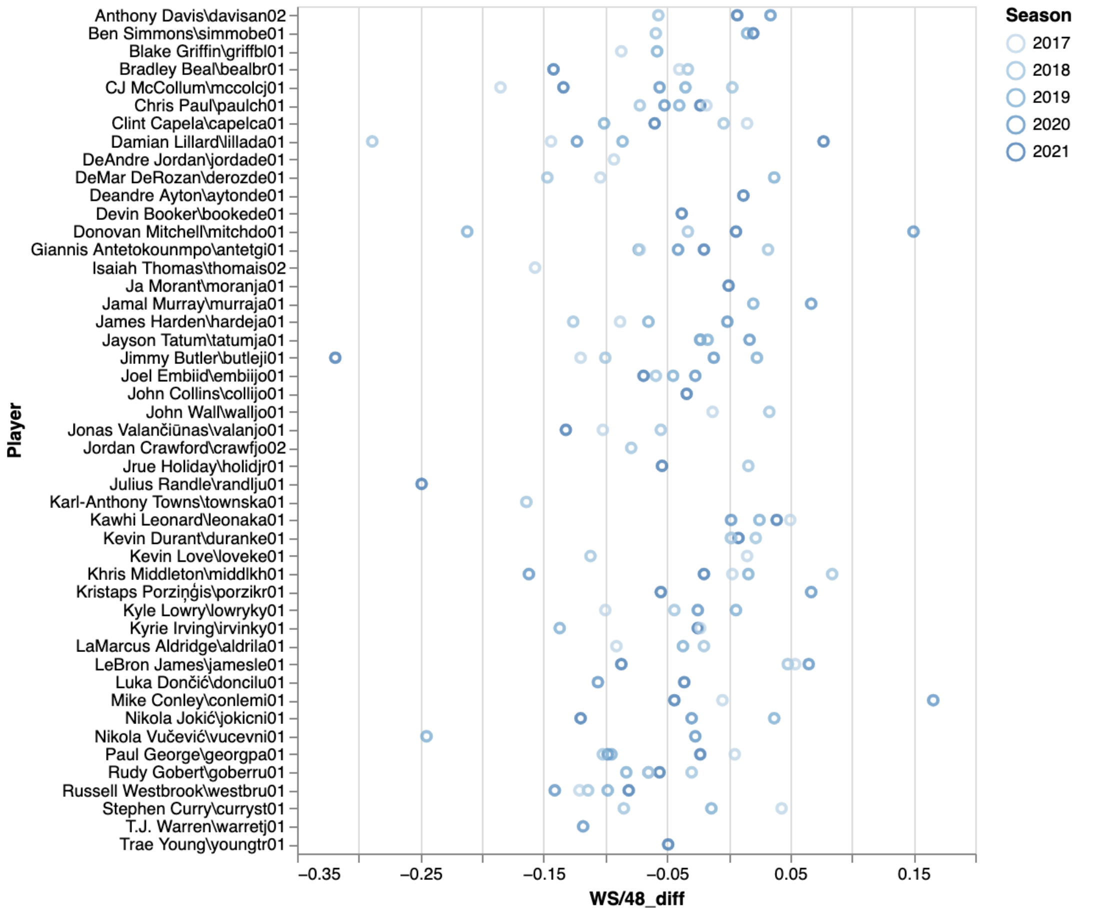
	Variable	Coefficient
0	Age	-0.000255095
1	G_regseason	0.000834597
2	MP_regseason	0.0027459
3	FG_regseason	0.128637
4	FGA_regseason	-0.05729
5	FG%_regseason	7.44395
6	3P_regseason	0.248138
7	3PA_regseason	-0.11586
8	3P%_regseason	-0.383467
9	eFG%_regseason	-6.42391
10	FT_regseason	0.0316826
11	FTA_regseason	0.0310604
12	FT%_regseason	0.550346
13	TRB_regseason	-0.087854
14	AST_regseason	-0.0833353
15	STL_regseason	0.447937
16	BLK_regseason	0.0420517
17	TOV_regseason	0.143373
18	PF_regseason	0.0275547
19	PTS_regseason	-0.0170599
20	PER_regseason	-0.0152513
21	TS%_regseason	-2.75179
22	3PAr_regseason	1.95061
23	FTr_regseason	-0.354563
24	TRB%_regseason	0.0462751
25	AST%_regseason	0.0166594
26	STL%_regseason	-0.330672
27	BLK%_regseason	-0.0117263
28	TOV%_regseason	-0.0273642
29	USG%_regseason	-0.0187182
30	WS/48_regseason	-0.16499
31	OBPM_regseason	-0.0468999
32	DBPM_regseason	-0.0700993
33	BPM_regseason	0.0987367
34	VORP_regseason	-0.0249496

The coefficients of the variables used in the linear regression model to explain the difference between Win Shares per 48 (playoffs) and Win Shares per 48 (regular season). After inspection, none of the variables had any significant effect on this difference in performance. The variables that appear to have large coefficients (FG% and eFG%) only appear this way because their data was in decimal format rather than percentage format (e.g. 40% was input as 0.4 instead of 40), while all other percentage variables were in percentage format. Our R^2 value was 0.33, confirming that these variables are not explaining the difference in WS/48 to a significant degree, and other factors are involved in.

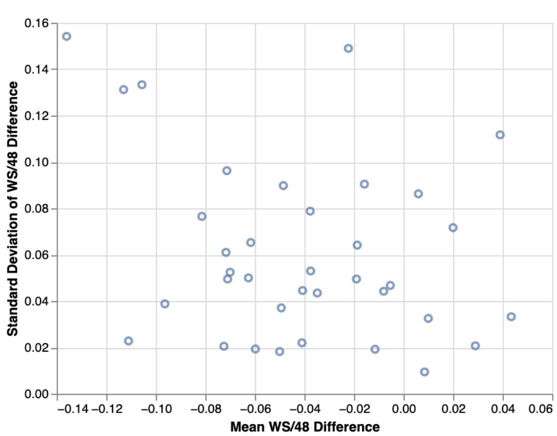


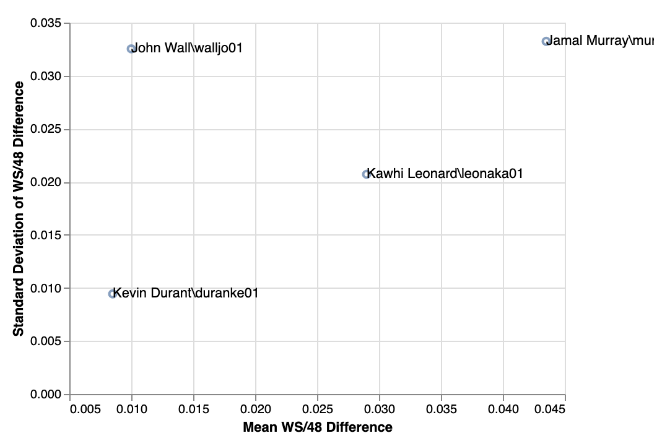
FG% in Regular Season against Difference in Playoff and Regular Season Performance. While trying to discover why FG% had such a large coefficient, we created a scatter plot. There is no apparent linear trend. However, it seems that players who shoot a higher percentage (and shoot less attempts) show little change in performance in the playoffs, while players who shoot a lower percentage (and with more frequency) are more unpredictable. Closer inspection found that the low frequency, low volatility players were taller players who shoot closer to the basket, and the high frequency, high volatility players were shorter players who shot from a farther distance.

Checking for Consistency Across Seasons



Differences in Playoff vs. Regular Season performance for our chosen 47 players. Most players seem to regress slightly in the playoffs, and most players' performance difference varies from season to season. A few (Kawhi Leonard, Kevin Durant) are consistent with little variance and little chance in performance. One player (Russell Westbrook) appears to perform consistently worse in the playoffs.





Mean WS/48 Difference in Performance against STD of WS/48 Difference in Performance. Noticing differences in consistency in our last graph, this chart is our attempt to find players who had the highest difference in playoff vs. regular season performance (increased level of play in the playoffs) and also a small variance (their changes in performance didn't vary widely). The second graph zooms in on the bottom right section of the first. By this analysis, we determine Kevin Durant, Kawhi Leonard, John Wall, and Jamal Murray to be the best players at raising their performance in the playoffs consistently across seasons. It's worth noting that this "increase" in performance is miniscule, and they're rather the best players at keeping their playoff performance on the same level as their regular season performance.

Discussion

By having a sample size of 1,063, we were looking forward to being able to compare how a player's regular season statistics would differ from their playoff statistics. After deciding that our sample was too large, we decided to cut the sample to only the top 47 players in the NBA currently in order to get a deeper understanding of how the selected variables varied amongst the better players in the league. Ultimately, our attempt to use a linear regression model gave us inconclusive results; we were unable to find any variables that appeared to explain the difference in playoff performance and regular season performance. A player's ability to increase play in the postseason is simply too difficult to predict and likely random or due to other factors, as confirmed by our low R^2 value. From our analysis for consistency across seasons, It does appear that the best players in terms of high performance "difference" were more so keeping their performance constant rather than showing a significant increase. This could suggest that a player who keeps playing the way that they have been playing would help the team get further in the playoffs, compared to one who decided to change up the way they play.

As seen in the plot of players' win shares per 48 minutes, the majority of the players become slightly worse in the playoffs; which can be attributed to extra pressure put on them by the fans, and even by the organization itself. However, two standout players, Kawhi Leonard and Kevin Durant, who have respectively won more than one championship, have stayed consistent throughout the playoffs, allowing them to win their teams a championship. Referring to the plot once again, Russell Westbrook is an interesting player, who tends to allow the heckling from fans to get the best of him. The playoffs, where the oppositions die hard fans come out to support their team clearly has played a significant role in this specific player's playoff record, since he has yet to win a championship.

If we were to do this project again, we believe that we should take statistics from more seasons and average them out to see how a player's performance has either improved or declined over the years. This could allow us to inspect when a player truly reaches the 'prime' of their career, where they are playing to the best of their ability. Furthermore, it would increase the sample size that we have, which could help out spread the data in our plots, and make it more clear on how playoff basketball affects a player's statistics compared to a regular season game.